

ESCUELA POLITÉCNICA NACIONAL

**ESCUELA DE POSGRADO EN CIENCIAS
ADMINISTRATIVAS Y ECONÓMICAS**

**METODOLOGÍA DE CLASIFICACIÓN POR SERVICIOS A
AGENCIAS BANCARIAS PRIVADAS**

**TESIS PREVIA A LA OBTENCIÓN DEL GRADO DE MÁSTER (MBA) MENCIÓN
GESTIÓN DE PROYECTOS**

DORIS NATALIA CAICEDO LONDOÑO

DIRECTOR: Doctor PhD. HUGO BANDA GAMBOA

QUITO, DICIEMBRE 2005

DECLARACIÓN

Yo, Doris Natalia Caicedo Londoño, declaro que el trabajo aquí descrito es de mi autoría, que no ha sido anteriormente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos de propiedad intelectual, según lo establecido por la Ley de Propiedad Intelectual y la normatividad institucional vigente.

Doris Caicedo L.

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por: Doris Natalia Caicedo Londoño, bajo mi supervisión:

Doctor PhD. Hugo Banda

AGRADECIMIENTO

Agradezco de manera especial Al Dr. Hugo Banda por su guía en el desarrollo de esta tesis. Además, a María Belén, Vinicio y Daniela por su ejemplo.

DEDICATORIA

A la memoria de mi padre, a mi madre y a César, mi esposo.

RESUMEN

Actualmente, las instituciones bancarias construyen Data Ware House con datos de sus clientes sobre: las transacciones que realizan, los productos que adquieren, etc., pero cómo estos datos se convierten en información?. Una solución es el desarrollo de metodologías de clasificación de comportamientos aplicando métodos multivariantes.

Estas metodologías son muy diferentes de los esquemas de clasificación tradicionales, los cuales son intuitivos y son definidos en base a pocas variables, puesto que el trabajo de combinarlas es extenso y su interpretación muy compleja.

En esta tesis se trata de ilustrar y mostrar la apropiada aplicación de una metodología de clasificación de las agencias de una institución bancaria basada en pasos lógicos aplicando métodos multivariantes.

Una de las consecuencias de esta metodología es la transformación de los datos en información relevante para la toma de decisiones. Esta información se aplica para identificar y solucionar los problemas objeto de investigación. Esto implica que la información debe contribuir a reducir la incertidumbre, ha de ser susceptible de influir en las decisiones y justificar su coste.

La clasificación de las agencias permite una disminución de costos en la institución bancaria y la aplicación de planes estratégicos eficientes, que dependen de las características de cada uno de los grupos.

PRESENTACIÓN

La presente tesis trata de la aplicación de una metodología de clasificación de agencias bancarias privadas basada en técnicas estadísticas multivariadas como son: el análisis factorial, el análisis de componentes principales y el análisis discriminante, de una forma general y práctica complementariamente el análisis cluster y el análisis de correlaciones.

Este tipo de metodología es completa y permite al investigador mostrar que los segmentos formados tienen un mejor acercamiento a la realidad, siendo la toma de decisiones más acertada.

En el primer capítulo se presenta la justificación de la aplicación de esta metodología y los objetivos de la investigación. Seguidamente, el sustento técnico matemático y estadístico de la propuesta, el mismo que es de fácil entendimiento. Otro capítulo corresponde a la explicación de la metodología de clasificación.

Como parte práctica de la tesis se analizan los datos recolectados por las agencias de una institución bancaria privada. Los datos corresponden al número de transacciones que realizan los clientes en las agencias, el número de clientes según la segmentación interna de la institución bancaria, número de colaboradores que tiene el banco en cada agencia, índices de eficiencia y rentabilidad, entre otros. Estos datos serán aplicados en el proceso lógico de la metodología propuesta.

El análisis de los datos y la aplicación de la metodología se aplica en el paquete estadístico SPSS.

ÍNDICE

CAPITULO I	14
1.1 INTRODUCCIÓN	14
1.2 JUSTIFICACIÓN	26
1.3 OBJETIVOS.....	29
1.3.1 OBJETIVOS ESPECÍFICOS	29
CAPITULO 2	30
2.1 MARCO TEÓRICO Y CONCEPTUAL	30
2.1.1 IDENTIFICACIÓN DE LAS FUENTES DE DATOS	31
2.1.2 DEFINICIÓN DE VARIABLES	32
2.1.3 VALIDACIÓN DE DATOS	32
2.1.4 TRATAMIENTO DE DATOS ATÍPICOS Y FALTANTES	33
2.2 ANÁLISIS CLUSTER	35
2.2.1 MÉTODO DE AGRUPACIÓN.....	36
2.2.2 MÉTODO DE AGREGACIÓN.....	37
2.2.3 MÉTODO JERÁRQUICO	38
2.2.4 MÉTODO NO JERÁRQUICO	40
2.3 ANÁLISIS FACTORIAL	41
2.3.1 MODELO MATEMÁTICO DEL ANÁLISIS FACTORIAL	42
2.3.2 FASES DEL ANÁLISIS FACTORIAL.....	46
2.4 ANÁLISIS DISCRIMINANTE.....	50
2.5 HERRAMIENTA ESTADÍSTICA – PAQUETE SPSS	56
2.5.1 PARTES DEL PAQUETE ESTADÍSTICO SPSS.....	57
CAPITULO 3	59
3.1 PROPUESTA TEÓRICA – PRÁCTICA.....	59
3.2 APLICACIÓN DE LA METODOLOGÍA DE CLASIFICACIÓN DE LAS AGENCIAS DE UNA INSTITUCIÓN BANCARIA PRIVADA	65
3.2.1 UBICACIÓN GEOGRÁFICA DE LAS AGENCIAS	66
3.2.2 MODELO DE NEGOCIO ESTÁNDAR.....	66
3.2.3 INFRAESTRUCTURA DE LAS AGENCIAS.....	68
3.2.3.1 Tecnología.....	69
3.2.4 CARACTERIZACIÓN DE LOS CLIENTES	69
3.2.5 CARACTERIZACIÓN DE LAS TRANSACCIONES.....	70

CAPITULO 4	72
4.1 DIAGNÓSTICO DE LA SITUACIÓN DE LAS AGENCIAS	72
4.2 TIPIFICACIÓN DEL PROBLEMA.....	75
4.2.1 DEFINICIÓN DE VARIABLES.....	75
4.2.2 SELECCIÓN DE VARIABLES.....	76
4.2.2.1 Análisis de Correlaciones.....	84
4.2.2.2 Aplicación del Análisis de Componente Principales (ACP)	86
CAPITULO 5	90
CLASIFICACIÓN DE LAS AGENCIAS SEGÚN VARIABLES TRANSACCIONALES Y DE NEGOCIO.....	90
5.1 ANÁLISIS CLUSTER.....	90
5.2 CARACTERIZACIÓN DE LOS GRUPOS FORMADOS	92
5.3 DEFINICIÓN DE LOS MODELOS DE ATENCIÓN AL CLIENTE	95
5.4 PRONÓSTICO DEL COMPORTAMIENTO POR AGENCIA.....	97
5.4.1 ANÁLISIS DISCRIMINANTE	97
5.5 RESULTADOS.....	111
CAPÍTULO 6	117
6.1 CONCLUSIONES.....	117
6.2 RECOMENDACIONES	121
REFERENCIAS BIBLIOGRÁFICAS	122

ÍNDICE DE GRÁFICOS

Gráfico 1: Tasa de interés en el Ecuador.....	16
Gráfico 2: Distribución de la población en el Ecuador	17
Gráfico 3: Distribución de la pobreza en el Ecuador	20
Gráfico 4: Propuesta de la metodología.....	59
Gráfico 5: Organigrama estándar	67
Gráfico 6: Distribución de la variable Montos totales per cápita.....	78
Gráfico 7: Distribución de la variable eficiencia financiera	78
Gráfico 8: Distribución de la variable transacciones per cápita.....	79
Gráfico 9: Distribución de la variable eficiencia transaccional	79
Gráfico 10: Distribución de la variable transacciones no core	80
Gráfico 11: Distribución de la variable clientes Banca Privada y Alto Valor	80
Gráfico 12: Distribución de la variable clientes jurídicos.....	81
Gráfico 13: Distribución de la variable clientes resto	81
Gráfico 14: Distribución de la variable clientes per cápita	82
Gráfico 15: Distribución de la variable transacciones jurídicos.....	82
Gráfico 16: Distribución de la variable transacciones Banca Privada y Alto Valor.....	83
Gráfico 17: Distribución de la variable montos totales de clientes Alto Valor y Pymes.....	83
Gráfico 18: Nube de variables para dos factores	88
Gráfico 19: Nube de individuos para dos factores.....	89
Gráfico 20: Gráfico de Líneas.....	93
Gráfico 21: Distribución por cluster y grupos.....	99
Gráfico 22: Cluster 1	107
Gráfico 23: Cluster 2	107
Gráfico 24: Cluster 3	108
Gráfico 25: Funciones Discriminantes	108
Gráfico 26: Clasificación por cluster según comportamiento transaccional y de negocio	112

ÍNDICE DE TABLAS

Tabla 1: Tasa de desempleo en el Ecuador.....	19
Tabla 2: Cuadro de Amenazas y Oportunidades.....	22
Tabla 3: Cuadro de Fortalezas y debilidades.....	24
Tabla 4: Distribución del número de agencias por región.....	66
Tabla 5: Promedio de transacciones por segmento.....	72
Tabla 6: Estadísticas Descriptivas.....	77
Tabla 7: Matriz de correlaciones.....	85
Tabla 8: KMO y Bartlett.....	86
Tabla 9: Comunalidad de las variables para un factor.....	86
Tabla 10: Valores Propios del ACP con un factor.....	87
Tabla 11: Puntuaciones factoriales del ACP con un factor.....	87
Tabla 12: Centros Finales de Cluster.....	91
Tabla 13: Tabla ANOVA.....	92
Tabla 14: Número de casos en cada Cluster de K-means.....	92
Tabla 15: La matriz de medias.....	94
Tabla 16: Tabla ANOVA – test de linealidad.....	95
Tabla 17: Distribución de los grupos por cluster.....	98
Tabla 18: Resumen de casos.....	99
Tabla 19: Test de igualdad de medias.....	100
Tabla 20: Test M de Box.....	101
Tabla 21: Método Stepwise.....	101
Tabla 22: Variables incluidas en el análisis.....	102
Tabla 23: Análisis de la significatividad de las variables independientes del Modelo.....	102
Tabla 24: Tabla de comparación de grupos emparejados.....	103
Tabla 25: Eigenvalues o Valores Propios.....	104
Tabla 26: Wilks' Lambda.....	104
Tabla 27: Coeficientes Estandarizados Canónicos de las Funciones Discriminantes.....	105
Tabla 28: Matriz Estructural.....	105
Tabla 29: Coeficientes canónicos.....	105
Tabla 30: Funciones en los centroides de los grupos.....	106
Tabla 31: Probabilidades a priori.....	106
Tabla 32: Clasificación de las agencias.....	109
Tabla 33: Matriz de Confusión.....	110
Tabla 34: Matriz de confusión del grupo de test o prueba.....	111
Tabla 35: Distribución geográfica de las agencias por grupo.....	113
Tabla 36: Agencias por segmento y regional.....	114
Tabla 37: Regional por segmento.....	115
Tabla 38: Activos fijos per cápita en dólares por segmento.....	116

ÍNDICE DE ANEXOS

ANEXO 1: Tamaño de la Economía	123
ANEXO 2: Crecimiento económico.....	124
ANEXO 3: Distribución de hombres y mujeres en el Ecuador	125
ANEXO 4: Población Económicamente Activa.....	126
ANEXO 5: Ingreso Total personal según quintiles	127
ANEXO 6: Pobreza por Necesidades Básicas Insatisfechas.....	128
ANEXO 7: Distribución Bono Solidario	129
ANEXO 8: IMPORTAR DATOS DESDE EXCEL A SPSS	130
ANEXO 9: AGENCIAS ELIMINADAS SEGÙN CRITERIOS.....	133
ANEXO 10: PARTICIÓN DE LA BASE DE DATOS	135
ANEXO 11: ANÁLISIS DE COMPONENTES PRINCIPALES (ACP)	137
ANEXO 12: ANÁLISIS DE COMPONENTES PRINCIPALES PARA 4 Y 2 FACTORES	142
ANEXO 13: ANÁLISIS CLUSTER	148
ANEXO 14: ANÁLISIS DE CLUSTER DE K-MEDIAS con 3 cluster.	152
ANEXO 15: CARACTERIZACIÓN DE LOS GRUPOS	153
ANEXO 16: ANÁLISIS DISCRIMINANTE	155
ANEXO 17: ENCUESTA	164

CAPITULO I

1.1 INTRODUCCIÓN

Los bancos o instituciones financieras concentran los depósitos de la población y según las políticas propias y las regulaciones de los organismos, proceden a colocarlos o prestarlos a aquellos miembros de la comunidad que lo necesitan para invertir en sus negocios o para aumentar su patrimonio personal.

Para llevar a cabo este proceso, los bancos tradicionales por lo general, poseen la siguiente estructura:

- Canales de distribución para el proceso de ventas y de post venta: una red de Sucursales y Agencias para atender a la mayoría del público, bancas especializadas para aquellos segmentos que estratégicamente define como claves o potenciales,
- Un Call Center o Centro de Atención Telefónica y
- Una página de Internet para realizar transferencias, pagos, etc.
- Centros de Servicios como Recursos Humanos, Contraloría y Auditoría, Contabilidad, Seguridad, Tesorería, Mercadeo, etc.

Por otro lado, el actual contexto de la Banca Privada en Ecuador está sufriendo una nueva transformación caracterizada por un clima económico más restringido que impacta en sus utilidades y en sus clientes. En la banca privada se produce una disminución de sus ingresos y los clientes se ven beneficiados por tasas activas competitivas enfocando su interés ahora en servicios bancarios de calidad y variedad de productos.

Otro aspecto importante que se debe tomar en cuenta en la actividad bancaria es la interacción banco-cliente-entorno, donde la ubicación Geográfica de las instituciones bancarias y la actividad económica de los clientes genera factores exógenos que afectan el crecimiento de los bancos y el desarrollo de la población.

¹Los ejes en los cuales se analiza el entorno en el cual se desarrollan las agencias de una institución bancaria son: el crecimiento económico y tamaño de la economía en forma general para el Ecuador, población (densidad demográfica, clasificación por hombres y mujeres, migración según sexo), educación (analfabetismo y escolaridad), empleo (población económicamente activa y desempleo), pobreza por necesidades básicas insatisfechas, el índice de desarrollo humano, índice de pobreza humano, índice de vulnerabilidad social, ingreso real del hogar según quintiles.

Economía: El tamaño de la economía de Ecuador, se refiere a la magnitud de la actividad económica de un país, es decir, de su producción de bienes y servicios. Se mide a través del producto interno bruto o PIB. Puede expresarse en términos nominales o reales. El PIB per cápita es la relación entre el PIB y la población total del Ecuador.

En el ANEXO1, se observa que en 1997 el PIB per cápita fue de 1655 dólares, pero en el año 2000 alcanza un monto de 1079 dólares, lo que refleja la grave crisis que sufrió la economía del país, existía menos producción para un mismo tamaño de población, para el año 2001 existe cierta recuperación, llegando a 1396 dólares. Pero esto no refleja la realidad de un país, puesto que existen polos de pobreza extrema que son compensados por los pocos ecuatorianos con un ingreso muy alto.

En el ANEXO 2 se puede observar la tasa de crecimiento del PIB, ratificando lo anteriormente mencionado.

¹ Fuente: SIISE Sistema Integrado de Indicadores Sociales del Ecuador

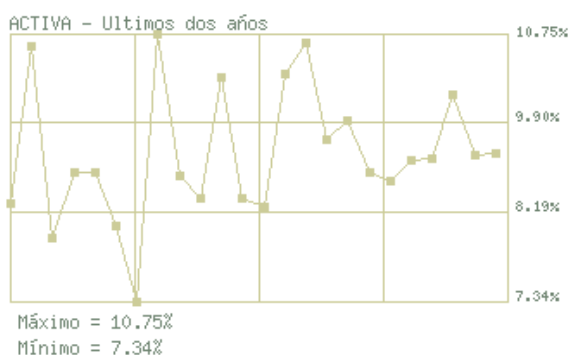
Las instituciones financieras enfrentan la realidad del país, a través del manejo de las tasas de interés, que en las mismas épocas mencionadas fueron altas. “La crisis financiera experimentada durante los años 1998 y 1999 fue una de las más agudas de la historia del Ecuador, con consecuencias económicas y sociales que trascienden la imagen de los indicadores del período, como por ejemplo la reducción del PIB en 7.9% en sures constantes y de alrededor del 30% en dólares en 1999, o el deterioro de los indicadores sociales de desempleo abierto y del porcentaje de la población bajo la línea de pobreza”.²

Tasa de Interés: En casi todos los bancos, el principal ingreso de la institución son los ingresos por intereses. En este rubro el secreto es controlar dos factores clave: el precio de la operación (tasas activas) y la calidad de la cartera de crédito.

La Banca Privada en el Ecuador ha experimentado una disminución drástica en sus ingresos por concepto de la tasa de interés activa, pero una tasa de interés pasiva sin altas variaciones, cuya evolución se muestra a continuación:

Gráfico 1: Tasa de interés en el Ecuador

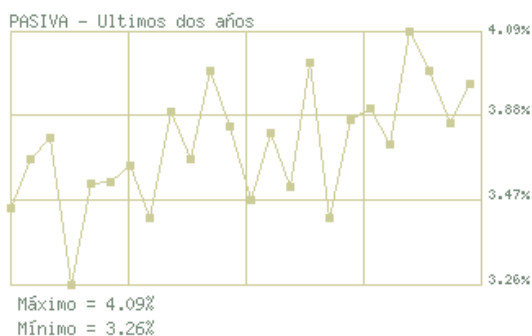
Tasa Activa



² Fuente: Banco Central del Ecuador: “UNA PROPUESTA DE PLAN ESTRATÉGICO DE DESARROLLO DE LARGO PLAZO PARA EL ECUADOR”

Tasa Pasiva

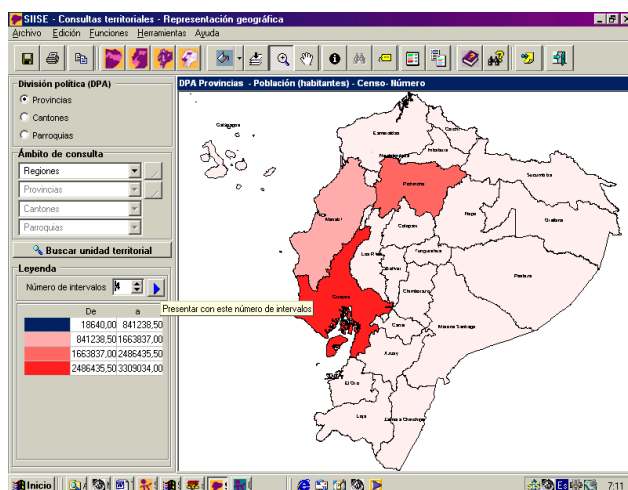
Fuente: Banco Central del Ecuador. Período: Feb/2005 hasta Jul/2005, datos semanales.



Para Julio, 2005 la tasa de interés activa es de 9.23% y la tasa pasiva de 3.52%, ratificando así la disminución de la rentabilidad que puede recibir la Banca Privada por este concepto. Una alternativa para mejorar su eficiencia y rentabilidad ha sido a través de la disminución de los costos internos y sobre todo mejorar su relación con los clientes, focalizando efectivamente las estrategias de mercadeo. Dentro de la focalización de las estrategias de mercadeo y la disminución de los costos está el total conocimiento de los puntos de contacto entre los clientes y la institución bancaria. Uno de estos puntos de contacto son las Agencias o Canales Físicos, en las cuales se genera todo tipo de información que responde a la interacción del cliente y el banco.

Población: La población en el Ecuador está distribuida de la siguiente forma:

Gráfico 2: Distribución de la población en el Ecuador



Como se puede apreciar el mayor número de habitantes está concentrado en las provincias de Pichincha y Guayas (de 2 millones a 3 millones de habitantes), seguida por Manabí y el resto de provincias. La distribución de hombres y mujeres en el Ecuador se puede observar en el ANEXO 3.

En las provincias de Pichincha y Guayas se concentra un alto número de agencias de la mayoría de Bancos en el Ecuador, correspondiendo a un porcentaje aproximadamente, del 60% de sus clientes.

Educación: En el Ecuador existe un 9% de analfabetismo respecto a las personas mayores de 12 años y de escolaridad máximo de 7 años, es decir que en promedio los ecuatorianos están en una institución educativa formal durante 7 años³, este índice fue elaborado en relación a la población mayor a los 24 años. Los bancos están inmersos en esta realidad e influye en la promoción de productos, siendo las estrategias más acertadas aquellas que requieren menos trámites y documentación y fácil acceso a créditos. Otro dato importante corresponde a la población que no se matricula cada año, y de ellas el 69% lo hace por razones económicas, lo que muestra la realidad de la educación en nuestro país.

Empleo: En el ANEXO 4 tenemos la distribución de la población económicamente activa en el Ecuador por provincia, como se observa la mayor concentración están en Pichincha y Guayas, y además en Quito y Guayaquil, ratificando la ubicación de un mayor número de agencias en estas ciudades.

En el Ecuador el 45% de la población económicamente activa, es decir “las personas en edad de trabajar (12 años y más), trabajaron al menos una hora durante el período de referencia de la medición (por lo general, la semana anterior); o aunque no trabajaron sí tenían trabajo; o, si no tenían trabajo, estaban en disponibilidad de trabajar”⁴.

Otra cifra importante que se debe analizar es la tasa de desempleo en el país, el porcentaje de desempleo es el número de personas de 12 años y más que están desocupadas, expresado como porcentaje del total de la población

³ Fuente SIISE: Educación

⁴ Fuente SIISE: Educación

económicamente activa (PEA) de ese grupo de edad en un determinado año, a nivel nacional se distribuye entre hombres y mujeres de la siguiente forma:

Tabla 1: Tasa de desempleo en el Ecuador

TASA DE DESEMPLEO

Fuente: Encuestas Urbanas de empleo y desempleo (EUED)-INEC

Año: 2001

Elaboración: SIISE

Región	Sexo	Porcentaje	desempleo	Población Económicamente Activa mayor a 12 años
Costa	Hombres	6.438	125,536	1,949,965
Costa	Mujeres	19.64	219,556	1,117,965
Sierra	Hombres	4.47	74,735	1,672,091
Sierra	Mujeres	8.2	118,518	1,446,118
Amazonía	Hombres	2.03	3,186	156,702
Amazonía	Mujeres	5.06	5,614	111,009
PAÍS		8.5	547,145	6,453,850

Pobreza: La distribución del ingreso por quintiles muestra la realidad de las familias ecuatorianas, (ver ANEXO 5), el 59% del ingreso en el Ecuador está en el quintil más alto, mientras que el 2.7% en el más bajo, ratificando la polarización y brecha entre la población con mayores recursos económicos y la población de bajos recursos económicos.

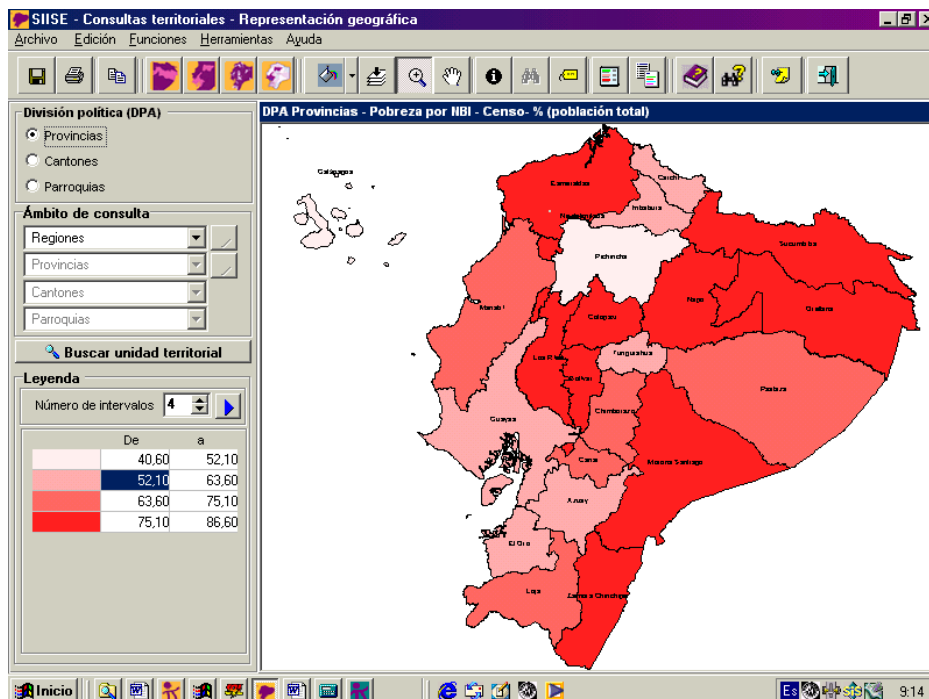
Las instituciones financieras tienen focalizados claramente sus mercados objetivos, tanto para los grupos económicos fuertes como la población que enfrenta la pobreza y la dificultad de acceder a la banca formal.

La pobreza por necesidades básicas insatisfechas en el Ecuador es del 61.3% medida en el 2001⁵. Ver ANEXO 6. Además, en el mapa se puede

⁵ Fuente SIISE

observar las provincias con mayor porcentaje de pobreza y se muestran en color rojo intenso.

Gráfico 3: Distribución de la pobreza en el Ecuador



El Bono Solidario se paga a través de las instituciones financieras con las cuales se han firmado convenios, como es el caso del Banco en estudio, en el ANEXO 7, se observa el número total de personas que lo reciben, en el Ecuador.

Los aspectos mencionados muestran las diferencias que existen entre provincias y ciudades de nuestro país; la influencia del entorno en el comportamiento transaccional y de negocio en una agencia bancaria.

Luego de recoger estos aspectos importantes en el ambiente que rodea una organización, se presenta un análisis complementario con un enfoque estratégico.

En la Fase I, PERSPECTIVA FILOSÓFICA, se menciona la misión, visión y los valores de la institución bancaria estudiada.

Misión: “Somos un equipo líder, que contribuimos al desarrollo del Ecuador, apoyando las necesidades financieras de las personas, de sus instituciones y de sus empresas”⁶.

Visión: “La institución financiera es el líder de su mercado en imagen, participación, productos y calidad de servicio, enfocando sus esfuerzos hacia el cliente, anticipándose a sus necesidades, desarrollando a su personal, otorgando rentabilidad sostenible a los accionistas”⁷.

Entre los valores y principios se menciona: “disciplina, orientación al cliente, austeridad, trabajo en equipo, autenticidad, desarrollo profesional y sensibilidad social.”

En la Fase II se realiza un estudio de la situación actual de la institución bancaria privada a través de tres vertientes:

- Un macroentorno
- Un ambiente interno y
- Un microambiente o microentorno

El análisis de cada vertiente permite elaborar un FODA en función de factores críticos de éxito, que son parte de la planificación estratégica y permiten determinar la empatía del sistema e identificar posibles mercados iniciales.

En el **Macroentorno** existen factores críticos definidos como amenazas y oportunidades que determinan el tipo de planificación estratégica a aplicar. A continuación se mencionan algunos factores que influyen en la institución, caso de estudio⁸.

⁶ Fuente: Página interna de la institución bancaria

⁷ Fuente: Página interna de la institución bancaria

⁸ Fuente: “Estrategias de Marketing”, del autor Díaz Santos.

Tabla 2: Cuadro de Amenazas y Oportunidades

AMENAZAS					
	Factores Críticos de Éxito	Probabilidad de ocurrencia	Efecto en el Banco		Impacto/ Pesos finales
1	Grupos económicos que atacan con críticas el funcionamiento	0.8	0.136	5	0.678
2	Tasas de interés activas hacia la baja y restricción en el cobro	0.6	0.102	4	0.407
3	Disminución de la capacidad de ahorro de los ecuatorianos y de	0.8	0.136	4	0.542
4	Inestabilidad Política	0.8	0.136	2	0.271
5	Marco jurídico politizado en el país por partidos políticos tradici	0.4	0.068	2	0.136
				TOTAL	2.034
OPORTUNIDADES					
	Factores Críticos de Éxito	Probabilidad de ocurrencia	Efecto en el Banco		Impacto/ Pesos finales
1	El banco pertenece al grupo financiero más grande del país	0.7	0.119	8	0.949
2	Alto índice de migración en el país	0.8	0.136	5	0.678
3	Avances tecnológicos en el corto plazo	0.4	0.068	4	0.271
4	Variedad de nichos de mercado	0.6	0.102	5	0.508
				TOTAL	2.407

Para calcular el Efecto en el Banco se divide cada probabilidad de ocurrencia para la sumatoria de las probabilidades de ocurrencia de las amenazas y de las oportunidades.

Una vez identificadas las oportunidades y amenazas que influyen en la institución financiera seleccionada, el siguiente paso es determinar el tipo de ambiente competitivo en el cual se encuentra, de la siguiente forma: se suman los puntajes de impacto de las amenazas y del mismo modo de las oportunidades, se realiza una comparación entre las dos sumatorias y se determina si es un ambiente competitivo ideal, especulativo, maduro o problemático. Por lo tanto, se obtiene:

$$\sum \text{Amenazas} = 2.034$$

$$\sum \text{Oportunidades} = 2.407$$

Como se observa, la sumatoria de los pesos de las oportunidades es superior a la sumatoria de las amenazas, entonces la estructura que debe tener la planificación estratégica debe ser progresista, que defina los lineamientos para el cambio, considerando un ambiente competitivo problemático y un medio ambiente turbulento, puesto que las decisiones políticas del país afectan a todas las

instituciones. El sector financiero actualmente es muy competitivo y el banco debe aprovechar su actual liderazgo.

En el **Microentorno** se identifica el resto de actores que existen en el esquema de fuerzas competitivas.

El poder de negociación está en la institución financiera seleccionada, el escoge sus proveedores y direcciona sus productos a varios consumidores. En cuanto a sus competidores, el banco tiene el mayor número de clientes y mayor cobertura a nivel nacional, pero presenta la misma variedad de productos y canales. Además, no existe entre las instituciones financieras diferencias importantes en la calidad de la atención al cliente.

En el ciclo de vida de una institución financiera se puede determinar la etapa de desarrollo de la empresa. En el análisis del microambiente, el Banco se encuentra en una fase II⁹, donde la variación de sus ventas es siempre positiva y acelerada, el número de productos va creciendo, la tendencia en precio es a la baja, en cuanto a las tasas de interés activas, pero el precio por servicios aumenta, el gasto en promoción y producción es alto y tiene amplitud y variedad de canales de comercialización. En conclusión, las estrategias del banco deben estar enfocadas a:

- Posicionar y reutilizar sus productos
- A disminuir sus costos y
- A mejorar la calidad del servicio al cliente en cada uno de sus canales.

En el **Análisis Interno** se define la estructura de fortalezas y debilidades, y se muestra en el siguiente cuadro:

⁹ El ciclo de vida de una institución tiene 4 partes, y depende de la evolución de algunos factores como: Venta, número de productos, precio, gasto en promoción y producción y amplitud en el canal de comercialización. Fuente: "Estrategias de Marketing", autor: Díaz Santos.

FORTALEZAS

	Factores Críticos de Éxito	Probabilidad de ocurrencia	Efecto en el Banco		Impacto/ Pesos finales
1	El banco tiene el mayor porcentaje de participación en el banca ecuatoriana, con un millón de clientes activos aproximadamente.	0.8	0.0708	5	0.354
2	Tiene una segmentación de clientes eficiente	0.6	0.0531	4	0.212
3	El banco está posicionado como una institución confiable y es conocido a nivel nacional	0.7	0.06195	5	0.310
4	Estructura financiera sólida y con liquidez	0.8	0.0708	5	0.354
5	A pesar de las crisis económicas surgidas en el país, el banco sigue adelante.	0.8	0.0708	5	0.354
6	La tecnología aplicada es de alto nivel y moderna	0.9	0.07965	4	0.319
7	El nivel de estudios promedio es secundario, pero con cierto porcentaje de universitarios	0.6	0.0531	2	0.106
8	Estructura organizacional cambiante	0.8	0.0708	5	0.354
				TOTAL	2.363

DEBILIDADES

	Factores Críticos de Éxito	Probabilidad de ocurrencia	Efecto en el Banco		Impacto/ Pesos finales
1	Existen varios competidores en el mercado financiero	0.8	0.0708	4	0.283
2	Sus productos no se diferencian de la competencia	0.8	0.0708	4	0.283
3	Tiene alta productividad y una eficiencia media.	0.6	0.0531	3	0.159
4	La rotación de empleados es baja, excepto en los cambios de estructura	0.8	0.0708	4	0.283
5	El desarrollo de competencias depende de cada persona	0.6	0.0531	2	0.106
6	Monitoreo de la calidad del liderazgo	0.8	0.0708	5	0.354
7	La cultura organizacional influye en la atención del cliente interno y externo	0.9	0.07965	5	0.398
				TOTAL	1.867

Tabla 3: Cuadro de Fortalezas y debilidades

Las variables cualitativas, es decir las debilidades y fortalezas, son transformadas en variables cuantitativas. El resultado, de las dos tablas muestra que la sumatoria de los pesos de las Fortalezas es mayor a la sumatoria de los pesos de las debilidades.

De acuerdo a los dos resultados anteriores, se obtiene que las oportunidades tienen mayor peso que las amenazas y las fortalezas que las debilidades, en conclusión la empatía o forma de adaptación que debe tener la organización con el medio es Agresiva, es decir, sus estrategias deben afectar al mercado y a sus competidores.

En la fase III, están los objetivos estratégicos y las maniobras estratégicas o proyectos.

Dentro de las instituciones financieras las estrategias y maniobras dependen de cada área y de los objetivos estratégicos de la organización. A continuación se mencionan algunas áreas:

- Un área de negocio encargada de las estrategias para cada familia de productos, y del cumplimiento presupuestario de las agencias.
- Un área de canales, con canales virtuales y canales de autoservicio (cajeros)
- Centros de servicios, formados por el resto de áreas que aportan a la cadena de valor de los productos: Administración y logística, Recursos Humanos, Seguridad, etc.
- Unidades de control como riesgo, auditoría, etc.

El objetivo de la presente tesis no es describir los planes estratégicos de cada área, pero se pondrá especial énfasis al área de canales.

Los canales son: canales virtuales, agencias, teléfono, cajeros automáticos, etc. Dentro de las estrategias de gestión de los canales se debe considerar: la disminución de costos, el mejoramiento de la calidad de servicio al cliente en cada canal y aplicar tecnologías acorde a los avances tecnológicos.

Para cada una de las estrategias mencionadas se deben elaborar maniobras o proyectos. La estrategia de mejoramiento de la calidad del servicio al cliente en cada canal estudia la estructura física o virtual del canal, el nivel de educación del personal y su capacitación y de las herramientas tecnológicas.

El canal del Banco más utilizado por los clientes es el canal físico o agencias, siendo este canal el más costoso. Entre los proyectos que se proponen para mejorar la atención del cliente en el canal agencias están:

- Disminución del tamaño de las colas, a través de la optimización del número de cajeros y disminución del tiempo de espera del cliente en la cola.
- Capacitación del personal y mejorar sus competencias en atención al cliente, para ofrecer un servicio de calidad.
- Mejorar los tiempos de respuesta de los sistemas tecnológicos para aumentar la tasa de efectividad en las cajas.
- Diferenciación de las agencias, entre transaccionales, negocio y mix, para disminuir costos.

Cada uno de estos proyectos podría ser un tema de tesis, y en este caso el proyecto a ser cubierto es la “diferenciación de las agencias”, puesto que la fase del ciclo de vida en la cual se encuentra el banco propone la disminución de costos.

1.2 JUSTIFICACIÓN

La clasificación de las agencias responde al interés del negocio de aprender y responder activamente a las necesidades del cliente. Los negocios que entienden lo que sus clientes quieren, cómo y cuándo lo quieren pueden tomar decisiones informadas para guiar las estrategias de sus negocios, crear conciencia y atraer y retener a sus clientes.

En el banco existen iniciativas implementadas con diversos objetivos respecto a sus agencias, lo que no permite obtener resultados concretos y cumplir sus metas ya sean económica o de eficiencia.

Esta tesis permitirá enfocar claramente las prioridades generadas en cada agencia; ayuda a reducir el desgaste de sus clientes más valiosos, ya sea por un mal servicio o por estructuras de negocio complicadas que producen pérdida de tiempo al cliente en la agencia.

La finalidad es incrementar la rentabilidad del banco, a través de la diferenciación de las agencias que converge a la disminución de los costos y a ofrecer un servicio bancario que se ajuste a las necesidades del cliente.

La metodología de clasificación propuesta es una iniciativa técnica para el análisis de los datos de la interacción del cliente con las agencias. En su mayoría las variables generadas en las agencias son cuantitativas (número de clientes, número de transacciones, eficiencia transaccional, capacidad instalada, etc.), pero pueden ser complementadas por variables cualitativas (tipo de clientes, tipo de transacciones, variables demográficas y geográficas). Para incluir variables cualitativas se debe aplicar otra metodología.

Las variables cuantitativas que recogen información de las agencias y permite diferenciarlas unas de otras tienen como directrices principales a:

- Los montos en dólares que generan los clientes en cada agencia
- Los clientes segmentados según criterios internos del banco
- El número de transacciones que genera cada agencia y clasificadas según segmentos internos del banco
- Índices de eficiencia financiera y transaccional del recurso humano de cada agencia.

En el desarrollo de la tesis se explican las variables que fueron utilizadas específicamente.

La metodología de clasificación se basa en un conjunto de métodos estadísticos que analizan de forma simultánea dos o más variables observadas, es decir, permiten tener una visión global de los fenómenos y estudiar las posibles interacciones que puedan existir entre los diversos factores.

El análisis multivariante permite analizar simultáneamente toda la información relativa al comportamiento de los clientes en las agencias, existe una mínima pérdida de información y ofrece una representación gráfica de la

interacción entre variables. Los métodos de análisis multivariante serán explicados en el Capítulo 2, en el Marco Teórico.

Una vez que se identifican grupos de agencias excluyentes con los métodos mencionados, es decir una agencia pertenece únicamente a un grupo, se caracterizan los grupos por las variables seleccionadas y se establecen nuevas estructuras de negocio individualizadas para cada grupo.

Esta tesis es una aportación técnico-práctica a corto plazo que responde a la implementación de una solución efectiva que será útil a largo plazo, el proceso analítico de clasificación que será aplicado permite al gerente de negocios y gerente de la red de oficinas implementar campañas de negocio focalizadas por tipo de agencia, cambiar la infraestructura de las agencias, permitiendo una disminución de los gastos por eficiencia y mejorar la relación del banco con sus clientes.

Esta metodología de clasificación tiene algunos usos, como por ejemplo:

- Manejar una compleja y extensa base de datos del comportamiento de agencias.
- Un punto de partida en el cual los objetivos y estrategias de mercadeo pueden ser más potencializados en cada agencia.
- Ofrece medidas que permiten ser más efectivo en el mercadeo de productos en las agencias.
- Identificar grupos de comportamientos para focalizar estructuras de negocio según las características dominantes en cada agencia.
- A continuación se mencionan los objetivos de esta tesis.

1.3 OBJETIVOS

- Elaborar una metodología de clasificación por servicios a agencias bancarias para mejorar el servicio al cliente.

1.3.1 OBJETIVOS ESPECÍFICOS

- Analizar las variables transaccionales y de negocio de cada agencia que expliquen la interacción entre el cliente y la agencia.
- Clasificar las agencias aplicando técnicas estadísticas que identifiquen claramente los grupos formados.
- Conceptuar cada grupo de agencias según su correlación en mayor o menor medida con las variables estudiadas.
- Establecer nuevas estructuras funcionales de negocio de acuerdo a los grupos formados por las agencias

En el siguiente capítulo está el sustento técnico de la propuesta metodológica para la clasificación de agencias bancarias y la definición de conceptos relacionados con el análisis multivariante y el tratamiento de bases de datos.

CAPITULO 2

2.1 MARCO TEÓRICO Y CONCEPTUAL

La clasificación de las agencias depende del número y tipo de variables seleccionadas. Existen diferentes técnicas de clasificación, para clasificación supervisada están: los árboles de decisión, las redes neuronales, el análisis discriminante, etc., y para clasificación no supervisada están los clusters.

El problema de dimensionalidad no permite que se pueda clasificar en forma empírica o manual un conjunto de variables y es necesario aplicar técnicas y herramientas estadísticas avanzadas; las técnicas aplicadas corresponden al Análisis de datos Multivariante. Además, para la clasificación de las agencias se aplica un análisis cluster y para la selección de variables y pronóstico de la probabilidad de pertenecer a un grupo se aplica el análisis discriminante.

La información aplicada por agencia existe y está almacenada en un DataWareHouse.

Este capítulo contiene material técnico y estadístico que será útil para comprender la metodología de clasificación que se propone en esta tesis. Varios conceptos son revisados y tienen el siguiente orden:

- Identificación de las fuentes de datos
- Definición de variables
- Validación de los datos
- Tratamiento de los datos atípicos y faltantes
- Partición de Datos
- Métodos de Análisis Multivariante.

2.1.1 IDENTIFICACIÓN DE LAS FUENTES DE DATOS

Para el caso de estudio los datos corresponden a información histórica generada internamente en el banco; sino se cuenta con ésta, el técnico debe crear la base de datos con otras técnicas. Los datos son obtenidos del DATAWAREHOUSE (DWH) que es un almacén de información distribuido en diferentes cubos. El DWH necesita de tres áreas bien diferenciadas entre sí pero que se integran para lograr un buen funcionamiento:

- Los sistemas origen o fuentes legados .
- El área de transformación de los datos (en lo adelante nos referiremos como ATD)
- El **Servidor** de presentación del DWH

Las fuentes legados: son los sistemas de procesamiento que capturan las transacciones del negocio. Deben ser sistemas confiables y consistentes, aunque entre ellos hay marcadas diferencias en los formatos y las **estructuras** de los datos; quedan fuera del DWH por lo que no se tiene el **control** sobre su contenido.

El área de transformación de los datos (ATD) consta tanto del área de **almacenamiento** como del conjunto de **procesos** que se usan frecuentemente para la extracción, transformación y carga. Es generalmente la parte más compleja de esta **arquitectura**. Es todo lo que se presenta entre la fuente externa y el área de presentación de los datos.

El servidor de presentación es la fuente de consulta del banco, donde se organizan los datos del DWH, se almacenan y están disponibles para hacer consultas, reportes y acceder mediante aplicaciones.

Concretamente, el área de transformación de datos puede estar constituida por un **grupo** de **servicios** que pueden ser implementados para mantener actualizada la información del **almacén**. Las bases de datos de casos anteriores

deben ser medidos en las mismas condiciones que los que se quiere predecir y no debe tener ningún sesgo relevante. En esta parte se realiza una primera selección de variables y unificación de los datos en una sola base de datos.

2.1.2 DEFINICIÓN DE VARIABLES

Una variable es un fenómeno medible que varía a través del tiempo, o que difiere de un lugar a otro o de un individuo a otro. Las variables pueden ser: cualitativas – en escala nominal, valores no numéricos con ausencia de orden entre ellos- y cuantitativa. Las variables cuantitativas pueden ser: en **escala ordinal**, valores no numéricos con presencia de orden entre ellos; en **escala de intervalo o razón**, que a su vez se clasifica en: cuantitativa discreta: únicamente puede tomar una cantidad finita o numerable de valores numéricos; cuantitativa continua, puede tomar cualquier valor numérico en un intervalo; en escala de intervalo, únicamente tiene sentido la diferencia entre sus valores; en escala de razón, sí, además de la diferencia, tiene sentido la razón entre sus valores.

Para el caso de estudio las variables estudiadas son cuantitativas continuas. No es factible trabajar con tantas variables por: tiempos de procesamiento, información redundante, problemas de convergencia en la aplicación de las técnicas, complejidad para interpretar los resultados, etc. Estos métodos requieren que el número de variables cuantitativas sea máximo una variable por cada diez casos, es decir, si se requiere realizar el análisis con 5 variables el número de casos mínimo que debe contener la base de datos es 50 casos.

2.1.3 VALIDACIÓN DE DATOS

En cada variable los casos deben tomar un único valor que debe ser numérico o simbólico, se debe minimizar la redundancia en las variables, es decir que varias variables expliquen lo mismo, y evitar valores confusos, es decir que para un mismo factor en una variable reciba varias codificaciones

2.1.4 TRATAMIENTO DE DATOS ATÍPICOS Y FALTANTES

Para cada una de las variables seleccionadas se debe realizar un análisis estadístico descriptivo (rango, mínimo, máximo, diagrama de cajas, media, moda, sesgo, curtosis, etc.).

Un dato faltante o missing puede ser: un campo no rellenado por olvido, campo no rellenado por imposibilidad o valor especial, en estos casos se debe: borrar el patrón, borrar la columna, sustituir missing por un valor numérico especial pero cuidando de no romper la topología, sustituir missing por valor promedio de los valores cercanos o tratar de predecirlo.

Un dato atípico es una observación con características diferentes a las demás, que pueden ser válidas o no, además un pequeño número de casos atípicos puede alterar totalmente la distribución de una variable y sus estadísticos. La forma de detectarlos es gráficamente o mediante detección analítica considerando: para número de datos menor a 80, el valor absoluto de la variable está a 2.5 desviaciones atípicas, y para un número de casos superior o igual a los 80, el valor está a 3 desviaciones atípicas. Es importante conocer la naturaleza de un valor atípico y optar por una acción concreta.

2.1.5 PARTICIÓN DE DATOS

De un mismo conjunto de datos pueden salir dos conjuntos: de entrenamiento y de validación, o se pueden formar 3 grupos, un conjunto de entrenamiento, uno de validación y otro llamado test, todo depende si el conjunto test o de prueba puede o no salir del conjunto de datos iniciales, depende de la estrategia de aplicación del modelo. El conjunto de entrenamiento corresponde al 60% de los datos, el conjunto de validación al 20% de los datos y 20% de los datos es el conjunto test, estos porcentajes pueden cambiar según los objetivos de la investigación. El error del modelo aplicado corresponderá a los errores en las corridas del modelo para el conjunto de entrenamiento y de validación.

2.1.6 ANÁLISIS MULTIVARIANTE

Se puede distinguir dos grandes bloques de métodos o técnicas multivariantes: métodos descriptivos y métodos explicativos.

Métodos descriptivos: (o de interdependencia): estos métodos no distinguen entre variables dependientes o a explicar y variables independientes o explicativas. Esto quiere decir que todas las variables están relacionadas unas con otras, es decir, están interrelacionadas. Dentro de los métodos descriptivos se pueden distinguir los siguientes grupos de técnicas:

Métodos de clasificación: los cuales permiten agrupar individuos, empresas, etc., a partir de un grupo de variables de partida. En concreto, en este tema, se verá el análisis cluster, además existen los métodos como: Escalas multidimensionales, análisis de correspondencias y modelos log-lineales.

Otro bloque de métodos descriptivos son los factoriales: son técnicas que permiten resumir información, y dentro de los cuales se estudiará el análisis de componentes principales.

Métodos explicativos (o de dependencia): son aquellos que explican unas variables en función de otras, es decir, hay variables dependientes o a explicar y variables independientes o explicativas. En este grupo está: la regresión lineal, análisis de supervivencia, análisis de la varianza, correlación canónica como por ejemplo: análisis discriminante, análisis de regresión logística y análisis conjunto. Como se menciona, dentro de este grupo está el análisis discriminante, que además permite obtener la predicción de casos nuevos a través de las funciones discriminantes.

2.2 ANÁLISIS CLUSTER

El análisis de cluster es una técnica de exploración de datos observados con el fin de formar segmentos de mercado o grupos naturales y relativamente homogéneos, basados en su proximidad. Un cluster o conglomerado debe ser lo más homogéneo posible en sí y lo más heterogéneo posible con los demás. Los datos observados estarán en un mismo cluster o en otro, dependiendo de la homogeneidad entre ellos. Debe haber una justificación teórica y una lógica que fundamentara y que guiara el análisis de agrupamiento. Teniendo en cuenta que es difícil evaluar la calidad de la agrupación, dado que no hay pruebas estadísticas para asegurar que el grupo obtenido no represente una aleatoriedad pura, es útil el análisis del analista sobre la medida del criterio y lo razonable de los resultados. Es importante preguntarse si tienen sentido los clusters formados. No hay que creer que este método provee una solución perfecta, completa, única y definitiva, aunque sí ampliamente confiable. Debido a que es un método exploratorio, los resultados deben considerarse provisionales hasta que sean confirmados mediante otra muestra independiente.

La información puede estar expresada en forma de frecuencia o en términos binarios o ser de tipo continua. El escalamiento de las variables es un aspecto importante, ya que las diferencias en el mismo pueden afectar a las soluciones de conglomeración. Si las variables se caracterizan de diferentes maneras (por ejemplo, una variable se mide en dólares y la otra, en años), se recomienda la previa estandarización de las mismas, para que una no domine a las otras. Otro aspecto a tener en cuenta es evitar el uso de variables nominales (ej. 1 para casado, 2 para soltero y 3 para otros). Estas deberían convertirse en variables dummy o dicotómicas (ej. 0 para casado y 1 para otros). También debe evitarse la existencia de multicolinealidad¹⁰ que se da cuando dos variables están

¹⁰Existencia de una relación lineal entre las variables independientes de un modelo de regresión.

altamente correlacionadas¹¹ porque una depende o es combinación de la otra. Deben incluirse todas las variables relevantes en el análisis. Si se omiten variables de interés, la solución obtenida puede ser equívoca.

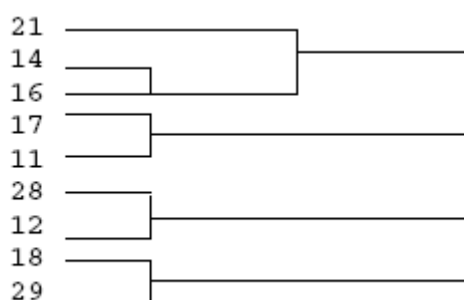
Durante el proceso, el analista tendrá que tomar dos decisiones: cuál será el método de agrupación y cuál, el de agregación. Con la primera, se observa de qué manera los individuos se unen en un mismo grupo. Con la segunda, se determina las semejanzas o diferencias entre clusters.

2.2.1 MÉTODO DE AGRUPACIÓN

El investigador debe decidirse sobre un criterio de distancia, para saber lo próximos o parecidos entre sí que están los distintos datos (o individuos) a analizar. Esto es pues, la elección de cómo se calcularán las distancias entre cada observación. Una vez calculadas las distancias, se procede a juntar los datos más parecidos según un método de agrupación, llegando a grupos homogéneos en sí y diferentes grupo a grupo.

Gráficamente, resulta:

Caso o individuo



Este es el resultado gráfico que muestra el proceso de agrupación de los casos. Vemos, por ejemplo, cómo el caso 14 se asemeja al 16 y ambos al 21.

¹¹ Relacionadas entre sí.

Según sea el tipo de dato (de frecuencia, binario o continuo), existe una gran variedad de coeficientes que permiten obtener la distancia y/o similitud entre los datos observados. La explicación de cada criterio sobrepasa los límites de esta tesis.

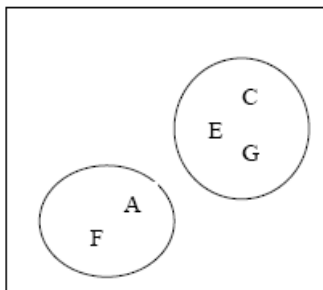
Cabe aclarar que no hay un cálculo de la distancia que sea el mejor para todas las situaciones posibles. Uno puede ser preferible a otro según las condiciones del análisis y del criterio del analista. En general, el promedio de todas las distancias euclídeas al cuadrado es el más usado.

Es importante que quede claro que cualquiera sea la distancia que se use, se debe considerar la calidad de la agrupación.

2.2.2 MÉTODO DE AGREGACIÓN

Consiste en la decisión de cómo se van a separar los distintos clusters, a través de la elección del estadístico a usar sobre el cual se mide la distancia entre los grupos. Nótese la diferencia con la agrupación, donde se determina el tipo de distancia. Mientras que la agregación define sobre qué casos de cada grupo se va a aplicar esta distancia y distingue los grupos entre sí.

Gráficamente, resulta:



Vemos como los individuos A y F por su cercanía forman un cluster y como los individuos C, E y G forman otro conglomerado. Existen dos enfoques

alternativos: jerárquico y no jerárquico. Ambos permiten llegar a distintas soluciones a partir del mismo grupo de observaciones iniciales.

Así es como el tipo de método de agregación a usar influye más que la decisión sobre cómo agrupar los datos, en el resultado final del análisis. Se puede realizar los dos tests (jerárquico y no jerárquico) y luego comparar los resultados usando una tabla de doble entrada.

2.2.3 MÉTODO JERÁRQUICO

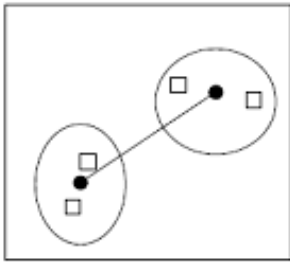
La agrupación jerárquica empieza con todos los objetos dentro de un grupo y los divide hasta que todos los objetos estén en su propio conjunto. Cada uno de los objetos es asociado con otro dentro de un agrupamiento. Esto hace que requiera de una gran cantidad de cálculos, lo que en ocasiones limita su posibilidad de aplicación en muestras muy grandes.

La primera combinación o separación de objetos restringirá el resto del análisis, a raíz de que cuando dos datos se juntan no se separan más. Esto hace que el agrupamiento resulte ser relativamente inestable y poco confiable. Por todo lo mencionado anteriormente, no garantiza una solución óptima.

Dentro de este enfoque, existen distintas maneras de evaluar la distancia entre dos clusters. Se puede elegir entre:

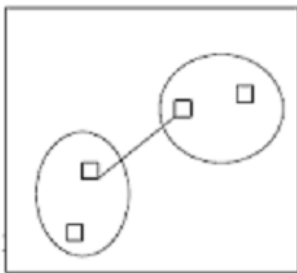
- La distancia entre las medias de cada grupo (centroid method)

Gráficamente,



- La distancia entre las medianas de cada grupo
- La distancia más corta entre un par de puntos, uno de cada grupo (nearest-neighbor)

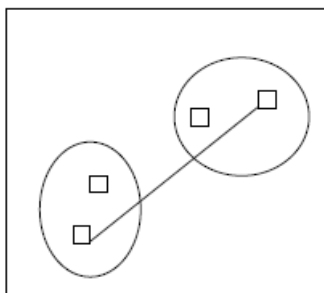
Gráficamente,



Distancias máximas

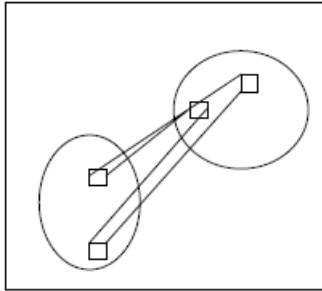
- La distancia más larga entre un par de puntos, uno de cada grupo (furthest-neighbor)

Gráficamente,



- El promedio de todos los posibles pares de puntos (average linkage between groups)

Gráficamente,



- La suma de las distancias al cuadrado (método de Ward).

Por ejemplo, para formar los conglomerados se puede medir la Distancia Euclídea (método de agrupación) entre las medias (método de agregación). Se denomina Distancia Euclídea a la línea recta formada por dos puntos. Según el criterio que elija, pueden obtenerse diferentes resultados. Para seleccionar la mejor solución, es importante el conocimiento que se tenga de sus clientes y de los productos y/o servicios que quiera ofrecer.

2.2.4 MÉTODO NO JERÁRQUICO

Un agrupamiento no jerárquico difiere tan solo en que permitirá a los objetos, que dejen un grupo, que se unan a otro, a medida que se forman los conglomerados o clusters. De este modo el criterio de agrupamiento se va mejorando. Los objetos se mueven de grupo a grupo siguiendo un proceso de optimización. El proceso de asignación de las observaciones a los grupos es iterativo y no permanente entre las distintas iteraciones. Cada observación se asigna a distintos clusters hasta llegar a una solución óptima (lo que no se logra con los métodos jerárquicos).

El método más usado es el denominado K - medias. Como cada grupo queda representado por su media, no requiere de gran cantidad de cálculos, por

lo cual es posible aplicarlo en bases grandes. Se utilizan especialmente variables continuas. Tiende a ser más confiable, las muestras divididas tenderán a verse más similares que en el método jerárquico.

Un problema central es determinar el número apropiado de conglomerados. Con este método, se tiene la posibilidad de especificar la cantidad de grupos que desee formar, porque lo conoce o porque se deriva de un uso planeado de los mismos. En caso contrario, puede partir del patrón de grupos generados por el programa. Pero si ha seleccionado un número inapropiado de conglomerados o ha omitido variables relevantes, los resultados podrían ser equívocos. Del mismo modo, puede definir el perfil de los mismos (si conoce de antemano esta información) y cada caso es comparado con dicho perfil.

Dicho de otro modo, se elige la cantidad de clusters a formar. Puede especificar las medias iniciales de los conglomerados, si conoce de antemano dicha información. Si la desconoce, el centro de cada conglomerado se puede formar mediante la iteración de los casos. Durante el proceso los datos pueden cambiar de un grupo a otro (nótese la diferencia con el enfoque jerárquico). Se realizan tantas iteraciones (según la cantidad deseada) hasta que cada cluster sea lo más homogéneo posible y ningún punto cambie de lugar.

2.3 ANÁLISIS FACTORIAL

El análisis factorial (AF) es una técnica de análisis multivariante que se utiliza para el estudio e interpretación de las correlaciones entre un grupo de variables. Parte de la idea de que dichas correlaciones no son aleatorias sino que se deben a la existencia de factores comunes entre ellas. El objetivo del AF es la identificación y cuantificación de dichos factores comunes.

El análisis factorial es una técnica para analizar las asociaciones lineales entre las variables. Si las variables no estuvieran asociadas linealmente, las

correlaciones entre ellas serían nulas y, en consecuencia, la matriz de correlaciones sería igual a la matriz identidad. Recíprocamente, si las correlaciones entre variables fueran nulas, no existirían asociaciones lineales entre las variables y, por consiguiente, carecería de sentido realizar un análisis factorial.

Por ejemplo, hay fenómenos como estilo de vida, imagen de un producto, actitudes de compra, nivel socioeconómico, que es necesario conocer pero que no se pueden medir con una sola pregunta, porque se trata de fenómenos complejos que se manifiestan en infinidad de situaciones, sentimientos, comportamientos y opiniones concretas. Estos fenómenos son el resultado de la medición de un conjunto de características. El AF nos permitirá combinar preguntas de manera que podamos obtener nuevas variables o factores que no son directamente medibles pero que tienen un significado.

Se trata de una técnica adecuada para el caso de variables continuas altamente correlacionadas.

2.3.1 MODELO MATEMÁTICO DEL ANÁLISIS FACTORIAL

El modelo matemático del AF supone que cada una de las p variables observadas es función de un número m factores comunes ($m < p$) más un factor específico o único. Tanto los factores comunes como los específicos no son observables y su determinación e interpretación es el resultado del AF.

Analíticamente, supondremos un total de p variables observables tipificadas y la existencia de m factores comunes. El modelo se define de la siguiente forma:

$$\begin{aligned} X_1 &= l_{11}f_1 + l_{12}f_2 + l_{1m}f_m + e_1 \\ X_2 &= l_{21}f_1 + l_{22}f_2 + l_{2m}f_m + e_2 \\ &\dots \\ X_p &= l_{p1}f_1 + l_{p2}f_2 + l_{pm}f_m + e_p \end{aligned}$$

que podemos expresar de forma matricial como: $X = LF + E$ donde:

X: es el vector de las variables originales.

L: es la matriz factorial. Recoge las cargas factoriales ó (saturaciones).

l_{jh} : es la correlación entre la variable j y el factor h.

F: es el vector de factores comunes.

E: es el vector de factores únicos.

Como tanto los factores comunes como los específicos son variables hipotéticas, supondremos, para simplificar el problema, que:

- Los factores comunes son variables con media cero y varianza 1. Además se suponen incorrelacionados entre sí.
- Los factores únicos son variables con media cero. Sus varianzas pueden ser distintas. Se supone que están incorrelacionados entre sí. De lo contrario la información contenida en ellos estaría en los factores comunes.
- Los factores comunes y los factores únicos están incorrelacionados entre si Esta hipótesis nos permite realizar inferencias que permitan distinguir entre los factores comunes y los específicos.

Basándonos en el modelo y en las hipótesis formuladas, podemos demostrar que la varianza (información contenida en una variable) de cada variable se puede descomponer en:

- Aquella parte de la variabilidad que viene explicada por una serie de factores comunes con el resto de variables que llamaremos comunalidad de la variable.
- La parte de la variabilidad que es propia a cada variable y que, por tanto, es no común con el resto de variables. A esta parte se le llama factor único o especificidad de la variable.

$$\text{Var}(x_j) = 1 = l_{j1}^2 \text{Var}(f_1) + l_{j2}^2 \text{Var}(f_2) + \dots + l_{jm}^2 \text{Var}(f_m) + \text{Var}(e_j) = l_{j1}^2 + l_{j2}^2 + l_{jm}^2 + \text{Var}(e_j)$$

donde:

l_{jh}^2 = representa la proporción de varianza total de la variable X_j explicada por el factor h .

$h_j^2 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jm}^2$: es la comunalidad de la variable X_j y representa la proporción de varianza que los distintos factores en su conjunto explican de la variable X_j . Es, por tanto, la parcela de esa variable que entra en contacto con el resto de variables. Varía entre 0 (los factores no explican nada de la variable) y 1 (los factores explican el 100% de la variable).

$Var(e_j)$: es lo que llamamos especificidad y representa la contribución del factor único a la variabilidad total de X_j .

$l_{1h}^2 = l_{1h}^2 + l_{2h}^2 + \dots + l_{ph}^2 = g_h$: es lo que se llama eigenvalue (autovalor) y representa la capacidad del factor h para explicar la varianza total de las variables. Si las variables originales estuviesen tipificadas, la varianza total sería igual a p y $g_{h/p}$ representaría el porcentaje de varianza total atribuible al factor h .

El objetivo del AF será, por tanto, obtener los factores comunes de modo que expliquen una buena parte de la variabilidad total de las variables.

¿Cuándo es adecuado realizar un AF?

Un AF resultará adecuado cuando existan altas correlaciones entre las variables, que es cuando podemos suponer que se explican por factores comunes. El análisis de la matriz de correlaciones será pues el primer paso a dar. Analíticamente, podemos comprobar el grado de correlación con las siguientes pruebas o test:

2.3.1.1 Test de esfericidad de Barlett

Es necesario suponer la normalidad de las variables. Contrasta la H_0 de que la matriz de correlaciones es una matriz identidad (incorrelación lineal entre las variables). Si, como resultado del contraste, no pudiésemos rechazar esta H_0 , y el tamaño de la muestra fuese razonablemente grande, no se debería considerar la realización de un AF, ya que las variables no están correlacionadas.

El estadístico de contraste del test de Bartlett es:

$$B = - (n - 1 - (2p + 5)/6) \ln | R^* |, \text{ bajo la hipótesis nula resulta: } \chi^2_{(p^2 - p)/2}$$

donde:

p : es el número de variables y

$| R^* |$: es el determinante de la matriz de correlaciones muestrales.

2.3.1.2 Índice KMO(Kaiser-Meyer-Olkin) de adecuación de la muestra

KMO se calcula como:

$$KMO = \frac{\sum_{i \neq j} \sum_{j \neq i} r_{ji}^2}{\sum_{i \neq j} \sum_{j \neq i} r_{ji}^2 + \sum_{i \neq j} \sum_{j \neq i} a_{ji}^2}$$

donde:

r_{ji} - coeficiente de correlación observada entre las variables j y h .

a_{ji} - coeficiente de correlación parcial entre las variables j y h .

Estos coeficientes miden la correlación existente entre las variables j y h , una vez eliminada la influencia que las restantes variables ejercen sobre ellas. Estos efectos pueden interpretarse como los efectos correspondientes a los factores comunes, y por tanto, al eliminarlos, a_{ji} - representará la correlación entre los factores únicos de las dos variables, que teóricamente tendría que ser nula. Si hubiese correlación entre las variables (en cuyo caso resultaría apropiado un AF), estos coeficientes deberían estar próximos a 0, lo que arrojaría un KMO próximo a 1. Por el contrario, valores del KMO próximos a 0 desaconsejarían el AF.

Está comúnmente aceptado que:

- Si $KMO < 0.5$ no resultaría aceptable para hacer un AF.
- Si $0.5 < KMO < 0.6$ grado de correlación medio, y habría aceptación media.
- Si $KMO > 0.7$ indica alta correlación y, por tanto, conveniencia de AF.

2.3.1.3 Medida de adecuación de la muestra para cada variable (MSA)

Este índice es similar al KMO, pero para cada variable. La j -ésima variable de MSA viene dada por la siguiente expresión:

$$MSA(j) = \frac{\sum_{j \neq i} r_{ji}^2}{\sum_{j \neq i} r_{ji}^2 + \sum_{j \neq i} a_{ji}^2}$$

Si el valor del MSA fuera pequeño, no se aconsejaría el AF. Por el contrario, valores próximos a 1 indicarían que la variable X_j es adecuada para incluirla con el resto en un AF. En muchas ocasiones, se eliminan las variables con MSA muy bajo. (diagonal principal de la matriz de correlación anti-imagen).

2.3.1.4 Correlación antiimagen AIC

El coeficiente de correlación antiimagen es el negativo del coeficiente de correlación parcial entre dos variables. Si existiesen factores comunes, esperaríamos pequeños coeficientes de correlación parcial. Por ello, el AF es aplicable cuando en la matriz de correlaciones antiimagen hay muchos coeficientes pequeños.

2.3.2 FASES DEL ANÁLISIS FACTORIAL

2.3.2.1 Extracción de Factores Comunes

Existen distintos métodos de estimación de los coeficientes de la matriz factorial L : los más comunes (para un AF exploratorio) son el método de las Componentes Principales y el método de Ejes Factoriales.

Para proceder a la estimación de los coeficientes factoriales vamos a partir de la identidad fundamental del AF:

$$R = LL' + w$$

donde:

R : es la matriz de correlaciones entre las variables.

W : es la matriz de varianzas y covarianzas de los factores únicos.

Consideramos una transformación de R : $R - w = LL' = R^*$ (matriz de correlación reducida) cuyos elementos diagonales son las comunalidades y el resto, los coeficientes de correlación lineal entre las variables originales. La idea consiste en determinar L partiendo de alguna estimación para R^* , y a partir de ella calcular los coeficientes de la matriz L . Podemos optar por dos métodos:

Método 1 - AF de Componentes Principales (ACP)

El método de componentes principales se basa en suponer que los factores comunes explican el comportamiento de las variables originales en su totalidad de manera que el modelo es:

$$X = LF$$

Las comunalidades iniciales de cada variable son igual a 1, porque el 100% de la variabilidad de las p -variables se explicará por los p factores. Evidentemente, carecería de interés sustituir las p -variables originales por p - actores que, en ocasiones, son de difícil interpretación. No obstante, si las correlaciones entre las p -variables fuesen muy altas, sería de esperar que unos pocos factores explicasen gran parte de la variabilidad total. Supongamos que decidimos seleccionar r factores. La comunalidad final de cada variable indicará

la proporción de variabilidad total que explican los r factores finalmente seleccionados.

La estimación de los coeficientes I_j se obtiene diagonalizando la matriz de correlaciones.

Método 2 - AF de Ejes Factoriales (PAF)

En este método partimos de la base de que sólo una parte de la variabilidad total de cada variable depende de factores comunes y, por tanto, la comunalidad inicial no será 1. Estima dichas comunalidades mediante los coeficientes de determinación múltiple de cada variable con el resto. Se sustituyen estos valores en la diagonal principal de la matriz R^* y se procede a efectuar un ACP. Una vez obtenido el resultado, se estiman de nuevo las comunalidades, se vuelven a sustituir en la diagonal principal de la matriz R^* y el proceso se retroalimenta hasta alcanzar un criterio de parada (por ejemplo cuando la diferencia entre las comunalidades de dos iteraciones sucesivas sea menor que una cantidad prefijada).

La elección de uno u otro método (ACP o PAF) depende de los objetivos del AF. Así el ACP es adecuado cuando el objetivo es resumir la mayoría de la información original (varianza total) con una cantidad mínima de factores con propósitos de predicción. El AFC resulta adecuado para identificar los factores subyacentes o las dimensiones que reflejan qué tienen en común las variables. El inconveniente del método PAF es que el cálculo de las comunalidades requiere mucho tiempo y muchos recursos informáticos y, además, no siempre se pueden estimar o, incluso, pueden ser no válidas (comunalidades menores que 0 o mayores que 1).

Empíricamente, se llega a resultados muy parecidos cuando el número de variables excede de 30 o las varianzas compartidas exceden de 0.6 para la mayoría de las variables.

2.3.2.2 Rotación de Factores

La interpretación de los resultados del AF se basará en el análisis de las correlaciones entre las variables y los factores que como sabemos viene dado por las cargas factoriales.

Para que dicha interpretación sea factible, es recomendable que: las cargas factoriales de un factor con las variables estén cerca de 0 ó de 1. Así, las variables con cargas próximas a 1 se explican en gran parte por el factor, mientras que las que tengan cargas próximas a 0 no se explican por el factor.

Una variable debe tener cargas factoriales elevadas con un sólo factor. Es deseable que la mayor parte de la variabilidad de una variable sea explicada por un solo factor. No debe haber factores con similares cargas factoriales

Así, si con la solución inicial no se consiguiese una fácil interpretación de los factores, éstos pueden ser rotados de manera que cada una de las variables tenga una correlación lo más próxima a 1 con un factor y a 0 con el resto de factores. Como hay menos factores que variables, conseguiremos que cada factor tenga altas correlaciones con un grupo de variables y baja con el resto. Si examinásemos las características de las variables de un grupo asociado a un factor, se podrían encontrar rasgos comunes que permitan identificar el factor y darle una denominación que responda a esos rasgos comunes. Así, conseguiremos desvelar la naturaleza de las interrelaciones existentes entre las variables originales. Los tipos de rotaciones más habituales son la ortogonal y la oblicua.

La rotación ortogonal permite rotar los factores estimados inicialmente, de manera que se mantenga la incorrelación entre los mismos. El método más utilizado de rotación es la varimax (Varianza máxima), ideado por Kaiser. La rotación oblicua no mantiene la ortogonalidad de los factores, lo que nos lleva a aceptar que dos o más factores expliquen a la vez una misma realidad. Las communalidades finales de cada variable permanecen inalteradas con la rotación.

2.3.2.3 Cálculo de las Puntuaciones factoriales

Una vez estimados los factores comunes, es importante calcular las puntuaciones de los sujetos (individuos u objetos) investigados para saber cuánto puntúan en cada factor. Así, podremos:

Sustituir los valores de las p variables originales para cada sujeto de la muestra por las puntuaciones factoriales obtenidas. En la medida en que el número de factores es menor que el número de variables iniciales, si el porcentaje de explicación de la varianza total fuese elevado, dichas puntuaciones factoriales podrían sustituir a las variables originales en muchos problemas de análisis o predicción. Además, muchas técnicas estadísticas se ven seriamente afectadas por la correlación entre las variables originales. En la medida en que las puntuaciones factoriales estén incorrelacionadas podrán utilizarse en ulteriores análisis.

Colocar a cada sujeto en una determinada posición en el espacio factorial y conocer qué sujetos son los más raros o extremos, dónde se ubican ciertos grupos de la muestra, los más jóvenes frente a los mayores; los de clase alta frente a los de clase media o baja; los creyentes frente a los no creyentes, etc obteniendo en qué factores sobresalen unos y otros.

2.4 ANÁLISIS DISCRIMINANTE

Es una técnica estadística multivariable que permite explicar la pertenencia de individuos u objetos a grupos preestablecidos (por ejemplo, fumador o no fumador, comprador habitual, comprador ocasional o no comprador, etc.). El principal objetivo del análisis discriminante es predecir la probabilidad de pertenencia de un sujeto u objeto a un grupo o clase, basándose en las distintas variables predictoras.

El AD es una técnica que pretende, por un lado, explicar la pertenencia de un elemento a un grupo dado en función de las VI disponibles y, por otro, predecir a qué grupo pertenecerá un elemento del que se conocen los valores de una serie de variables.

Dicho de otro modo, el AD pretende:

Por un lado, crear un modelo explicativo (expresado por la ecuación de "función discriminante") compuesto por las VI capaces de discriminar de forma significativa entre los casos pertenecientes a las categorías de la VD y
Por otro, calcular la probabilidad de que un caso sea incluido en una categoría de la VD, en función de los valores de VI.

En el Análisis de Conglomerados o Cluster (AC) hay que definir los grupos, en el AD, los grupos están ya predefinidos. El AC es un paso previo al AD: el AC forma los grupos para que, luego, el AD determine qué variables influyen en la diferenciación de los mismos.

Tipos de Análisis Discriminante

Análisis Factorial Discriminante. Explica qué VI contribuyen significativamente a la diferenciación en 2 o más grupos. Permite predecir la pertenencia de un caso a cierto grupo y realizar predicciones, asignando a cada sujeto al grupo más cercano a su puntuación factorial.

Análisis de las Funciones Discriminantes. Predice la pertenencia de un individuo a un cierto grupo, en función de la probabilidad calculada, conocidos una serie de datos. Calcula las probabilidades de pertenecer a un determinado grupo, según técnicas de decisión bayesianas.

Dada una variable dependiente cualitativa y un conjunto de una o más variables independientes cuantitativas, el análisis discriminante consiste en obtener unas funciones lineales de las variables independientes, denominadas

funciones discriminantes, que permitan clasificar a los individuos en una de las subpoblaciones o grupos establecidos por los valores de la variable dependiente.

Sea, $\{x_{i1}, \dots, x_{ip}\}, i = 1, \dots, n$, la muestra de n observaciones de las variables independientes cuantitativas X_1, \dots, X_p , en los k grupos de individuos establecidos por los valores de la variable dependiente cualitativa Y . Si n es el número de individuos en la muestra y p el número de variables independientes, la tabla de datos establecida por las $n \times p$ observaciones tendrá n filas y p columnas. Cada fila puede ser considerada como un punto en un espacio de p dimensiones. Las coordenadas de cada punto se obtendrán a partir de los valores en las p variables para el individuo correspondiente. A partir de la representación de las n filas se trata de extraer un nuevo espacio de pequeña dimensión tal que, al proyectar la nube de puntos sobre dicho espacio, por un lado, los puntos correspondientes a individuos del mismo grupo estén próximos y, por otro, los correspondientes a individuos de distintos grupos estén alejados. Los ejes de este nuevo espacio serán las funciones discriminantes.

La expresión de una función discriminante, D_s , será:

$$D_s = B_{s1}X_1 + \dots + B_{sp}X_p + B_{s0}$$

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

$$d_s = B_{s1}x_{i1} + \dots + B_{sp}x_{ip} + B_{s0}$$

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. El porcentaje de casos correctamente clasificados será un índice de la efectividad de las funciones discriminantes. Si dichas funciones son efectivas sobre la muestra observada, es de esperar que también lo sean cuando se trate de clasificar a un individuo para el que se desconoce a cuál de los grupos pertenece.

Selección de Variables: En el caso del análisis discriminante el criterio que se considera para la selección de variables es el lambda de Wilks..

El lambda de Wilks para un conjunto de p variables independientes mide las desviaciones dentro de cada grupo respecto a las desviaciones totales sin distinguir grupos, en el espacio p -dimensional generado por los valores de las p variables. Si su valor es pequeño, la variabilidad total será debida a las diferencias entre grupos, y por tanto, el conjunto de variables correspondiente discriminará los grupos. Por el contrario, si su valor es próximo a 1 los grupos estarán mezclados y el conjunto de variables independientes no será adecuado para construir las funciones discriminantes. En particular, en el primer paso, la variable candidata a ser seleccionada será aquella tal que el valor de la Lambda de Wilks sea el mínimo de entre los obtenidos sobre cada uno de los conjuntos formados por cada uno de las variables independientes. Además del criterio de selección existe un criterio de eliminación (F de entrada, F de salida y Tolerancia).

El estadístico F de entrada evalúa la disminución que se produciría en la Lambda de Wilks si la variable correspondiente fuera seleccionada. Si su valor es grande la disminución será significativa y la selección adecuada. El estadístico F de salida evalúa el incremento que se produciría en lambda de Wilks si la variable correspondiente fuera eliminada. La tolerancia de una variable X_j con las variables $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p$ se define como:

$$Tol_j = 1 - R_j^2.$$

Si el valor de la tolerancia es igual a 0, la variable X_j será una combinación lineal de las restantes variables.

El método Stepwise para la selección de variables sigue el siguiente proceso:

En el primer paso se introduce la variable que proporcione el mínimo valor de Lambda de Wilks, siempre y cuando verifique el criterio de entrada. En caso

contrario, el proceso finalizará sin que ninguna variable sea seleccionada y, en consecuencia, no será posible construir las funciones discriminantes a partir de la información de las variables independientes.

En el segundo paso se introduce la variable que, junto con la primera seleccionada, proporcione el mínimo valor de la Lambda de Wilks, siempre que verifique el criterio de entrada. En caso contrario, el proceso finalizará, y el conjunto de funciones discriminantes se reducirá a una única función que coincidirá con la variable independiente seleccionada en el primer paso.

En el siguiente paso se introduce la variable que, junto con las previamente seleccionadas, proporcione el mínimo valor de Lambda de Wilks, siempre que verifique el criterio de entrada. Si al seleccionar una variable, el mínimo valor de la F de salida para las variables previamente seleccionadas verifica el criterio de eliminación, antes de proceder a la selección de una nueva variable, se eliminará la variable correspondiente.

Cuando ninguna variable verifique el criterio de eliminación se vuelve a la etapa anterior. La etapa anterior se repite hasta que ninguna variable no seleccionada satisfaga el criterio de selección y ninguna de las seleccionadas satisfaga el de eliminación, o se alcance el máximo número de pasos.

El estadístico de Lambda de Wilks permite contrastar la hipótesis nula de que los centros de los grupos son iguales. En concreto, al rechazarse la hipótesis nula se puede concluir que la información que aportará cada una de las funciones a la hora de clasificar a los casos será estadísticamente significativa al nivel 0.05.

La correlación canónica y el autovalor asociado a una función discriminante son dos medidas, relacionadas con la Lambda de Wilks, que permitirán evaluar la información que aportará 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 totales 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.

La clasificación de los individuos se realiza a partir de las funciones discriminantes tipificadas, o lo que es equivalente, de las funciones discriminantes expresadas en términos de las variables independientes tipificadas. El eje construido a partir de la tipificación de las variables pasará por el origen y, por tanto, el término independiente en la función discriminante correspondiente será igual a cero. A partir de las funciones discriminantes es posible obtener una regla que permita clasificar a los individuos en uno de los k grupos. Una técnica que se utiliza frecuentemente se base en la regla de Bayes. La probabilidad estimada de que un individuo i , con puntuaciones discriminantes $d_{i1}, d_{i2}, \dots, y d_{ic}$, pertenezca al grupo j se denota como $P(G_j / D)$ y se obtiene mediante la expresión:

$$P(G_j / D) = \frac{P(D / G_j)P(G_j)}{\sum_{j=1}^k P(D / G_j)P(G_j)} \quad j = 1, \dots, k$$

donde:

$$D = (d_{i1}, \dots, d_{is}, \dots, d_{ic})$$

$P(G_j)$ es la probabilidad a priori de pertenecer al grupo j .

$P(D / G_j)$ es la probabilidad de que, supuesto que el individuo pertenece al grupo j , sus puntuaciones en las funciones discriminantes sean $d_{i1}, d_{i2}, \dots, y d_{ic}$,

Un individuo será clasificado en el grupo para el que la probabilidad será máxima, es decir, será clasificado en G_j si:

$$P(G_j / D) = \text{máx} [P(G_1 / D), \dots, P(G_k / D)]$$

El porcentaje de casos correctamente clasificados será un índice de la efectividad de la función discriminante. En cualquier caso, al evaluar este índice deberá tenerse en cuenta la tasa de clasificaciones correctas esperadas según las probabilidades a priori. Las probabilidades a priori, si son conocidas, deben ser indicadas al inicio del análisis. En caso contrario, se considerará que todos los grupos tienen la misma probabilidad o, lo que es equivalente, que la probabilidad a priori de cada uno de los k grupos es igual a $1/k$. En ocasiones, las probabilidades a priori son desconocidas pero se supone que la muestra observada es representativa de la población objeto de análisis, por lo que un criterio adecuado sería considerar que la probabilidad a priori de un determinado grupo coincide con la proporción de casos en dicho grupo.

Para la predicción de los casos se aplica el criterio para clasificarlos el de máxima probabilidad a posteriori. Por tanto, puede suceder que dos casos fueran clasificados en un mismo grupo y sin embargo, las probabilidades de pertenecer a dicho grupo fueran muy distintas.

En este sentido, a la hora de clasificar a los individuos sería conveniente conocer no sólo la máxima probabilidad sino, incluso, las probabilidades de pertenecer a cualquiera de los restantes grupos.

2.5 HERRAMIENTA ESTADÍSTICA – PAQUETE SPSS

Un paquete estadístico es un conjunto de programas y subprogramas organizado de forma que cada uno de ellos está interconectado con todos los demás. De esta forma, la utilidad del conjunto integrado es mayor que la suma de las partes. Un paquete estadístico es muy similar a un paquete ofimático (por ejemplo, el paquete Microsoft Office).

Otro modo de definir un paquete estadístico es como colección de procedimientos conectados de manera que funcionan de manera conjunta; es decir, para pasar de uno a otro no se necesita salir del programa y volver a él. En

su lugar, un paquete estadístico permite que se puedan aplicar a un mismo fichero de datos un conjunto ilimitado de procedimientos estadísticos de manera sincronizada, sin salir nunca del programa. SPSS es uno de los principales paquetes estadísticos del mercado. Otros ejemplos importantes de paquetes estadísticos son **SAS**, **Statistica**, **Matlab**, **Statgraphics** y **Minitab**. En el pasado, había otros paquetes estadísticos, como BMDP y Systat, absorbidos por SPSS.

2.5.1 PARTES DEL PAQUETE ESTADÍSTICO SPSS

SPSS es un paquete estadístico (no un programa como puede serlo Microsoft Word), que está compuesto por varias rutinas y subrutinas dependientes todos ellos del núcleo, llamado módulo SPSS Base, sin cuya instalación no funcionan los demás módulos. El módulo SPSS Base tiene, además de funciones estadísticas, la importante misión de crear los menús de apertura de ficheros (sean de datos, de resultados o de sintaxis), de edición de textos, de creación de gráficos, etc. Es, por tanto, del todo indispensable para que toda la maquinaria de SPSS funcione.

- Un paquete estadístico tiene muchas utilidades, ya que puede ser utilizado como:
- Hoja de Cálculo que permite realizar operaciones aritméticas, algebraicas y trigonométricas.
- Gestor de Bases de Datos que permite gestionar de modo dinámico los datos, pues se pueden actualizar los cambios operados o realizar informes personalizado.
- Generador de Informes que permite preparar atractivos informes de una investigación realizada, incorporando en un mismo archivo el texto del informe, las tablas y resultados estadísticos que se necesite presentar e, incluso, los gráficos.

- Analizador de datos que tiene la capacidad de extraer de un fichero toda la información, permitiendo ejecutar procedimientos estadísticos descriptivos, inferenciales y multivariantes.
- Ejecutor de Minerías de Datos que puede llevar a cabo búsquedas inteligentes, para extraer información oculta, elaborando árboles de decisión, segmentaciones de mercados o diseños de redes neuronales de inteligencia artificial.
- Data Mining es una metodología de trabajo específicamente concebida para descubrir "filones" de información en el interior de una "montaña" de datos.

Luego de actualizar algunos conocimientos estadísticos y de análisis de datos se presenta a continuación la propuesta metodológica.

CAPITULO 3

3.1 PROPUESTA TEÓRICA – PRÁCTICA

El presente capítulo comprende la explicación de la metodología y el proceso conceptual que se debe ir desarrollando. A continuación se observa el proceso en 10 pasos:

Gráfico 4: Propuesta de la metodología

METODOLOGÍA	Responsable Canal Físico y Gerente de Marketing	Especialista
Diagnóstico de situación de las agencias		
Definición de Variables		
Selección de variables y análisis de datos		
Partición de la base de datos		
Aplicación de un Análisis de Componentes Principales		
Definición de clusters		
Descripción de los grupos		
Definición de los modelos de atención al cliente		
Aplicación de un Análisis Discriminante		
Validación de Resultados		

Los segmentos formados desde los datos de comportamiento de las agencias envuelven un algoritmo analítico o metodología que comprende los siguientes pasos:

Paso 1: Diagnóstico de la situación de las agencias

En este paso se debe analizar la información relevante que tiene cada agencia,

¿cuántos clientes atiende?

¿cuántas transacciones realizan los clientes?

¿si es o no eficiente el recurso humano en cada agencia?

¿qué rentabilidad genera al banco cada agencia?

Paso 2: Definición de variables

En una actividad conjunta con los funcionarios del banco se definieron algunos puntos de vista acerca de los aspectos de comportamiento de las agencias que son de interés para el Gerente del Canal Físico y para representantes del campo de negocio (productos, canales, presupuesto, etc). Revisar y Seleccionar las variables desde la lista completa de variables disponibles en los datos de comportamiento, ésta debe ser una actividad conjunta entre el analista o investigador y el encargado de mercadeo o el gerente del canal físico. El encargado de mercadeo o el gerente del canal físico tienen algunos puntos de vista a priori acerca de cuáles aspectos del comportamiento son importantes para la clasificación que desea obtener y cuáles no son relevantes para sus planes de mercadeo o estrategias de estructuras de negocio. El analista debe tener una apreciación de cuáles variables se pueden incluir en el análisis según los valores y el procedimiento estadístico a ser aplicado. Esta metodología de clasificación puede generar resultados equivocados si se utilizan variables dicotómicas (0,1) directamente, o no se realizan las respectivas transformaciones. Las variables que fueron seleccionadas como principales y de interés para el estudio, como un aporte de esta tesis, fueron:

- Monto total que tiene el cliente en el banco: según criterios de anclaje para cada relación cliente-agencia y de todos los productos.
- Número de colaboradores del banco que se encuentran en cada agencia.
- Costos totales generados en cada agencia, costos fijos más costos variables.
- Margen financiero, otros ingresos y comisiones que genera cada agencia.
- Capacidad instalada en minutos, cantidad de minutos estándar que deben atender los cajeros o ventanillas.
- Capacidad utilizada en minutos, cantidad de minutos reales que atendieron los cajeros o ventanillas.
- Número de transacciones realizadas en cada agencia por cada segmento de clientes.
- Transacciones propias de los productos del banco.
- Transacciones por servicios prestados: número de transacciones realizadas por personas no clientes del banco o que corresponden a servicios.
- Número de clientes de todos los segmentos pertenecientes a una agencia.
- Número total de clientes que realizan sus transacciones en la agencia correspondiente.
- Número de transacciones realizadas por los clientes de todos los segmentos por agencia.
- Número de transacciones total.
- Monto total de los clientes de alto valor y pequeña empresa.

Paso 3: Selección de variables y Análisis de los datos

Las variables no deben tener datos erróneos, ni mal generados. Los datos en blanco o faltantes en las variables cuantitativas de transaccionalidad y número

de clientes deben ser reemplazados por el valor de cero; pero en las variables capacidad instala, capacidad utilizada, margen financiero y costos generales no pueden aparecer valores faltantes, puesto que implicaría datos no calculados o no generados, entonces se elimina la agencia correspondiente del análisis. En este paso se obtienen estadísticas descriptivas de cada variable, como son: media, moda, desviación estándar, rangos, máximos, mínimos, sesgo y kurtosis y es necesario identificar los puntos atípicos, es decir aquellos que están fuera del intervalo de la media ± 3 veces la desviación estándar. Si se identifican puntos atípicos o faltantes por errores se debe eliminar la agencia.

Paso 4: Partición de la base de datos

En este paso se divide la base en tres grupos: un grupo de entrenamiento formado por el 60% de las agencias, otro de validación con el 20% y finalmente uno de test o prueba con el 20%.

Paso 5: Aplicación de un Análisis de Componentes Principales

Se considera el 100% de la base para este análisis.

Análisis de Correlaciones: Se analiza una matriz de correlaciones y se observa que los valores sean altos, es decir cercanos a 1,

Además se aplica el test de esfericidad de Barlett, según el cual, si se rechaza la hipótesis nula de que la matriz de correlaciones es la identidad entonces se puede continuar con el análisis de componentes principales.

Luego se analiza el Índice KMO (Kaiser-Meyer-Olkin), y si es menor a 0.5 no se debe realizar el análisis de componentes principales.

Una vez que se obtienen los primeros resultados del análisis de componentes principales se debe analizar: la comunalidad que debe ser cercana a 1, cada valor propio que debe ser superior a 1. Si no se obtiene que el primer factor explique un porcentaje de variabilidad alto se corre un proceso VARIMAX, o de ortogonalidad y se validan nuevamente los parámetros mencionados.

Finalmente se seleccionan las puntuaciones factoriales para cada individuo y se forma un índice con puntajes para cada agencia.

Paso 6: Aplicar un Análisis Cluster

Formado el índice con el primer factor del análisis de componentes principales se realiza un análisis cluster exploratorio o jerárquico y se solicita una solución de cluster abierta, es decir, para un rango de soluciones de 2 a 5 clusters. Luego se genera un dendograma y como medida de similitud se utiliza la medida de intervalo o distancia euclídea, puesto que son variables cuantitativas continuas, también se aplica como método de agrupación la relación promedio entre grupos. Con los resultados de la tabla de pertenencia a los cluster, un análisis de frecuencias y el dendograma se define el número óptimo de clusters y se comprueba que el proceso de clustering asigne un número homogéneo y consistente de casos a cada uno. Cuando se obtiene el número óptimo de grupos se realiza un análisis de Cluster K-means y se analizan los resultados de: una tabla ANOVA donde el nivel de significancia de cada variable debe ser menor a 0.05 y el grado de contribución que se observa en el valor de F, mientras mayor sea el valor, mayor es la contribución.

Paso 7: Descripción de los grupos

Usar toda la información disponible en el conjunto de variables seleccionadas para describir cada uno de los grupos: consiste en explicar las características distintivas y dominantes de cada segmento o grupo y crear niveles basados en las descripciones encontradas. Esta caracterización permitirá entender cada uno de los complejos grupos de agencias para ser usados en el proceso de planificación de las estrategias. Los componentes principales también pueden contribuir con información para cada grupo, en especial si son de fácil interpretación los coeficientes de las funciones lineales. Seguidamente, se realiza: un gráfico de líneas múltiples, este permite representar de un modo visual los valores medios de los perfiles de los 3 clusters sobre las variables originales; un procedimiento de medias entre las variables originales y la variable que

contiene información de los grupos y el test de linealidad, que analiza si los grupos son homogéneos, y se debe rechazar esta hipótesis observando el nivel de significancia de la tabla Anova, si es menor a 0.05

Paso 8: Definición de los modelos de atención al cliente

En esta parte se identifican las características encontradas en el análisis cluster y se determinan las estructuras de negocio que se ajustan a cada grupo según el negocio y la transaccionalidad.

Paso 9: Aplicar un Análisis Discriminante

Para desarrollar reglas que permitan asignar el objeto a un segmento: Esta técnica toma la clasificación realizada por el análisis cluster como la variable dependiente y las variables, seleccionadas en el inicio de la investigación, como explicativas, y encuentra un conjunto de reglas lógicas o scoring para la clasificación. Esto significa que el objeto de estudio (agencias) puede ser clasificado en el futuro según las variables aplicadas en un grupo determinado, para este análisis se considera únicamente el grupo de entrenamiento. Primero se identifica la variable de agrupación, luego las variables independientes. Para la selección de las variables que influyen en el análisis se considera el método Stepwise y las probabilidades previas pueden ser: según el tamaño de cada grupo o como proporciones iguales. En los resultados se observa el último paso de selección de variables en el cual se analiza que el Lambda de Wilks sea cercano a 1 y los F de entrada superiores a 3.84, además el aporte de la función canónica discriminante cercano al 100%, se analiza los resultados del test de Box's M cuyo nivel de significancia debe ser menor a 0,05, los valores propios que expliquen más del 80% de la información, un test de las funciones de Lambda de Wilks, cuyo nivel de significancia debe ser menor a 0,05 y la matriz de confusión donde las agencias correctamente clasificadas debe ser superior al 80%. Los coeficientes de la función discriminante canónica determinan el peso de cada variable en cada función para clasificar a las agencias. Una vez que se ha corrido para el grupo de entrenamiento se corre el mismo proceso para el grupo de

validación junto con el de entrenamiento y se observan los cambios en los coeficientes de las funciones discriminantes, cuya variación debe ser mínima. Obtenidos los grupos de las agencias, para el conjunto de validación se estudia la tabla de doble entrada o tabla cruzada entre lo pronosticado y lo real, el porcentaje de acierto corresponderá a las agencias que están bien clasificadas.

En conclusión, las fases del Análisis Discriminante son:

- Se determina el espacio de discriminación
- Se obtienen las funciones discriminantes
- Se calculan las coordenadas de cada centroide
- Se asigna cada dato al grupo más próximo.

Paso 10: Validación de los resultados

Seguidamente se incluye en el análisis al grupo de prueba o test y se valida del mismo modo el porcentaje de aciertos y los coeficientes de las funciones discriminantes. Para complementar el análisis de clasificación de las agencias, se cruzan los grupos con la regional.

3.2 APLICACIÓN DE LA METODOLOGÍA DE CLASIFICACIÓN DE LAS AGENCIAS DE UNA INSTITUCIÓN BANCARIA PRIVADA

El grupo de agencias en las cuales se aplicará la metodología corresponde a una institución bancaria del Ecuador, con información del año 2003. Inicialmente se describe algunos aspectos importantes para estandarizar conocimientos sobre el funcionamiento de una agencia y sus características.

3.2.1 UBICACIÓN GEOGRÁFICA DE LAS AGENCIAS

Las agencias del banco analizadas en la presente tesis, tiene la siguiente distribución por región:

Tabla 4: Distribución del número de agencias por región

REGIÓN	%_agencias
ORIENTE	3.59%
MACHALA	5.39%
CUENCA	5.99%
SIERRA	16.17%
GUAYAQUIL	17.96%
COSTA	24.55%
QUITO	26.35%
Total general	100.00%

Como se puede observar, la mayor concentración de las agencias se observa en la ciudad de Quito, seguido por la región Costa que excluye a Guayaquil, luego la ciudad de Guayaquil, la región Sierra, que corresponde al resto de la Sierra excepto Quito, y con una menor concentración la zona Oriental.

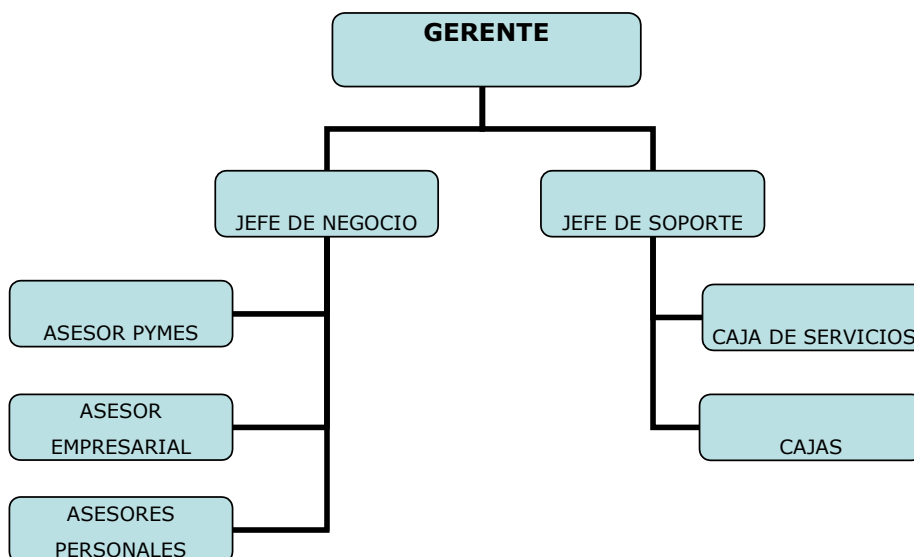
Las ciudades de Cuenca y Machala se consideran importantes dentro del desarrollo económico del País, así mismo, el Banco tiene una mejor cobertura en estas ciudades a comparación del resto, con excepción de Quito y Guayaquil.

3.2.2 MODELO DE NEGOCIO ESTÁNDAR

Un banco tiene una estructura funcional en cada sucursal y agencia, siendo una sucursal de mayor tamaño; además, la sucursal depende de la regional y la regional del Gerente del Canal Físico.

Las agencias, que son puntos de contacto entre el banco y el cliente, tienen un organigrama general y estándar, con mayor o menor número de colaboradores, como se muestra a continuación:

Gráfico 5: Organigrama estándar



Fuente: Institución Financiera

Los cajeros están distribuidos en ventanillas de extensión, punto pago, cajas de servicios, caja VIP, Autobanco y caja de negocios, y dependiendo del tamaño de la agencia son en mayor o menor número los colaboradores.

En una sucursal existen: Regional de Tecnología, Regional de Riesgo, Supervisor Operativo, Operadores (Cartera y Operaciones), Técnico Administrativo, Telefonista, persona de proveeduría, guardias y Servicios Tercerizados, quienes dan apoyo a las agencias. Actualmente, esta distribución ha cambiado.

El gerente de la agencia tiene entre sus funciones la elaboración del presupuesto, planes y acciones comerciales, participar en la calificación de cámara, participar en los comités para autorizar solicitudes de crédito, estar pendiente del recurso humano, prestar una buena atención al cliente, mejorar la eficiencia de los procesos operativos, elaborar reportes y dar firmas de aprobación.

El jefe de agencia está encargado de las acciones comerciales y de evaluar el cumplimiento de las metas comerciales por parte de los asesores.

El jefe de soporte tiene como función coordinar y organizar los procesos operativos y contables.

El asesor bancario debe organizar y coordinar las operaciones bancarias de sus clientes, conocer los productos y tipos de venta cruzada, brindar un buen servicio al cliente y participar en campañas de venta de productos.

Los cajeros pueden ser: **de servicios**, son aquellos que entregan toda la información que el cliente requiere, en el momento que la solicita, como por ejemplo: tarjeta de débito, chequeras, cheques certificados, entrega de estados de cuenta, etc.; de **negocios**, aquellos que atienden (asesores); y **transaccionales**, aquellos que tramitan ingresos y egresos de caja.

Este organigrama se aplica a todas las agencias y no depende de las características poblacionales de la ciudad donde está ubicada la agencia, el tipo de actividad económica de la población, tasas de desempleo o niveles socio-económicos, por otro lado, el tipo de clientes que atiende o tipo de transacciones que recibe la agencia, entonces es necesario su diferenciación para ofrecer un mejor servicio, ya sea aumentando los asesores comerciales o aumentando el número de cajeros.

3.2.3 INFRAESTRUCTURA DE LAS AGENCIAS

La infraestructura de las agencias está formada por un sector de puestos de trabajo amplios para los asesores y parte administrativa, y otro con puestos de trabajo pequeños para las cajas.

Las agencias cuentan con esta distribución estándar, pero se observa que no cubren las necesidades de atención eficientemente, puesto que el número de clientes por asesor es bajo, y la demanda de atención en las cajas sobrepasa la oferta de cajeros.

3.2.3.1 Tecnología

Cada agencia está formada por una red LAN. Las computadoras conectadas a la red tienen las siguientes características: Pentium 4, 2.8GHz, 128Mb en memoria RAM y 40 Gb en el disco. Cada computador tiene un cajetín en RACK que se comunica a un switch o HUB, luego a un router y a un DTU, ubicados en un cuarto de telecomunicaciones en cada agencia. Seguidamente, se conectan con un ÚLTIMA MILLA, puede ser Interactive, Inadatos o Tramadatos, las mismas ofrecen servicio de Internet, a través de fibra óptica en sectores de fácil acceso o caso contrario por señal satelital, como por ejemplo, en las provincias del Oriente; finalmente se conectan con granjas de servidores centralizados en Quito. Esta conexión sirve para los sistemas internos del Banco, como la Intranet e ingreso de datos generales del cliente, etc. Además de contar con esta interconexión, cada agencia tiene un servidor donde está un sistema distribuido que cuenta con lenguajes como SQL, para el manejo de las bases de datos de los clientes, y SNA que hace posible la comunicación con el HOST. En este servidor está todo el sistema de cajas con información de los clientes de la agencia correspondiente.

3.2.4 CARACTERIZACIÓN DE LOS CLIENTES

En una institución bancaria existen diferentes grupos de clientes, estos grupos se definen por una combinación de:

- Los montos que maneja el cliente en sus cuentas con el banco
- Número de Transacciones
- Número de cuentas que manejan con el Banco
- La rentabilidad generada al Banco
- Si pertenecen o no a un grupo económico
- Si son clientes naturales o jurídicos
- El número de canales del banco que utiliza el cliente.

En el Banco, caso de estudio, existe una primera clasificación de clientes en: personas naturales y jurídicos, una segunda clasificación que responde a un modelo heurístico aplicando varias de las características mencionadas anteriormente y es: ALTO VALOR, MEDIO VALOR y BAJO VALOR, siendo los clientes de alto valor aquellos que generan mayor rentabilidad, altos montos en sus cuentas y una alta interacción con el banco a través de sus canales.

Los clientes de una ciudad responden a una realidad socio-económica que difiere de una ciudad a otra, en consecuencia las agencias debe ser también diferentes, en algunos casos más enfocada a ofrecer productos a cierto tipo de clientes, y otros únicamente, con servicios de cajas para no clientes del banco o transacciones especiales.

3.2.5 CARACTERIZACIÓN DE LAS TRANSACCIONES

Las transacciones generadas en la institución bancaria considerada como caso de estudio, se dividen en:

- Transacciones core: aquellas que son propias de la actividad financiera que realiza el Banco, por ejemplo: depósito en cuenta corriente en efectivo, depósito en cuenta corriente en cheque, depósito en cuenta de ahorros en cheque, depósito en cuenta de ahorros en efectivo, retiro en cuenta de ahorros en cheque y depósito en cuenta de ahorros en efectivo y pago de cheques. Este tipo de transacciones, con mayor frecuencia, la realizan los clientes del banco.
- Transacciones no core: son aquellas que resultan de convenios interinstitucionales, por ejemplo: cobranzas empresas, impuestos SRI, pago de luz, pago matrículas autos, aporte IESS, matrículas universidades, pensiones colegios, aduanas, pago tarjetas DINERS,

VISA, MASTERCARD, sector público, bono solidario, pago a proveedores, pago a jubilados, roles de pago, entre otras. Estas transacciones son realizadas por personas no clientes del Banco en la mayoría de los casos.

Las transacciones son un factor importante en la clasificación de las agencias, puesto que determina si una agencia está más enfocada en servir a los clientes del banco, o tiene un mayor número de transacciones realizadas por personas no clientes, siendo en el primer caso un enfoque de servicio bancario y en el segundo un servicio transaccional únicamente.

CAPITULO 4

4.1 DIAGNÓSTICO DE LA SITUACIÓN DE LAS AGENCIAS

La Información aplicada en la presente tesis corresponde a la información de las agencias de una institución financiera, pero el análisis corresponde a una muestra de la población total de agencias.

Las agencias atienden a un promedio de 7759 clientes mensualmente cada una, estos clientes son atendidos en un promedio por segmento distribuido de la siguiente forma: clientes en segmento corporativo (en promedio 1), clientes en segmento empresarial (en promedio 12), clientes en segmento personas (en promedio 7447) y clientes en segmento privado (en promedio 20). En el segmento personas existen clientes rentables y no rentables para el banco que han sido segmentadas en 3 grupos. Además de los clientes que se acercan a cada agencia existen personas que no tienen cuenta en el banco, como por ejemplo, aquellas que cobran el bono solidario, o realizan pagos por servicios básicos. El número de transacciones que se realizan en promedio es de 27.838, de las cuales por segmento, aproximadamente, se muestra en la siguiente tabla:

Tabla 5: Promedio de transacciones por segmento

SEGMENTO	TRANSACCIONES PROMEDIO (#)
CORPORATIVO	3609
EMPRESARIAL	1333
PERSONAL	26819
PRIVADO	572
OTROS	4579

Fuente: institución financiera

Por otro lado, el número de personas o colaboradores que atienden en una agencia, dependiendo si las agencias son grandes, medianas o pequeñas, es en promedio 14 personas.

Un grupo de colaboradores del banco constituyen los asesores comerciales y de negocio y otro, formado por los cajeros. Cada agencia tiene un promedio de 5 asesores comerciales y 8 cajeros; ratificando lo mencionado anteriormente, este número varía de acuerdo al tamaño de la agencia, ya sea: grande, mediana o pequeña.

De acuerdo al promedio de transacciones por agencia se tiene que una ventanilla o caja laborando 25 días al mes y 8 horas diarias realiza una transacción en 3 minutos, entonces un cliente que está en la cola después de cinco personas aproximadamente debe esperar 15 minutos. Esta es una de las razones por la cual es necesario mejorar el servicio al cliente en las agencias, el tiempo que consumen los clientes y no clientes en la cola. La respuesta a este problema sería optimizar el número de cajeros, pero con una visión más global se trata de disminuir los gastos operativos con una mejor distribución del negocio.

Por otro lado, el tiempo que un asesor comercial utiliza con un cliente es aproximadamente 5 minutos, pero esto depende del cliente y del tipo de necesidad.

Resumiendo estos dos análisis, se tiene que existen agencias con sobrecobertura de servicio al cliente en cuanto a negocios, mientras otras tienen una subcobertura en cuanto al servicio en cajas; así por ejemplo el máximo número de transacciones que tiene una agencia es de 73.111, y para cubrir esta demanda debería tener aproximadamente 18 cajeros y un asesor, por tanto la agencia tiene una tendencia más transaccional que de negocio, pero el banco mantiene los mismos esquemas para la distribución del personal, sin considerar las necesidades de cada agencia.

Continuando con este diagnóstico, se tiene que por agencia el monto mensual en gastos es en promedio de 11.542 dólares, aproximadamente, por tanto lo que se pretende con esta tesis es lograr la diferenciación de las agencias entre negocio y servicio de cajas para obtener una disminución de los gastos según el tipo de clientes que atienden, el tipo de transacciones que realizan y la eficiencia que tiene cada agencia.

Otro aspecto crítico es el índice de productividad, la relación entre la capacidad utilizada y la capacidad instalada, mientras más cercano a 1 la agencia tiene una mejor utilización de la capacidad instalada, y más cercano a 0 una subutilización de los recursos. Este índice en promedio es de 0.84 para las agencias del banco, del caso de estudio, pero se observa que en algunas agencias existe una saturación del servicio con un índice de productividad de 2.42 y una subutilización con un índice de 0.036, en los dos casos las agencias muestran un índice de productividad ineficiente.

En el banco existe una rotación de clientes aproximadamente del 5% y un crecimiento mensual del 10%, por tanto, el crecimiento del banco es mayor a la deserción de sus clientes, entonces si se mantiene este crecimiento en pocos años el servicio a los clientes estará más deteriorado y existirá un crecimiento exagerado de los gastos y disminución de los ingresos debido a la disminución de los clientes por el mal servicio, convergiendo a la ineficiencia de las agencias.

La diferenciación de las agencias por el tipo de clientes que atienden también es un factor importante en la definición de nuevas estructuras de negocio. Los clientes ubicados en el segmento personas y que pertenecen a un nivel rentable dejan un ingreso promedio por agencia superior al resto de segmentos, entonces viene a cada uno la pregunta, ¿merecen estos clientes la misma atención en las agencias que el resto de clientes, desde el punto de vista del servicio? El ingreso que dejan cada uno de estos grupos en cada agencia depende también de su ubicación geográfica.

Las medidas que se utilizan para medir la utilidad de la aplicación de la clasificación de las agencias bancarias, son:

- La satisfacción el cliente: para medirla se debe realizar una encuesta, cuyo formulario debe ser aplicado antes y después de la clasificación, para comparar las dos situaciones y evaluar si el cliente percibe algún tipo de mejora. De la experiencia que se tiene en el banco, en cuanto al levantamiento de información, el servicio al cliente tiene deficiencias. La herramienta aplicada para el levantamiento de información es una encuesta.
- Costos: el costo inicial promedio, medido como el promedio del monto de activos fijos, que tiene una agencia es de **60.746 USD**.

La comparación de estas medidas, luego de aplicada la clasificación, indicará el mejor uso y distribución de los recursos del banco.

4.2 TIPIFICACIÓN DEL PROBLEMA

4.2.1 DEFINICIÓN DE VARIABLES

Las variables anteriormente mencionadas se resumen en las siguientes, según criterios e intereses de las Gerencias del Banco involucradas en el estudio:

- Índice del Monto total de los productos/Trans (APC): monto total de los productos que tiene la agencia sobre número total de transacciones.
- Eficiencia financiera (EF): costo total por agencia vs. margen financiero, otros ingresos y comisiones.
- Transacciones per cápita (TPC): número total de transacciones vs. Headcount (número de colaboradores del banco por agencia).
- Eficiencia transaccional (ET): capacidad utilizada (min) vs. Capacidad instalada (min).

- Transacciones no core (TNC): número de transacciones no core (pago por servicios, etc.) vs. Transacciones core.
- Índice de Clientes BP y AV (CBPAV): número de clientes de Banca Privada más clientes Alto Valor vs. Total de clientes.
- Índice de Clientes Jurídicos (CJ): número de clientes de Corporativos, Empresariales, Pymes y Otros Jurídicos vs. total de clientes.
- Índice de Clientes resto (CR): número de clientes total sin BP, AV, Corporativos, Empresariales, Pymes y Otros Jurídicos vs. Total de clientes.
- Clientes per cápita (CPC): el resto de clientes vs. La sumatoria de los clientes de privada, alto valor, corporativos y empresariales.
- Índice de Transacciones Jurídicos (TJ): Número de transacciones de Corporativos más Empresariales más Pymes más Otros Jurídicos vs. Total de transacciones.
- Índice de Transacciones de Banca Privada y Alto Valor (TBPAV): Número de transacciones Banca Privada más Clientes Alto Valor vs. Total de transacciones.
- Índice de Montos totales de los clientes de Alto Valor y Pymes (AAVPY): Montos totales de los clientes de Alto Valor y Pymes vs. Total de montos en la agencia.

4.2.2 SELECCIÓN DE VARIABLES

Análisis de los datos: se importan los datos al paquete SPSS, desde el Excel, el nombre de los campos debe ser corto y máximo de 4 dígitos. Una vez que la tabla está abierta se debe cambiar los valores faltantes (missing) de cada variable por el valor 0, a excepción de los campos: eficiencia financiera y eficiencia transaccional. El número de casos que son estudiados corresponden a una muestra de la base total de agencias, debido a la confidencialidad de la información.

En estas variables se procede a identificar los datos atípicos y estadísticas descriptivas que permitan aceptar o rechazar agencias y variables, además se estandarizan las variables (a cada observación se resta la media de la variable y se divide para su desviación estándar, para eliminar las diferencias de unidades cuando se comparan montos con porcentajes). Este proceso se puede observar en el Anexo 8.

En la tabla se muestran algunas medidas descriptivas:

Tabla 6: Estadísticas Descriptivas

Fuente: caso de estudio

Statistics												
	APC	EF	TPC	ET	TNC	CBPAV	CJ	CR	CPC	TJ	TBPAV	AAVY
N	118	118	118	118	118	118	118	118	118	118	118	118
	0	0	0	0	0	0	0	0	0	0	0	0
Mean	445568,8	,9917	2007,9939	,8640	,7194	6,085E-02	2,369E-02	,9155	488,5008	7,681E-02	7,751E-02	,3508
Median	365082,1	,9940	1970,7000	,8536	,6572	4,984E-02	1,944E-02	,9265	471,1790	7,128E-02	7,327E-02	,3615
Std. Deviation	314786,5	,2956	575,5724	,2551	,3406	4,482E-02	1,650E-02	5,841E-02	266,7338	4,106E-02	2,922E-02	,1085
Skewness	2,326	-,599	1,257	2,023	2,253	3,543	3,254	-3,907	1,076	2,023	,860	-,495
Std. Error of Skewness	,223	,223	,223	,223	,223	,223	,223	,223	,223	,223	,223	,223
Kurtosis	8,474	3,611	3,211	10,955	7,305	20,051	20,329	24,477	3,279	8,344	1,561	1,331
Std. Error of Kurtosis	,442	,442	,442	,442	,442	,442	,442	,442	,442	,442	,442	,442
Minimum	,00	,00	990,50	,35	,29	,00	,00	,49	,06	,02	,02	,00
Maximum	2029443	1,96	4431,00	2,43	2,43	,37	,14	1,00	1595,40	,31	,19	,62

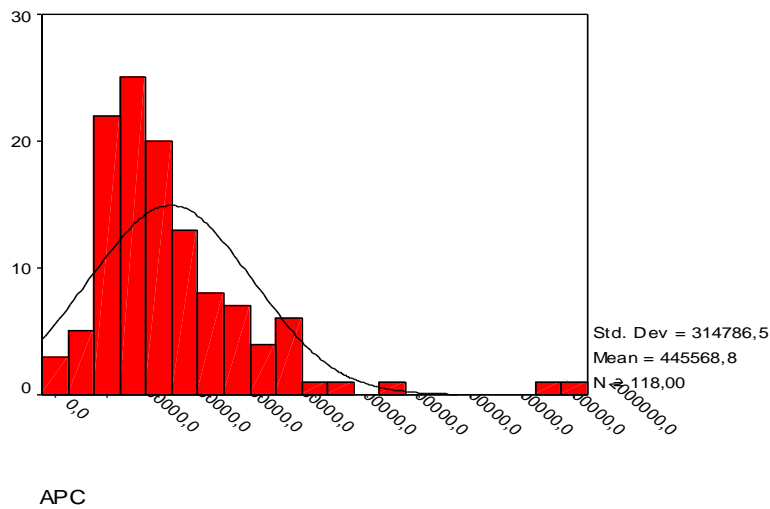
Para la variable montos totales por producto per cápita, se observa que existen datos atípicos, puesto que el valor máximo es superior al valor de $\mu + 3*\sigma$, por tanto las agencias que tienen montos per cápita superiores a 1'389.928 son eliminadas.

El sesgo se observa en aquellas agencias que tienen asignados los montos y transacciones de otras agencias.

En el gráfico se observa que se ratifica este criterio, puesto que existe un sesgo hacia la izquierda respecto a la mediana:

Gráfico 6: Distribución de la variable Montos totales per cápita

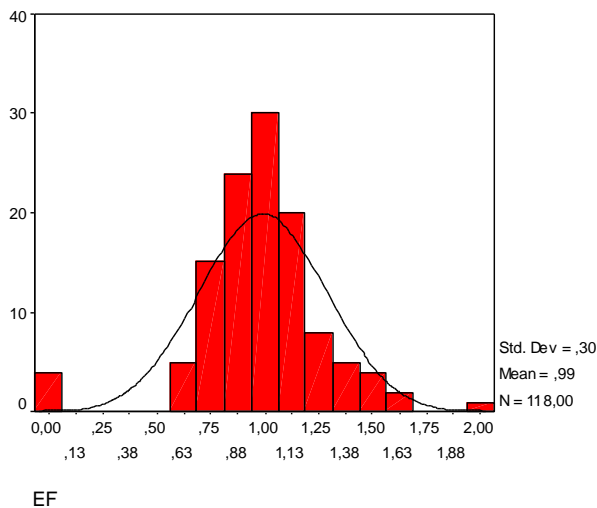
Fuente: caso de estudio



Para la variable eficiencia financiera no se observan comportamientos atípicos.

Gráfico 7: Distribución de la variable eficiencia financiera

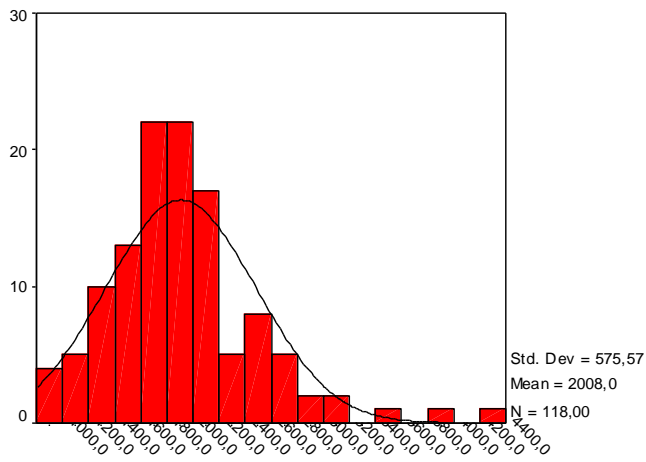
Fuente: caso de estudio



La variable transacciones per cápita no muestra agencias con comportamientos atípicos que determinen su eliminación, puesto que las agencias que muestran altos valores en esta variable no crean sesgos exagerados.

Gráfico 8: Distribución de la variable transacciones per cápita

Fuente: caso de estudio

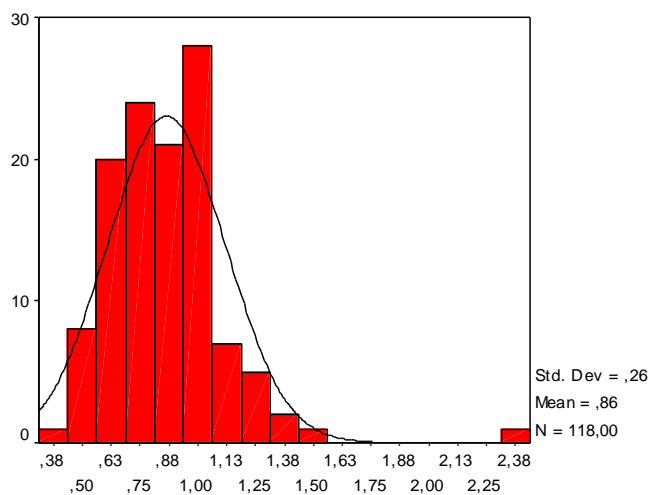


TPC

El comportamiento de las agencias en cuanto a su eficiencia transaccional muestra un sesgo en el valor máximo el cual debe ser eliminado, puesto que el resto de datos muestran una agrupación alrededor de la mediana:

Gráfico 9: Distribución de la variable eficiencia transaccional

Fuente: caso de estudio

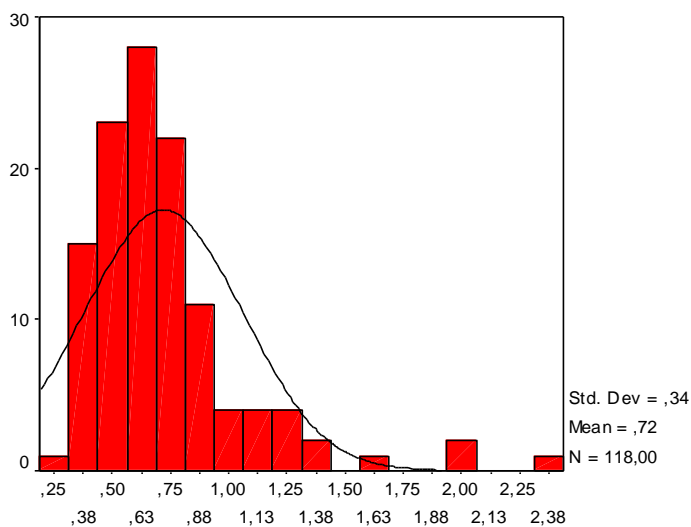


ET

En la variable transacciones no core se elimina también la agencia con el valor máximo puesto que es superior al valor de la media más tres veces la desviación estándar y es el polo superior del gráfico, pero la distribución de los valores de la variable no muestra saltos amplios entre los valores..

Gráfico 10: Distribución de la variable transacciones no core

Fuente: caso de estudio

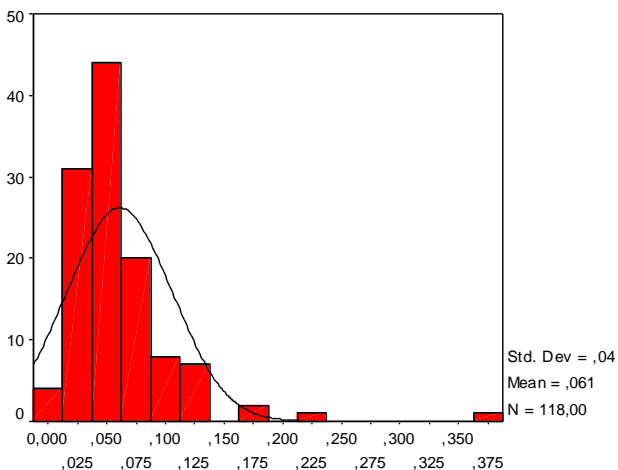


TNC

La variable “clientes Banca Privada y Alto valor” muestra altos sesgos en los valores de las agencias en el extremo superior, por tanto se eliminarán todas aquellas cuyo valor es superior a 0.20. En el gráfico se observa su distribución y el punto extremo:

Gráfico 11: Distribución de la variable clientes Banca Privada y Alto Valor

Fuente: caso de estudio

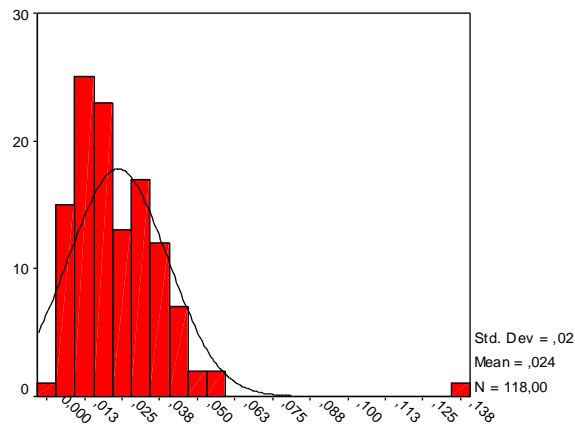


CBPAV

La variable “clientes jurídicos” muestra altos valores en las agencias con el valor del extremo superior, por tanto se eliminará la agencia con el valor máximo. En el gráfico se observa su distribución y el punto extremo:

Gráfico 12: Distribución de la variable clientes jurídicos

Fuente: caso de estudio

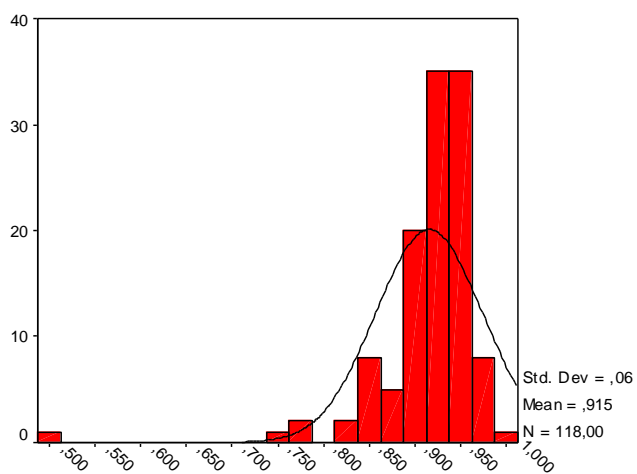


CJ

Para la variable clientes resto se realiza el mismo proceso y se elimina el punto mínimo que se observa., que es atípico, a pesar que la regla de la media menos tres veces la desviación estándar se cumple (igual a 69,52):

Gráfico 13: Distribución de la variable clientes resto

Fuente: caso de estudio

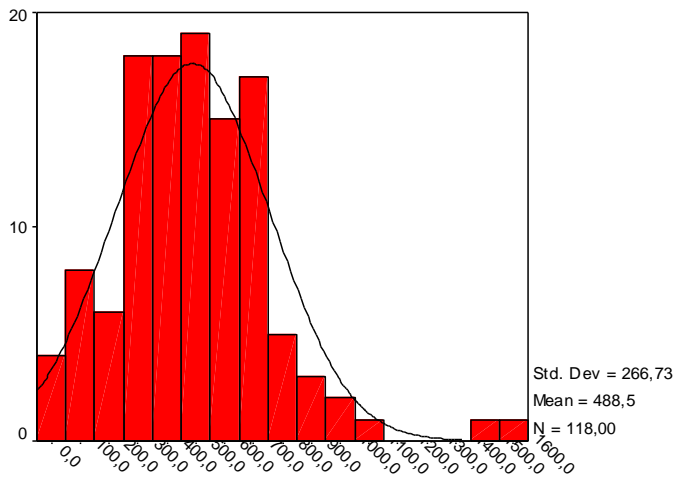


CR

La variable clientes per cápita no muestra datos atípicos.

Gráfico 14: Distribución de la variable clientes per cápita

Fuente: caso de estudio

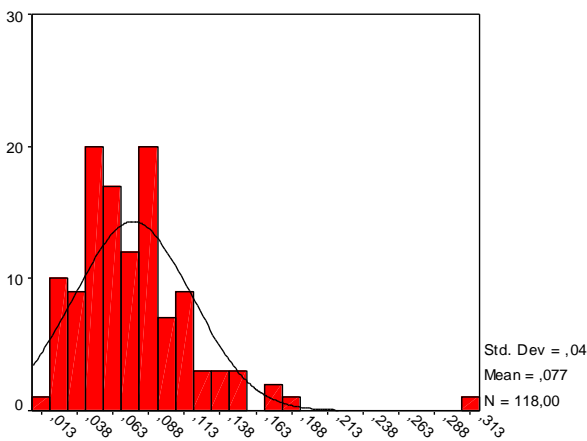


CPC

La variable “transacciones jurídicos” muestra altos valores en las agencias con el valor del extremo superior, por tanto se eliminará la agencia con el valor máximo. En el gráfico se observa su distribución y el punto extremo:

Gráfico 15: Distribución de la variable transacciones jurídicos

Fuente: caso de estudio



TJ

Para las variables transacciones Banca Privada y Alto Valor, montos por Productos de los clientes de Alto Valor y Pymes no se observan datos atípicos y el sesgo es muy pequeño.

Gráfico 16: Distribución de la variable transacciones Banca Privada y Alto Valor

Fuente: caso de estudio

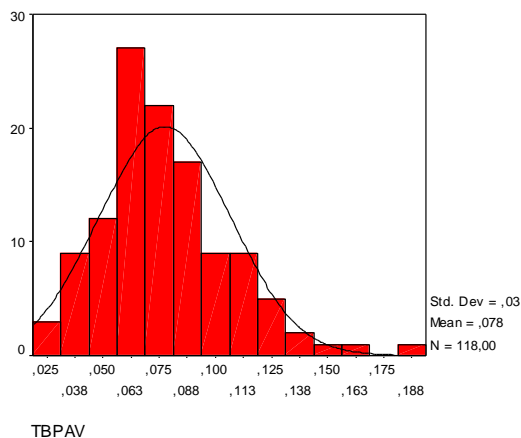
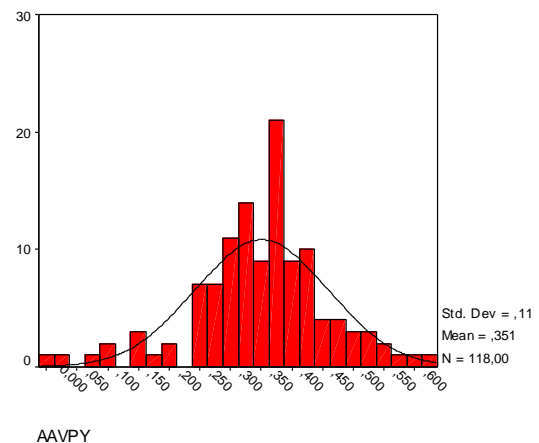


Gráfico 17: Distribución de la variable montos totales de clientes Alto Valor y Pymes

Fuente: caso de estudio



En cada uno de los gráficos se puede observar el ajuste a una distribución Normal, con un mayor o menor sesgo en cada variable, con el método Stepwise se podrá identificar las más convenientes para el análisis. En el Anexo 9 se aprecia el proceso de eliminación de las agencias en el paquete SPSS, según lo definido anteriormente. Se eliminan 6 agencias: la agencia 4, la agencia 8, la agencia 21, la agencia 76, la agencia 100 y la agencia 110.

Por tanto, el número total de agencias seleccionadas para el análisis son 112 agencias.

Partición de la Base de Datos: en esta parte se seleccionan los grupos de entrenamiento, de validación y de test o prueba para el Análisis Discriminante. El grupo de entrenamiento comprende el 60% de la población, es decir 67 agencias, el de validación, el 20% de la población, 23 agencias y el grupo de test o de prueba formado por 22 agencias. Los tres grupos deben ser excluyentes, es decir

que una agencia pertenece a un solo grupo. El proceso de partición de la base de datos se observa en el Anexo 10.

4.2.2.1 Análisis de Correlaciones

El análisis de componentes principales es una técnica para analizar las asociaciones lineales entre las variables. El análisis de correlaciones entre variables determina si es adecuada o no la aplicación del análisis de componentes principales, además si no existe correlación entre ellas la matriz de correlación sería igual a la identidad, y por consiguiente carecería de sentido realizar el análisis.

Observando la matriz de correlaciones entre las variables: Montos per cápita (APC), Eficiencia financiera (EF), transacciones per cápita (TPC), eficiencia transaccional (ET), transacciones no core (TNC), clientes banca privada y alto valor (CBPAV), clientes jurídicos (CJ), resto de clientes (CR), clientes per cápita (CPC), transacciones jurídicos (TJ), transacciones banca privada y alto valor (TBPAV) y montos totales alto valor y pymes (AAVPY) en el gráfico puede comprobarse que, por ejemplo, las variables CBPAV y CR tienen una asociación lineal inversa muy fuerte, mientras que la asociación de una de ellas respecto a una tercera cualquiera es más débil. Sin embargo, en otros casos las relaciones no son tan fuertes y, por tanto, no es fácil determinar subconjuntos de variables similares entre sí, ver anexo 11 para conocer el proceso en el paquete SPSS y los primeros resultados del ACP.

El análisis de componentes principales facilitará la interpretación de la información contenida en la matriz de correlaciones.

En la matriz de correlaciones se observa que la variable Clientes Resto y Clientes Banca Privada y Alto Valor tienen una correlación negativa alta, por lo cual se elimina la variable Clientes Banca Privada y Alto Valor.

Tabla 7: Matriz de correlaciones

Correlation Matrix

	CBPAV	CJ	TBPAV	AAVPY	EF	TPC	ET	TNC	CR	CPC	TJ	APC
Correlation	1,000	,572	,584	,300	-,159	-,287	-,305	-,194	-,967	-,347	,196	,385
CBPAV												
CJ	,572	1,000	,314	,275	-,157	-,244	-,169	,008	-,763	-,244	,390	,313
TBPAV	,584	,314	1,000	,440	-,281	-,219	-,159	-,516	-,558	-,105	,320	,486
AAVPY	,300	,275	,440	1,000	-,162	-,176	-,010	-,029	-,322	-,025	,196	,258
EF	-,159	-,157	-,281	-,162	1,000	,028	,290	-,103	,174	-,142	,070	-,460
TPC	-,287	-,244	-,219	-,176	,028	1,000	,333	,130	,303	,164	-,256	-,181
ET	-,305	-,169	-,159	-,010	,290	,333	1,000	,018	,293	,030	-,122	-,165
TNC	-,194	,008	-,516	-,029	-,103	,130	,018	1,000	,150	-,112	-,337	-,262
CR	-,967	-,763	-,558	-,322	,174	,303	,293	,150	1,000	,349	-,276	-,401
CPC	-,347	-,244	-,105	-,025	-,142	,164	,030	-,112	,349	1,000	-,051	,477
TJ	,196	,390	,320	,196	,070	-,256	-,122	-,337	-,276	-,051	1,000	,191
APC	,385	,313	,486	,258	-,460	-,181	-,165	-,262	-,401	,477	,191	1,000

Al incluir la variable Clientes Banca Privada y Alto Valor y Clientes Resto se genera una matriz no definida positiva y entonces no se obtiene la matriz de covarianza, ni los criterios de Barlett y el KMO. Una vez eliminada la variable Clientes Banca Privada se procede a realizar el ACP correspondiente, considerando los criterios de Barlett y KMO y el proceso Varimax. (Ver proceso en el Anexo 11).

El test de esfericidad de Bartlett permite demostrar que la matriz de correlaciones no es la identidad, por tanto, al analizar los resultados se rechaza esta hipótesis y se puede continuar con el ACP.

El índice de KMO (Kaiser-Meyer-Olkin) debe ser mayor a 0.5, caso contrario no se puede realizar el ACP.

Los resultados se muestran en la siguiente tabla:

Tabla 8: KMO y Bartlett

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,567
Bartlett's Test of Sphericity	Approx. Chi-Square	475,785
	df	55
	Sig.	,000

Se concluye que se puede realizar el ACP con las variables seleccionadas y la matriz de correlaciones no es la identidad.

4.2.2.2 Aplicación del Análisis de Componente Principales (ACP)

El ACP se realiza para un número de factores de 4, de 2 y de un factor, en el Anexo 12 se pueden observar los resultados para 4 y 2 factores. Entre los resultados del ACP, en el paquete SPSS, está el cálculo de la comunalidad, que es la proporción de variabilidad de una variable explicada por el conjunto de los k primeros factores y, en este caso concreto por el primer factor se obtiene:

Tabla 9: Comunalidad de las variables para un factor

Communalities		
	Initial	Extraction
CJ	1,000	,495
TBPAV	1,000	,606
AAVPY	1,000	,265
EF	1,000	,137
TPC	1,000	,227
ET	1,000	,153
TNC	1,000	,141
CPC	1,000	1,269E-02
TJ	1,000	,274
APC	1,000	,408
CR	1,000	,671

Extraction Method: Principal Component Analysis

Como se puede observar la variable que tiene menos proporción de variabilidad explicada por el primer factor es Clientes Per Cápita, y la mayor proporción explicada es Transacciones Banca Privada y Alto Valor. Luego, se

analizan los valores propios que sean superiores a 1 y la varianza explicada, para este caso de un factor, se tiene:

Tabla 10: Valores Propios del ACP con un factor

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,389	30,807	30,807	3,389	30,807	30,807
2	1,714	15,585	46,392			
3	1,401	12,739	59,130			
4	1,110	10,088	69,218			
5	,855	7,771	76,989			
6	,822	7,476	84,465			
7	,627	5,702	90,166			
8	,517	4,702	94,869			
9	,271	2,465	97,333			
10	,178	1,615	98,948			
11	,116	1,052	100,000			

Extraction Method: Principal Component Analysis.

El valor propio es superior a 1 (3,389) y explica el 31% de la varianza. Si se aplica esta metodología a otro caso de Agencias Bancarias se aconseja, si es posible, obtener una varianza mayor. Las puntuaciones factoriales por variable son:

Tabla 11: Puntuaciones factoriales del ACP con un factor

Component Matrix^a

	Component
	1
CJ	,703
TBPAV	,778
AAVPY	,515
EF	-,371
TPC	-,477
ET	-,391
TNC	-,375
CPC	-,113
TJ	,523
APC	,639
CR	-,819

Extraction Method: Principal Component Analysis

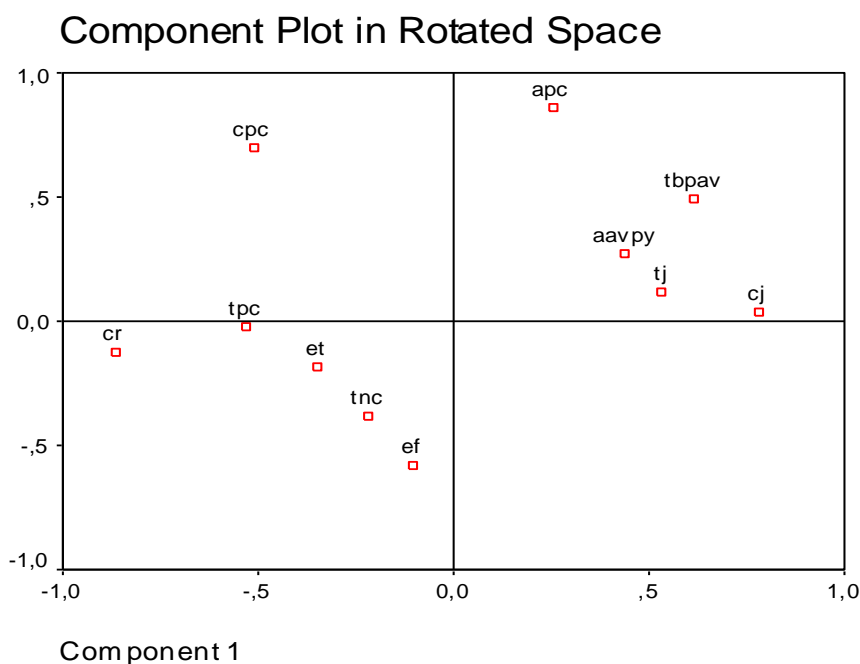
a. 1 components extracted.

En la matriz de componentes están las saturaciones que indican las variables correladas entre sí. La variable con un mayor coeficiente o puntuación

factorial es Clientes Resto. Este factor se puede definir como un factor de definición de tipos de clientes que transaccionan en las agencias, a medida que crece este factor está caracterizado por las variables CJ, TBPAV, AAVPY, TJ, APC y aquellas agencias que tienen altos valores en CR, CPC, TNC, ET, TPC, EF tienen un puntaje menor en el factor. El factor tiene características de tipo de clientes, tipo de transacciones y montos de los productos de las agencias, por tanto es útil para los fines de esta tesis, en la clasificación de agencias bancarias privadas porque cubre varias directrices que diversifican la caracterización de cada grupo. Cuando se realiza la rotación Varimax se obtiene una rotación ortogonal de los factores que siguen siendo incorrelados y se minimiza el número de variables con saturaciones altas en un factor.

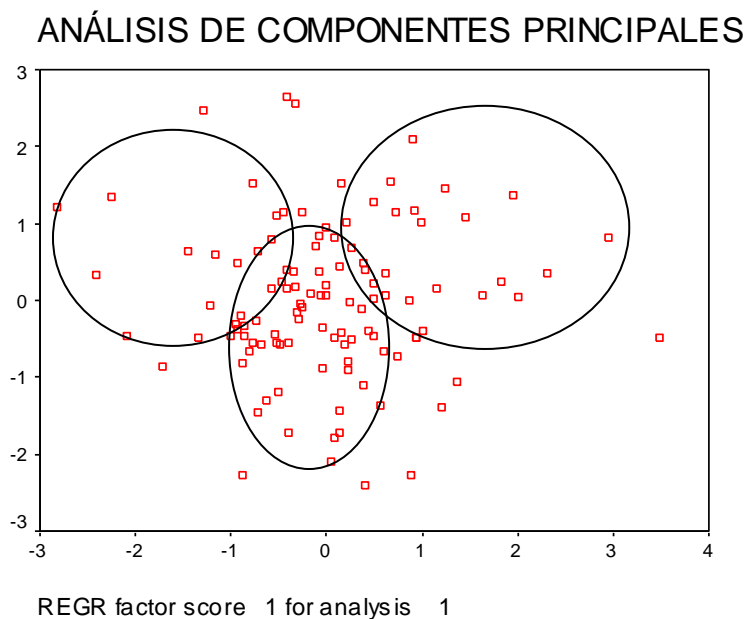
Analizando la representación gráfica de las variables de cada uno de los factores o componentes enunciados puede observarse que sus posiciones son tales que la distancia al origen es grande en todos los casos, y las variables de un mismo grupo están próximas entre sí. La representación de la nube de variables y la nube de puntos se muestran en los gráficos 18 y 19 respectivamente, donde se identifica la ubicación de las agencias respecto a los cuadrantes donde se ubican las variables, y se toman los valores para un ACP con dos factores.

Gráfico 18: Nube de variables para dos factores



La interpretación de una nube de variables depende de la ubicación de las variables respecto al factor o componente 1, mientras más cercano al eje y más lejano al centro las variables influyen o tienen un mayor peso en el factor, como es el caso de las variables CR, correlacionada con el factor negativamente y CJ, correlacionada positivamente.

Gráfico 19: Nube de individuos para dos factores



Las agencias muestran un alto agrupamiento en el centro, donde se ubican las variables ET, TNC, EF, que caracterizan a este grupo. Otro grupo de agencias están caracterizadas por las variables APC, TBPAV, AAVPY, TJ, CJ, es decir aquellas que tienen una tendencia de negocio y menos transaccional, y otro grupo de agencias que tiene como variables dominantes CR, TPC y CPC lo que muestra un comportamiento transaccional.

CAPITULO 5

CLASIFICACIÓN DE LAS AGENCIAS SEGÚN VARIABLES TRANSACCIONALES Y DE NEGOCIO

5.1 ANÁLISIS CLUSTER

El primer componente principal encontrado en el análisis de Componentes Principales es la variable segmentada por el análisis cluster, la misma que permitirá formar los grupos y caracterizarlos por las variables aplicadas en el análisis anterior.

A continuación, se realiza la elección de un procedimiento de Agrupación, es decir, optar por un procedimiento jerárquico (o exploratorio) y después un no jerárquico (llamado k-means). En el clustering, siempre se debe empezar por un procedimiento de clustering exploratorio y, después, otro, confirmatorio para estudiar cuál es el número óptimo de clusters para esta tabla de datos.

El análisis cluster Exploratorio se conoce como Análisis Cluster Jerárquico debido a la creación de relaciones jerárquicas entre las observaciones, de modo que, una vez que dos observaciones entran a formar parte de un cierto cluster, no pueden salir de él. En el Anexo 13 está el proceso de cálculo en el paquete SPSS. Los resultados de las variables de agrupación se guardan en la tabla inicial.

En el dendograma Anexo 13 se observa que existen 3 grupos que diferencian a las agencias unas de otras. A la vista de lo observado en este dendograma, se desprende que existen 3 clusters bien diferenciados. La tabla de pertenencia a los Cluster muestra el proceso de asignación de cada caso a uno de los clusters que se ha solicitado crear. Como se ha optado por crear una solución abierta, se deben ver en el Anexo 13, cada una de las columnas (donde se halla la solución de cada uno de los clusters) y comprobar que el proceso de clustering asigna un número homogéneo y consistente de casos a cada cluster.

El procedimiento de Tablas de Frecuencias permite ver el número de agencias asignados a cada uno de los clusters (Ver Anexo 13) o variables de agrupación generadas.

A partir del número de clusters que hemos considerado óptimo (3) se aplica el método No Jerárquico de Agrupación. En el Anexo 14, se observa el proceso corrido en el paquete SPSS.

Los resultados son los siguientes:

- Los centros finales de los clusters asignan un alto valor al factor o componente principal en el cluster 2, menor en el 3 y más bajo en el 1.

Tabla 12: Centros Finales de Cluster

Final Cluster Centers			
	Cluster		
	1	2	3
REGR factor score 1 for analysis 2	-1,05945	1,75425	,16787

- En la tabla Anova se realiza una prueba de comparación de grupos. Esta prueba consiste en enfrentar los tres grupos respecto de las variables que han intervenido en el proceso de agrupamiento. La hipótesis que pretendemos contrastar en esta prueba es la de que los 3 grupos son significativamente iguales entre sí respecto de la variable y

se rechaza cuando el nivel de significancia es menor a 0.05 y un alto valor de F. En la tabla se observa que se rechaza la hipótesis nula y la variable factor o componente principal es significativa:

Tabla 13: Tabla ANOVA

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
REGR factor score 1 for analysis 2	45,668	2	,180	109	253,153	,000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

- El número de casos por grupo es:

Tabla 14: Número de casos en cada Cluster de K-means

Number of Cases in each Cluster

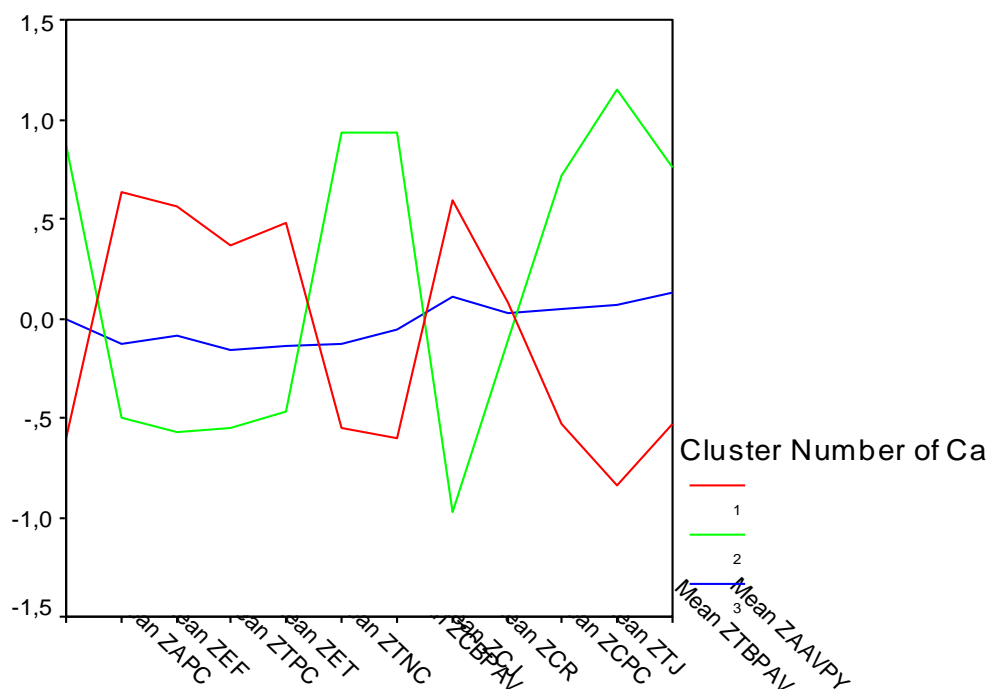
Cluster	1	36,000
	2	16,000
	3	60,000
Valid		112,000
Missing		,000

5.2 CARACTERIZACIÓN DE LOS GRUPOS FORMADOS

Un procedimiento útil para estudiar los perfiles de los clusters es el gráfico de líneas múltiples. Este permite representar de un modo visual los valores medios de los perfiles. En el ANEXO 15, se observa el proceso de generación en el paquete SPSS.

En el gráfico de la línea de puntos se observa:

Gráfico 20: Gráfico de Líneas



El cluster 1 está caracterizado por las variables: eficiencia financiera, índice de las transacciones per cápita, eficiencia transaccional, índice de las transacciones no core, el índice de los clientes resto.

El cluster 2 está caracterizado por: la variable montos totales en las agencias per cápita, el índice de los clientes Banca Privada y Alto Valor, el índice de los clientes jurídicos, el índice de las transacciones de los Clientes Jurídicos, el índice de las Transacciones de los clientes de Banca Privada y Alto Valor, los montos de los clientes de Alto Valor y Pymes y montos per cápita.

El cluster 3 está caracterizado por tener valores medios en todas las variables, entonces son aquellas agencias que no muestran características dominantes.

Por tanto, se ha obtenido de este gráfico una nueva constatación para creer que existen 3 y sólo 3 clusters. El primero de ellos estaría caracterizado por un comportamiento que se lo definirá como transaccional. El segundo lo estaría por un comportamiento de Negocios. El último lo estaría por un comportamiento general de negocios y transaccional.

Hasta ahora se tiene la tabla de frecuencias y un gráfico de líneas para describir los perfiles, entonces se realiza un procedimiento de medias para corroborar estadísticamente la hipótesis de que la división del factor o componente principal es en tres grupos. La variable independiente es la variable cluster y las variables dependientes las originales.

El aporte más importante del comando Means es el cálculo del coeficiente eta y el test de linealidad. Ver anexo 15.2.

Los resultados son:

Tabla 15: La matriz de medias

Cluster Number of Case	APC	EF	TPC	ET	TNC	CBPAV	CJ	CR	CPC	TJ	TBPAV	AAVPY
1	255,713.88	118%	2,332.06	96%	0.88	0.04	0.01	0.95	509.08	0.05	0.05	0.29
2	719,555.89	85%	1,680.46	72%	0.56	0.10	0.04	0.86	460.01	0.11	0.11	0.43
3	443,505.98	95%	1,954.94	82%	0.67	0.06	0.02	0.92	495.36	0.08	0.08	0.36
Total	422,579.94	101%	2,036.94	85%	0.72	0.06	0.02	0.92	494.72	0.07	0.08	0.35

Donde se ratifican los perfiles obtenidos en el gráfico de líneas.

Tabla 16: Tabla ANOVA – test de linealidad

ANOVA Table

Between Groups Linearity

	Sum of Squares	df	Mean Square	F	Sig.
APC * Cluster Number of Case	580553206316,466	1	580553206316,466	15,375	,000
EF * Cluster Number of Case	,989	1	,989	14,991	,000
TPC * Cluster Number of Case	2659517,921	1	2659517,921	9,398	,003
ET * Cluster Number of Case	,342	1	,342	9,496	,003
TNC * Cluster Number of Case	,846	1	,846	7,844	,006
CBPAV * Cluster Number of Case	,005	1	,005	9,336	,003
CJ * Cluster Number of Case	,001	1	,001	14,047	,000
CR * Cluster Number of Case	,011	1	,011	15,502	,000
CPC * Cluster Number of Case	2521,443	1	2521,443	,037	,848
TJ * Cluster Number of Case	,010	1	,010	9,675	,002
TBPAV * Cluster Number of Case	,012	1	,012	30,829	,000
AAVPY * Cluster Number of Case	,093	1	,093	11,371	,001

Todas las variables, excepto “índice de los clientes per cápita”, tienen un nivel de significancia menor a 0.05, lo que indica que contribuyen de modo significativo al proceso de agrupación.

5.3 DEFINICIÓN DE LOS MODELOS DE ATENCIÓN AL CLIENTE

Los modelos de atención responden a la caracterización de los perfiles definidos anteriormente y a las necesidades de cada tipo de grupo:

Las agencias del grupo 1 están más enfocadas a la transaccionalidad y a la atención de “clientes resto” (es decir, clientes no corporativos, empresariales, banca privada, alto valor), por tanto requieren un mayor número de ventanillas para depósitos y retiros de cuentas de los clientes, pagos por servicios públicos, matriculación universidades e instituciones educativas, cobro de cheques o depósitos realizadas por no clientes del banco, pago de roles de pago, depósitos de mensajeros de empresas y un mínimo número de balcones de servicios para entrega de chequeras, de certificados de estados de cuenta, de certificados bancarios, de cheques de gerencia, etc, además un mínimo número de asesores para venta de productos (puede omitirse en este tipo de agencias) o para actualización de datos.

Las agencias del grupo 2, son aquellas enfocadas al negocio, atienden a clientes más rentables para el banco (no se descarta que algunos clientes rentables realicen sus transacciones en las agencias del grupo 1), requieren de un mayor número de asesores para venta de productos y relaciones uno a uno con los clientes para manejar sus cuentas o asesoría bancaria/financiera, un menor número de ventanillas para transacciones no core, y balcones de servicios, puesto que el tipo de clientes en esta agencia manejan una mejor relación financiera con el banco.

Las agencias del grupo 3, son aquellas que atienden a todo tipo de clientes y tienen un considerable número de transacciones de todo tipo, (en menor medida que del tipo uno), por tanto deben tener ventanillas, asesores bancarios y balcones de servicios en un número que responda a sus necesidades, pero intermedio entre los del grupo uno y dos. Aquí se puede aplicar teoría de colas para identificar el número de cajeros o ventanillas óptimo para cada agencia o grupo.

En cuanto, al organigrama de cada agencia, los jefes de negocio y de servicios no siempre aparecerán, se reducen los ayudantes y dependiendo de cada caso aparecen técnicos de sistemas para mantenimiento del software transaccional y hardware en las ventanillas o cajas.

La plataforma tecnológica será más amplia en las agencias del grupo 1, para agilizar las transacciones, que son en mayor número y deben involucrar un menor tiempo de ejecución, y de interacción con otros sistemas, como SRI, ANDINATEL, universidades, etc. En el grupo 2, la tecnología debe responder a un mayor desarrollo en sistemas de seguimiento del comportamiento del cliente; el asesor requiere conocer a su cliente tanto en rentabilidad, número de productos que tiene en el banco, también el perfil demográfico, uso de canales, segmento al que pertenece según la segmentación interna del banco, criterios de deserción, si el cliente a disminuido su relación cuantitativamente, índices de lealtad, si es o no

un cliente potencial, etc. Para las agencias del grupo 3 se deben utilizar las características de los grupo uno y dos, tecnológicamente hablando.

En cuanto a la infraestructura, se aumentarán ventanillas, si son del grupo 1; disminuirán en el grupo 2, e inversamente para balcones de servicios y puestos para los asesores y en el grupo 3 tendrá ventanillas y balcones de servicios y puestos para asesores comerciales. La creación de una caja o ventanilla es menos costosa que crear puestos para asesores comerciales y balcones de servicios, tanto por la infraestructura como por la capacitación del colaborador. Un colaborador de ventanilla recibe capacitación menos costosa y cuenta con un espacio reducido, a comparación de un asesor comercial que requieren mayor espacio, materiales de oficina, impresoras y mayor conocimiento de los productos del banco y recibe capacitación externa. Las agencias del grupo 1, requieren de amplios espacios, en un solo piso, con anuncios informativos generales sobre las transacciones y menos decoración, pero las agencias del grupo 2 requieren mayor decoración, anuncios puntuales para cada tipo de cliente, espacios mejor distribuidos, y ubicación estratégica. Para el grupo 3, se debe mezclar los dos tipos de anuncios y su ubicación debe ser también estratégica.

Respecto a la ubicación geográfica de las agencias, que puede ser tema para otra tesis, el objetivo sería identificar los puntos óptimos de ubicación para la apertura de agencias según la ubicación geográfica de los clientes.

5.4 PRONÓSTICO DEL COMPORTAMIENTO POR AGENCIA

5.4.1 ANÁLISIS DISCRIMINANTE

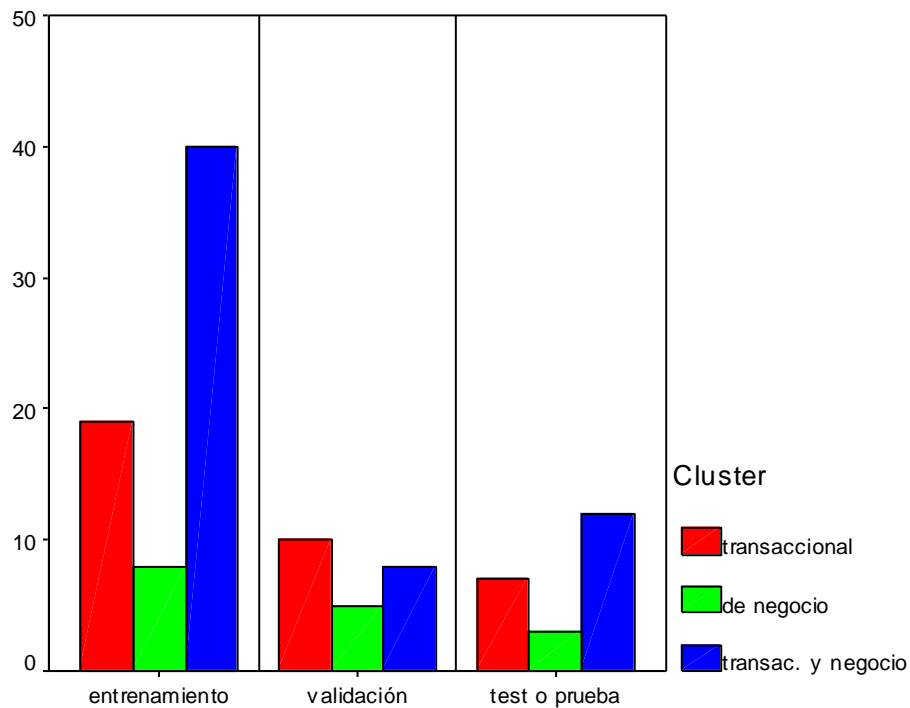
El análisis cluster creó una variable de clasificación a partir del cálculo de las distancias óptimas entre los sujetos. Una vez obtenida la variable clasificadora o de segmentación, llega el momento de utilizar esa variable para reclasificar los grupos, lo que se conoce como técnica discriminante o Análisis Discriminante.

El análisis discriminante no crea grupos de semejanza, sino que utiliza los ya existentes para ofrecer un análisis de la capacidad de las variables del análisis discriminante para clasificar o diferenciar entre las n agencias de los k grupos y se obtiene un esquema de clasificación. El proceso de generación del análisis discriminante en el SPSS se puede observar en el Anexo 16.

En la tabla que se muestra a continuación se observa la distribución de los grupos en cada cluster:

Tabla 17: Distribución de los grupos por cluster

			GRUPOS			Total
			entrenamiento	validación	test o prueba	
Cluster Number of Case	transaccional	Count	19	10	7	36
		% within Cluster Number of Case	52,8%	27,8%	19,4%	100,0%
		% within GRUPOS	28,4%	43,5%	31,8%	32,1%
		% of Total	17,0%	8,9%	6,3%	32,1%
	de negocio	Count	8	5	3	16
		% within Cluster Number of Case	50,0%	31,3%	18,8%	100,0%
		% within GRUPOS	11,9%	21,7%	13,6%	14,3%
		% of Total	7,1%	4,5%	2,7%	14,3%
	transaccional y de negocio	Count	40	8	12	60
		% within Cluster Number of Case	66,7%	13,3%	20,0%	100,0%
		% within GRUPOS	59,7%	34,8%	54,5%	53,6%
		% of Total	35,7%	7,1%	10,7%	53,6%
Total	Count	67	23	22	112	
	% within Cluster Number of Case	59,8%	20,5%	19,6%	100,0%	
	% within GRUPOS	100,0%	100,0%	100,0%	100,0%	
	% of Total	59,8%	20,5%	19,6%	100,0%	

Gráfico 21: Distribución por cluster y grupos

En la interpretación de los datos de salida, luego de corrido el proceso, se observan las siguientes conclusiones:

- Una tabla de resumen del procesamiento de casos

El número de casos a ser analizados son 67, y son aquellos que forman parte del grupo de entrenamiento, el resto, 45, pertenecen al grupo de validación y de prueba.

Tabla 18: Resumen de casos

Unweighted Cases	N	Percent
Valid	67	59,8
Excluded		
Missing or out-of-range group codes	0	,0
At least one missing discriminating variable	0	,0
Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0
Unselected	45	40,2
Total	45	40,2
Total	112	100,0

- Se informa del número de casos participantes y excluidos. Ver anexo 16

El test de igualdad de medias indica que las medidas de las variables en cada grupo del análisis son diferentes para los tres grupos, excepto para las variables transacciones no core, clientes per cápita, puesto que no se rechaza la hipótesis.

Tabla 19: Test de igualdad de medias

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
APC	,694	14,102	2	64	,000
EF	,817	7,181	2	64	,002
TPC	,816	7,206	2	64	,002
ET	,895	3,738	2	64	,029
TNC	,941	2,009	2	64	,143
CBPAV	,581	23,046	2	64	,000
CJ	,580	23,175	2	64	,000
CR	,454	38,416	2	64	,000
CPC	,973	,874	2	64	,422
TJ	,796	8,196	2	64	,001
TBPAV	,615	20,037	2	64	,000
AAVPY	,898	3,639	2	64	,032

- La matriz de covarianza (ver anexo 16) permite ver el grado de cohesión existente dentro de los grupos de análisis. Se observan medias y bajas correlaciones. No se puede interpretar como falta de colinealidad entre las variables. Ver anexo 16
- La tabla de resultados del Test de la M de Box se concluye que la varianzas son significativamente iguales, pues la significación del valor F es menor que 0,05. Ver anexo 16.

Tabla 20: Test M de Box**Test Results**

Box's M		63,979
F	Approx.	2,682
	df1	20
	df2	1623,701
	Sig.	,000

Tests null hypothesis of equal population covariance matrices:

En la tabla de variables introducidas/Extraídas se observa la minimización de Lambda de Wilks global y la prueba exacta de F. Se aprecia que las variables Clientes Resto, montos totales per cápita, las transacciones de jurídicos y la eficiencia financiera superan las pruebas de significatividad.

Tabla 21: Método Stepwise**Variables Entered/Removed^{a,b,c,d}**

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	CR	,454	1	2	64,000	38,416	2	64,000	,000
2	APC	,362	2	2	64,000	20,853	4	126,000	,000
3	TJ	,299	3	2	64,000	17,131	6	124,000	,000
4	EF	,261	4	2	64,000	14,589	8	122,000	,000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

- Maximum number of steps is 24.
- Minimum partial F to enter is 3.84.
- Maximum partial F to remove is 2.71.
- F level, tolerance, or VIN insufficient for further computation.

- La tabla sobre el Análisis de las variables ofrece información de las variables independientes ya incluidas en el modelo paso a paso, en virtud de sus significatividad y el menor Lambda de Wilks, así se selecciona inicialmente la variable Clientes resto, seguida por la variable montos totales per cápita, transacciones jurídicos y eficiencia financiera.

Tabla 22: Variables incluidas en el análisis

Step		Tolerance	F to Remove	Wilks' Lambda
1	CR	1,000	38,416	
2	CR	,994	28,896	,694
	APC	,994	8,042	,454
3	CR	,978	26,164	,551
	APC	,969	8,818	,384
	TJ	,962	6,539	,362
4	CR	,953	27,099	,493
	APC	,859	4,248	,298
	TJ	,910	7,973	,329
	EF	,802	4,411	,299

- Se extrae inicialmente la variable clientes resto porque tienen el menor valor de Lambda de Wilks y mayor valor de F de entrada y significativo. La tolerance (tolerancia) es el porcentaje de la variabilidad de una variable independiente no explicada por las otras y si la tolerancia es igual a la mínima tolerancia se la puede incluir en el modelo como sucede con el resto de variables pero no cumplen los otros criterios de selección.
- Ahora se contrasta la significatividad de las variables independientes dentro del modelo con el estadístico lambda de Wilks y para este modelo son muy significativas las variables incluidas.

Tabla 23: Análisis de la significatividad de las variables independientes del Modelo

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	,454	1	2	64	38,416	2	64,000	,000
2	2	,362	2	2	64	20,853	4	126,000	3,427E-13
3	3	,299	3	2	64	17,131	6	124,000	4,508E-14
4	4	,261	4	2	64	14,589	8	122,000	4,097E-14

- La tabla de comparaciones de grupos emparejados se confrontan por pares las cargas explicativas del modelo en cada paso para los grupos de las variables dependientes.
- El modelo discriminante conseguido tendrá tanta mayor fuerza, cuanto mayor sean los valores de F y mayor su significatividad. En este caso es un nuevo avance en ganancia de la calidad del modelo discriminante.

Tabla 24: Tabla de comparación de grupos emparejados

Pairwise Group Comparison^{a,b,c,d}

Step	Cluster	Number of Case		1,00	2,00	3,00
1	1,00	F			74,763	26,636
		Sig.			,000	,000
	2,00	F	74,763			32,450
		Sig.	,000			,000
	3,00	F	26,636	32,450		
		Sig.	,000	,000		
2	1,00	F			54,170	18,657
		Sig.			,000	,000
	2,00	F	54,170			24,045
		Sig.	,000			,000
	3,00	F	18,657	24,045		
		Sig.	,000	,000		
3	1,00	F			47,352	16,064
		Sig.			,000	,000
	2,00	F	47,352			21,220
		Sig.	,000			,000
	3,00	F	16,064	21,220		
		Sig.	,000	,000		
4	1,00	F			38,656	15,769
		Sig.			,000	,000
	2,00	F	38,656			16,113
		Sig.	,000			,000
	3,00	F	15,769	16,113		
		Sig.	,000	,000		

a. 1, 64 degrees of freedom for step 1.

b. 2, 63 degrees of freedom for step 2.

c. 3, 62 degrees of freedom for step 3.

d. 4, 61 degrees of freedom for step 4.

Como una de las partes importantes del resumen de Funciones Discriminantes Canónicas, se ofrece la Tabla de Valores Propios, que informa de

la calidad discriminante de las dos funciones discriminantes (% de varianza) y que suman entre ambas el 100%. Una de ellas es más significativa que la otra (98% frente a 2%). La correlación canónica es la proporción de la variabilidad total explicada por las diferencias entre los grupos.

Tabla 25: Eigenvalues o Valores Propios

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2,636 ^a	98,0	98,0	,851
2	,053 ^a	2,0	100,0	,224

a. First 2 canonical discriminant functions were used in the analysis.

- La tabla de lambda de Wilks es un test de la calidad de la función discriminante. En este caso se rechaza la hipótesis de que las agencias analizadas reaccionan por igual a las variables dependientes.

Tabla 26: Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,261	83,907	8	,000
2	,950	3,225	3	,358

- La tabla de los coeficientes estandarizados canónicos de las Funciones Discriminantes informa del grado de importancia de las variables independientes, así “montos totales per cápita” y “eficiencia financiera” caracterizan positivamente a la función discriminante 2; “clientes resto”, positivamente, y “transacciones jurídicos”, negativamente, a la función discriminante 1.

Tabla 27: Coeficientes Estandarizados Canónicos de las Funciones Discriminantes**Standardized Canonical Discriminant Function Coefficients**

	Function	
	1	2
APC	-,414	,602
EF	,376	1,045
CR	,825	-,055
TJ	-,561	-,010

- La Matriz Estructural muestra los coeficientes de correlación de las variables independientes con cada una de las dos funciones discriminantes.
- Los coeficientes de las variables en cada función discriminante se obtiene de la tabla de coeficientes. En este caso se observa que la variable montos totales per cápita tiene una influencia baja en las dos funciones, la variable eficiencia financiera tiene un mayor peso en la función discriminante 2, la variable clientes resto tiene mayor importancia en la función 1, la variable transacciones jurídicos tiene un mayor peso en la función discriminante 1, ratificando lo anteriormente mencionado.

Tabla 28: Coeficientes canónicos**Canonical Discriminant Function Coefficients**

	Function	
	1	2
APC	,000	,000
EF	1,606	4,460
CR	37,105	-2,451
TJ	-17,367	-,300
(Constant)	-33,773	-3,419

Unstandardized coefficients

$$D_1 = -33.77 + 0,00001 * APC + 1,606 * EF + 37,105 * CR - 17,367TJ$$

$$D_2 = -3.419 + 0,00001 * APC + 4,460 * EF - 2,451 * CR - 0,300TJ$$

- La tabla de funciones en los centroides de los grupos muestra las coordenadas de la proyección del centroide de cada grupo sobre la función discriminante.

Tabla 29: Funciones en los centroides de los grupos

Functions at Group Centroids

Cluster Number of Case	Function	
	1	2
1,00	1,972	,223
2,00	-3,394	,376
3,00	-,258	-,181

Unstandardized canonical discriminant functions evaluated at group means

- La probabilidad a priori aplicada a cada grupo es que son equiprobables, es decir que tienen la misma probabilidad de que una agencia pertenezca a un grupo.

Tabla 30: Probabilidades a priori

Prior Probabilities for Groups

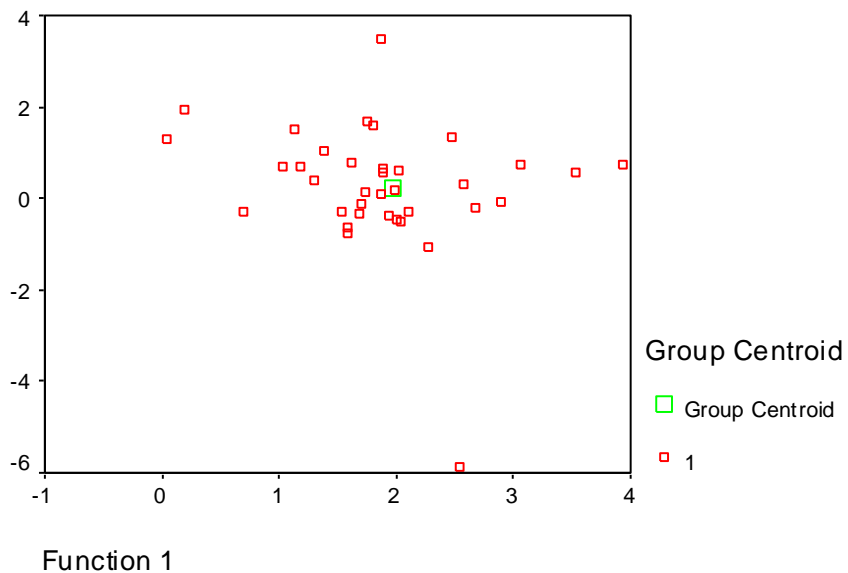
Cluster Number of Case	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1,00	,333	19	19,000
2,00	,333	8	8,000
3,00	,333	40	40,000
Total	1,000	67	67,000

- Las funciones discriminantes canónicas se pueden observar para cada cluster:

Gráfico 22: Cluster 1

Canonical Discriminant Functions

Cluster Number of Case = 1

**Gráfico 23: Cluster 2**

Canonical Discriminant Functions

Cluster Number of Case = 2

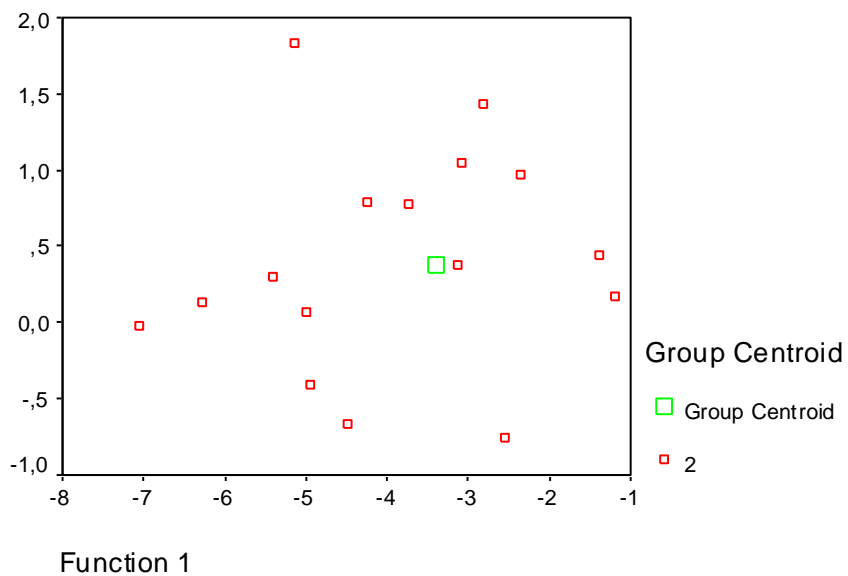


Gráfico 24: Cluster 3

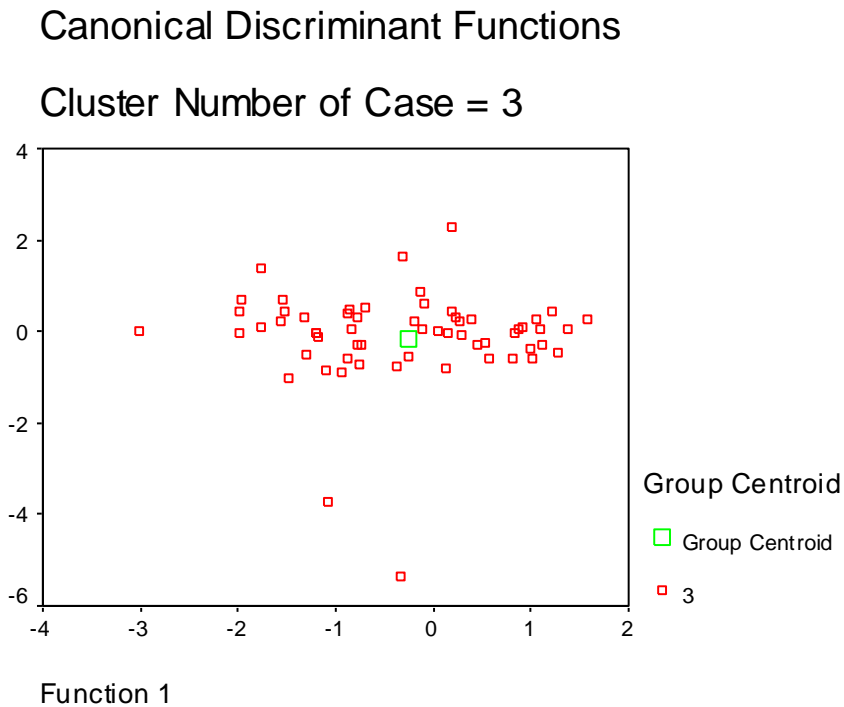
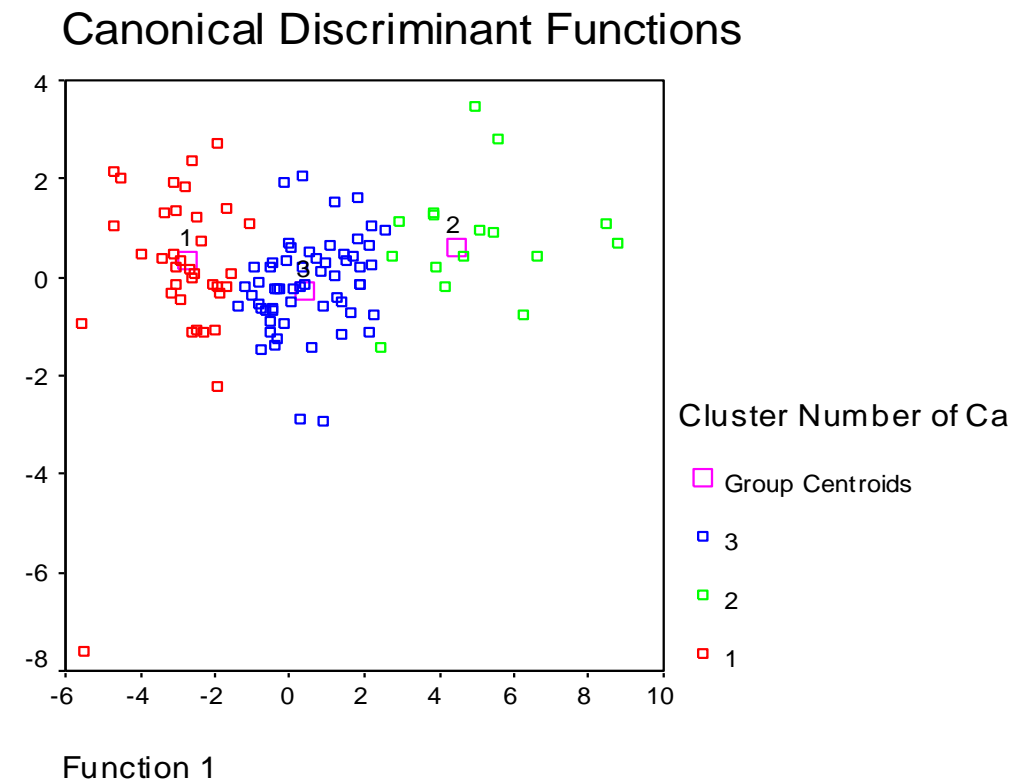


Gráfico 25: Funciones Discriminantes



Aquí se observa la distribución de las agencias según el grupo al que pertenecen. En el cluster1 están las agencias centradas pero en el cluster 2 las agencias son más dispersas, y en el cluster 3 también están centradas, lo que mejora la separación de los grupos.

La Matriz de Confusión muestra el número de casos de cada grupo o cluster. Se incluyen también tres columnas que indican el número de casos predichos de cada grupo. La celda en la que se cruza cada especie consigo misma es la más importante, pues si todos los casos observados se predicen correctamente, es una muestra de una buena predicción, como se puede observar en este caso:

Tabla 31: Clasificación de las agencias

Classification Results ^{b,c,d}							
Cluster Number of Case			Predicted Group Membership			Total	
			1,00	2,00	3,00		
Cases Selected	Original	Count	1,00	19	0	0	19
			2,00	0	7	1	8
			3,00	9	3	28	40
	%	1,00	100,0	,0	,0	100,0	
		2,00	,0	87,5	12,5	100,0	
		3,00	22,5	7,5	70,0	100,0	
	Cross-validated ^a	Count	1,00	19	0	0	19
			2,00	0	7	1	8
3,00			9	4	27	40	
%		1,00	100,0	,0	,0	100,0	
		2,00	,0	87,5	12,5	100,0	
		3,00	22,5	10,0	67,5	100,0	
Cases Not Selected	Original	Count	1,00	14	0	3	17
			2,00	0	7	1	8
			3,00	2	2	16	20
	%	1,00	82,4	,0	17,6	100,0	
		2,00	,0	87,5	12,5	100,0	
		3,00	10,0	10,0	80,0	100,0	

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 80,6% of selected original grouped cases correctly classified.

c. 82,2% of unselected original grouped cases correctly classified.

d. 79,1% of selected cross-validated grouped cases correctly classified.

Se observa que 9 agencias inicialmente ubicadas en el cluster 3 son clasificados ahora en el cluster 1, y 3 agencias que anteriormente pertenecían al

cluster 3 pertenecen ahora al cluster 2, en el cluster 3 sólo existe un cliente que era del cluster 2.

Continuando con el proceso de entrenamiento de las funciones discriminantes, se aplica lo anteriormente analizado con las variables fijas para el grupo de validación, se crea una variable que identifique al grupo de entrenamiento y de validación. El total de agencias es 90:

Tabla 32: Número de casos, Entrenamiento y Validación

Unweighted Cases	N	Percent
Valid	90	80,4
Excluded		
Missing or out-of-range group codes	0	,0
At least one missing discriminating variable	0	,0
Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0
Unselected	22	19,6
Total	22	19,6
Total	112	100,0

La matriz de Confusión del grupo de validación y de entrenamiento es:

Tabla 33: Matriz de Confusión

			Predicted Group Membership			Total
			1,00	2,00	3,00	
Cases Selected	Original	Count	1,00	2,00	3,00	
			29	0	0	29
			0	13	0	13
			3	2	43	48
		%	1,00	2,00	3,00	
			100,0	,0	,0	100,0
			,0	100,0	,0	100,0
			6,3	4,2	89,6	100,0
Cases Not Selected	Original	Count	1,00	2,00	3,00	
			27	0	2	29
			0	11	2	13
			4	4	40	48
		%	1,00	2,00	3,00	
			93,1	,0	6,9	100,0
			,0	84,6	15,4	100,0
			8,3	8,3	83,3	100,0
Cases Selected	Cross-validated ^a	Count	1,00	2,00	3,00	
			7	0	0	7
			0	3	0	3
			0	1	11	12
		%	1,00	2,00	3,00	
			100,0	,0	,0	100,0
			,0	100,0	,0	100,0
			,0	8,3	91,7	100,0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 94,4% of selected original grouped cases correctly classified.

c. 95,5% of unselected original grouped cases correctly classified.

d. 86,7% of selected cross-validated grouped cases correctly classified.

Las agencias están claramente ubicadas en los respectivos cluster, obteniendo un error del 13,3% en la clasificación total, lo cual ratifica la consistencia de la clasificación de las agencias.

5.5 RESULTADOS

Como se puede observar en la tabla 33, se tiene un nivel de acierto del 95,5%, para el grupo de test o prueba, cuyo error se debe a la mala clasificación de tres clientes del cluster 3 (inicialmente) que pasan en el pronóstico al cluster 1 y dos clientes del cluster 3 pasan al cluster 2. Nuevamente se observa que el nivel de acierto es alto. La varianza total explicada por la primera función discriminante es del 97%, que es alta.

La matriz de confusión para la ejecución del modelo según las variables seleccionadas de los grupos de entrenamiento y validación frente al grupo de test o prueba es:

Tabla 34: Matriz de confusión del grupo de test o prueba

Classification Results^d

Cross-validated^a

		Cluster Number of Case	Predicted Group Membership			Total
			1,00	2,00	3,00	
Cases Selected	Count	1,00	5	0	2	7
		2,00	0	2	1	3
		3,00	0	0	12	12
	%	1,00	71,4	,0	28,6	100,0
		2,00	,0	66,7	33,3	100,0
		3,00	,0	,0	100,0	100,0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

d. 86,4% of selected cross-validated grouped cases correctly classified.

En la tabla se identifica que el error cometido en los grupos de entrenamiento y validación al aplicar las funciones discriminantes, del 20%, mientras que el error cometido al aplicar las funciones discriminantes del grupo de entrenamiento al grupo de test o prueba es del 13,6, siendo un error aceptable.

Respecto a los costos por cada tipo de error, el más costoso sería crear una agencia con estructura para negocio pero el comportamiento sus clientes sea transaccional, es decir que sea del cluster 1 y reciba un tratamiento del cluster 2. En la matriz de confusión para las agencias del grupo de test o prueba se comete un error no costoso cuando a las agencias con comportamiento de negocio son clasificadas en el cluster tanto transaccional como de negocio. Así mismo, dos agencias del grupo transaccional son clasificadas en el grupo tanto transaccional como de negocio, por tanto en ninguno de los dos casos se produce un error grave.

Los resultados fueron sometidos a validación por los expertos en negocios y del canal físico del banco ratificando la clasificación de las agencias, conceptualizadas de la siguiente forma:

Gráfico 26: Clasificación por cluster según comportamiento transaccional y de negocio

transaccional	ALTO	TRANSACCIONAL (cluster 1)	TRANSACCIONAL Y DE NEGOCIO (cluster 3)
	BAJO	TRANSACCIONAL Y DE NEGOCIO (cluster 3)	NEGOCIO (cluster 2)
		BAJO	ALTO
		negocio	

Para complementar los resultados del análisis de la clasificación de las agencias se estudia la distribución geográfica de las agencias.

Tabla 35: Distribución geográfica de las agencias por grupo**GRUPOS * REGIONAL Crosstabulation**

Count

		GRUPOS			Total
		entrenamiento	validación	test o prueba	
REGIONAL	COSTA CENTRO	3	4		7
	COSTA NORTE	11	8	1	20
	CUENCA	6		1	7
	GUAYAQUIL	11	5	5	21
	MACHALA	3	1	2	6
	ORIENTE	3		2	5
	QUITO	13	3	7	23
	SIERRA CENTRO	10	1	1	12
	SIERRA NORTE	7	1	3	11
Total		67	23	22	112

Como se puede observar en el grupo de test o prueba no existen agencias de la regional Costa Centro. Las regionales con mayor representatividad son Quito, Guayaquil y Costa Norte.

La distribución de las agencias según el segmento al que pertenecen, transaccional, transaccional y negocio o únicamente negocio se muestra en la tabla 36.

Tabla 36: Agencias por segmento y regional

REGIONAL * Cluster Number of Case Crosstabulation

			Cluster Number of Case			Total
			transaccional	negocio	transaccional y negocio	
REGIONAL	COSTA CENTRO	Count	4		3	7
		% within REGIONAL	57,1%		42,9%	100,0%
	COSTA NORTE	Count	11	1	8	20
		% within REGIONAL	55,0%	5,0%	40,0%	100,0%
	CUENCA	Count			7	7
		% within REGIONAL			100,0%	100,0%
	GUAYAQUIL	Count	5	3	13	21
		% within REGIONAL	23,8%	14,3%	61,9%	100,0%
	MACHALA	Count	4		2	6
		% within REGIONAL	66,7%		33,3%	100,0%
	ORIENTE	Count	2		3	5
		% within REGIONAL	40,0%		60,0%	100,0%
	QUITO	Count	2	10	11	23
		% within REGIONAL	8,7%	43,5%	47,8%	100,0%
	SIERRA CENTRO	Count	5	2	5	12
		% within REGIONAL	41,7%	16,7%	41,7%	100,0%
	SIERRA NORTE	Count	3		8	11
		% within REGIONAL	27,3%		72,7%	100,0%
Total		Count	36	16	60	112
		% within REGIONAL	32,1%	14,3%	53,6%	100,0%

El mayor porcentaje de agencias transaccionales se observa en la regional Costa Centro, seguida por Costa Norte, Machala. Las agencias de las regionales Cuenca, Guayaquil y Sierra Norte muestran altas agrupaciones en las agencias con características transaccional y de negocio.

El mayor número de agencias enfocadas al servicio del negocio está en Quito. La distribución de las agencias por cada segmento según la regional, es decir la relación por columnas se muestra en la tabla 37:

Tabla 37: Regional por segmento

REGIONAL * Cluster Number of Case Crosstabulation

			Cluster Number of Case			Total
			transaccional	negocio	transaccional y negocio	
REGIONAL	COSTA CENTRO	Count	4		3	7
		% within Cluster Number of Case	11,1%		5,0%	6,3%
	COSTA NORTE	Count	11	1	8	20
		% within Cluster Number of Case	30,6%	6,3%	13,3%	17,9%
	CUENCA	Count			7	7
		% within Cluster Number of Case			11,7%	6,3%
	GUAYAQUIL	Count	5	3	13	21
		% within Cluster Number of Case	13,9%	18,8%	21,7%	18,8%
	MACHALA	Count	4		2	6
		% within Cluster Number of Case	11,1%		3,3%	5,4%
	ORIENTE	Count	2		3	5
% within Cluster Number of Case		5,6%		5,0%	4,5%	
QUITO	Count	2	10	11	23	
	% within Cluster Number of Case	5,6%	62,5%	18,3%	20,5%	
SIERRA CENTRO	Count	5	2	5	12	
	% within Cluster Number of Case	13,9%	12,5%	8,3%	10,7%	
SIERRA NORTE	Count	3		8	11	
	% within Cluster Number of Case	8,3%		13,3%	9,8%	
Total		Count	36	16	60	112
		% within Cluster Number of Case	100,0%	100,0%	100,0%	100,0%

La mayor concentración de las agencias en el segmento transaccional se observa en la regional Costa Norte. Así mismo, la mayor concentración del segmento de negocio está en Quito. En el segmento transaccional y de negocio la mayor concentración se observa en Guayaquil, seguido por Quito.

Un análisis complementario que ratifica la caracterización de las agencias está relacionado con los activos fijos que tiene cada agencia según el grupo al que pertenece. En la tabla 38, se observan algunos resultados que permiten concluir que las agencias que pertenecen al segmento de negocio tienen activos per cápita en mayor monto que el resto de segmentos, seguido por el caracterizado por un comportamiento de negocio y transaccional y finalmente por el segmento de agencias con enfoque transaccional.

Tabla 38: Activos fijos per cápita en dólares por segmento

SEGMENTO	Activos per cápita (\$)
Transaccional	35.248,35
Negocio	108.908,18
Transaccional y Negocio	63.202,90
Total general	60.746,83

Fuente: caso de estudio 2003

Comparando estos resultados con el dato de línea de base dado para el costo o activos fijos (60.746 USD), existe una mejor distribución de los recursos dependiendo del tipo de agencia que se desea crear, es decir que crear una agencia únicamente transaccional tiene menos costo que crear una agencia enfocada al negocio. Por tanto, la aplicación de la clasificación de las agencias permite distribuir la inversión del banco en cada agencia en forma correcta y acorde a las necesidades de transaccionalidad y de negocio que tenga la agencia.

En cuanto a las agencias transaccionales, el banco tendría un ahorro de 917.945 dólares que pueden ser invertidos en mejorar la atención de los clientes en las agencias clasificadas como negocio y –negocio y transaccionalidad- y aumentar la rentabilidad del banco.

La medición de la satisfacción del cliente se debe realizar a través de una encuesta, cuyo formulario se muestra en el Anexo 17.

CAPÍTULO 6

6.1 CONCLUSIONES

- La metodología de clasificación de agencias bancarias desarrollada en esta tesis se basó en criterios estadísticos de análisis de datos y pronóstico, que permitieron fundamentar técnicamente la toma de decisiones y direccionar correctamente los recursos de la organización de donde se obtuvo el caso de estudio. El proceso lógico aplicado puede ser utilizado en distintos campos de acción donde la generación de datos y su transformación en información es lo más importante. Desde la recolección de los datos, la generación y selección de variables hasta el pronóstico del comportamiento de nuevos individuos, hace de esta metodología una fusión de criterios y pasos que se pueden generalizar a las necesidades de cada institución bancaria para conocer sus agencias y mejorar sus servicios con una inversión más focalizada.
- El caso de estudio aplicado en esta tesis ofreció información suficiente para el análisis que fue en beneficio del banco y del mejoramiento de la calidad del servicio al cliente. Algunas agencias cambiaron su estructura física, según las necesidades del cliente ya sea transaccional o de negocio.
- El análisis cluster, el análisis de componentes principales y el análisis discriminante consienten una pérdida de información, pero se obtiene una ganancia en significación, estudian la estructura de bases de datos

grandes suministrando una representación simplificada de la misma, pero más significativa, sencilla y fácil de ver.

- La clasificación de las agencias por el tipo de servicio que ofrecen al cliente respondió a las necesidades del área de marketing y de canales, ya sea por la optimización de los recursos como por el mejoramiento del servicio al cliente. A los gerentes de estas áreas les permitió realizar estudios concretos en cada uno de los 3 grupos, cruzarlos con otras variables que consideraban relevantes de la interacción diaria de la agencia con el cliente. Los resultados fueron favorables para la clasificación y ratificó la asignación de la mayoría de las agencias a cada grupo.
- Las herramientas utilizadas en la metodología propuesta son exhaustivas por tratar toda la información disponible. Permitieron analizar de manera simultánea gran cantidad de datos, lo cual, es importante en marketing. El estudio inicia con 11 variables seleccionadas por los gerentes e involucrados en el estudio, y la metodología propuesta determinó la selección de 4 variables como relevantes, es decir significativas, y son: Aums per cápita, clientes resto, eficiencia financiera y transacciones jurídicos. En conclusión, la reducción de variables ayudó a entender mejor la caracterización de los grupos. Estas variables recogen información de los ejes de negocio y transaccional, siendo los campos Aums per cápita y eficiencia financiera consideradas de negocio; y clientes resto y transacciones jurídicos como transaccional.
- La conceptualización de cada grupo se basó en los resultados obtenidos del análisis Cluster y del análisis discriminante. Un grupo está caracterizado por una alta transaccionalidad, otro con un monto de negocio intermedio y transaccionalidad media y otro con un mayor enfoque de negocio. A través de la aplicación de la clasificación, el banco tendría un ahorro importante de novecientos mil dólares,

aproximadamente, en activos fijos, en el grupo de agencias de características transaccionales, con la posibilidad de mejorar los activos fijos de las agencias de los otros grupos.

- Con la clasificación de las agencias se puede diferenciar el tipo de servicio que se debe brindar al cliente del banco. Si un cliente va a una agencia clasificada como de negocio recibirá asesoría comercial y financiera, mientras que en una agencia transaccional no la encontrará, por tanto los clientes con necesidades de negocio se concentrarán en los grupos de agencias de negocio y –negocio y transaccional-; pero las personas que no pertenecen al banco realizarán sus transacciones en las agencias calificadas como transaccionales. Las agencias que pertenecen al grupo transaccional y de negocio recibirán a clientes y no clientes del banco, pero internamente ofrecerán también un servicio diferenciado, puesto que es necesario reducir el tiempo de espera del cliente en la cola del banco y ofrecer un mejor servicio.
- El paquete estadístico utilizado en la aplicación de los diferentes métodos es el SPSS, es un paquete de fácil uso y muestra una visión sintética de las relaciones entre las filas y columnas de grandes tablas. Se convierte así en un nuevo medio de comunicación, entre el investigador y los datos para obtener información.
- La metodología de clasificación recoge la importancia y relevancia que tienen las herramientas de análisis de datos en el mundo de la información, es una respuesta a la necesidad de segmentar todo tipo de datos. Día a día, grupos de investigación con distintos enfoques agrupan la información y la clasifican, por ejemplo, se conoce cuántas personas pertenecen a niveles de pobreza en el mundo, qué países son más ricos que otros, qué cantones en el Ecuador son más vulnerables, que empresas según un conjunto de índices son las más productivas, cuáles son mis clientes más rentables, cuáles puntos de venta atienden a mis clientes importantes, etc. Esta metodología

responde a la necesidad de clasificar los datos para convertirlos en información de una manera lógica y técnica, basándose en conocimientos matemáticos de análisis de datos.

6.2 RECOMENDACIONES

- Al analizar datos reales, se debe tomar en cuenta que existen factores que intervienen y que no son controlables, por ejemplo, la ausencia de datos y falta de flujo de la información de un área a otra. El efecto de estos factores no puede ser cuantificado, por lo que no son medidos por el paquete, pero deben ser mencionados.
- La metodología de clasificación aplicada al caso práctico analiza únicamente variables cuantitativas, tanto en el análisis de correlaciones, análisis de componentes principales, análisis cluster y análisis de discriminante. Para el caso del análisis discriminante si se requieren variables cualitativas se recomienda que sus valores sean recodificados, mediante la creación de nuevas variables, a valores numéricos que correspondan en algún sentido a las categorías originales. Cuando una variable tiene dos categorías se puede codificar como 0 (ausencia de la cualidad) y 1 (presencia de la cualidad), si tiene más de dos categorías, debe generarse tantas variables como total de características menos uno.
- Los métodos de análisis de datos aplicados para realizar la clasificación de las agencias bancarias tienen criterios técnicos que deben ser conocidos y perfectamente manejados por el investigador, puesto que la falta de validez de algún criterio estadístico puede invalidar la clasificación y los resultados no responderán a la realidad.
- Se recomienda a las empresas que tienen interés en mejorar y optimizar sus recursos y focalizar la toma de decisiones, utilicen herramientas estadísticas en sus estudios y análisis de las bases de datos. De esta manera tendrán un sustento matemático y estadístico que combinado con su experiencia explicarán científicamente la realidad de su entorno.

REFERENCIAS BIBLIOGRÁFICAS

- Grande, Idelfonso y Abascal, Elena, “MÉTODOS MULTIVARIANTES PARA LA INVESTIGACIÓN COMERCIAL”, Editorial Ariel S.A., Barcelona, 1989.
- Ferrán, Magdalena, “SPSS para Windows, programación y análisis estadístico”, Mc.Graw-Hill, España, 1996.
- Santesmases, Miguel, “DYANE, diseño y análisis de encuestas en investigación social y de mercados”, Ediciones Pirámide, España, 2001.
- Westphal, Blaxton Teresa, “DATA MINING SOLUTIONS”, métodos y herramientas para resolver problemas del mundo real, Editorial John Wiley & Sons, Inc, USA, 1998.
- Peña Daniel, “ANÁLISIS DE DATOS MULTIVARIANTES”, Editorial Mc. Graw Hill, Madrid – España, 2002.
- Joseph F. Hair, Jr., “ANÁLISIS MULTIVARIANTE”, Editorial Prentice Hall, 5ta. Edición, España, 1999.
- Sallenave, Juan Pablo, “LA GERENCIA INTEGRAL”, Editorial Norma, Quito, 1994.
- Sapag Chain, Nassin, “EVALUACIÓN DE PROYECTOS”, Editorial CIADE, 1997.

ANEXO 1: Tamaño de la Economía

ANEXO 2: Crecimiento económico

ANEXO 3: Distribución de hombres y mujeres en el Ecuador

nombre_provincia	nombre_sexo	poblacion_total
Azuay	Hombres	279,792
	Mujeres	319,754
Bolívar	Hombres	83,156
	Mujeres	86,214
Cañar	Hombres	95,010
	Mujeres	111,971
Carchi	Hombres	75,834
	Mujeres	77,105
Chimborazo	Hombres	190,667
	Mujeres	212,965
Cotopaxi	Hombres	169,303
	Mujeres	180,237
El Oro	Hombres	266,716
	Mujeres	259,047
Esmeraldas	Hombres	197,150
	Mujeres	188,073
Galápagos	Hombres	10,204
	Mujeres	8,436
Guayas	Hombres	1,648,398
	Mujeres	1,660,636
Imbabura	Hombres	167,818
	Mujeres	176,226
Loja	Hombres	197,595
	Mujeres	207,240
Los Ríos	Hombres	335,279
	Mujeres	314,899
Manabí	Hombres	596,502
	Mujeres	589,523
Morona Santiago	Hombres	57,425
	Mujeres	57,987
Napo	Hombres	40,284
	Mujeres	38,855
Orellana	Hombres	46,798
	Mujeres	39,695
Pastaza	Hombres	31,988
	Mujeres	29,791
Pichincha	Hombres	1,167,332
	Mujeres	1,221,485
Sucumbios	Hombres	70,139
	Mujeres	58,856
Tungurahua	Hombres	213,513
	Mujeres	227,521
Zamora Chinchipe	Hombres	39,662
	Mujeres	36,939
Zonas no delimitadas	Hombres	37,788
	Mujeres	34,800
TOTAL		12,156,608

ANEXO 4: Población Económicamente Activa

Fuente: SIISE

Ordenar	Provincias	Indicador	Grupo
1	Azuay	192,688	2
2	Bolívar	51,250	1
3	Cañar	65,168	1
4	Carchi	48,277	1
5	Cotopaxi	93,169	1
6	Chimborazo	126,608	1
7	El Oro	145,068	1
8	Esmeraldas	91,990	1
9	Guayas	877,737	5
10	Imbabura	89,589	1
11	Loja	122,361	1
12	Los Ríos	166,443	1
13	Manabí	302,241	2
14	Morona Santi	28,358	1
15	Napo	19,764	1
16	Pastaza	15,515	1
17	Pichincha	661,992	4
18	Tungurahua	138,107	1
19	Zamora Chinc	22,953	1
20	Galápagos	4,759	1
21	Sucumbíos	27,948	1
22	No delimitada	18,557	1
23	Orellana	17,008	1

ANEXO 5: Ingreso Total personal según quintiles

ANEXO 6: Pobreza por Necesidades Básicas Insatisfechas

Fuente: SIISE

nombre_provincia	porcentaje	numero	poblacion_total
Azuay	53	318,834	599,546
Bolívar	77	129,602	169,370
Cañar	70	144,593	206,981
Carchi	60	91,207	152,939
Chimborazo	67	271,834	403,632
Cotopaxi	76	265,338	349,540
El Oro	58	302,411	525,763
Esmeraldas	76	292,925	385,223
Galápagos	41	7,563	18,640
Guayas	61	2,027,210	3,309,034
Imbabura	58	200,150	344,044
Loja	68	276,165	404,835
Los Ríos	77	502,608	650,178
Manabí	75	887,380	1,186,025
Morona Santiago	76	87,538	115,412
Napo	77	60,986	79,139
Orellana	83	71,524	86,493
Pastaza	67	41,346	61,779
Pichincha	41	970,474	2,388,817
Sucumbios	82	105,430	128,995
Tungurahua	61	270,475	441,034
Zamora Chinchipe	77	58,754	76,601
Zonas no delimitadas	87	62,864	72,588
TOTAL	0.612606	7,447,211	12,156,608

ANEXO 7: Distribución Bono Solidario

Fuente: SIISE

Región	Provincia	Madres	Tercera edad	Discapacitados	Total beneficiarios
Costa	El Oro	54,537	11,893	664	67,094
Costa	Esmeraldas	36,358	6,851	367	43,576
Costa	Guayas	292,116	56,799	1,792	350,707
Costa	Los Ríos	69,216	10,913	218	80,347
Costa	Manabí	126,461	30,165	1,410	158,036
Costa	Galápagos	630	110	0	740
Costa	No delimitadas	2,505	241	11	2,757
Sierra	Azuay	37,146	10,559	523	48,228
Sierra	Bolívar	19,053	6,240	122	25,415
Sierra	Cañar	15,621	4,542	136	20,299
Sierra	Carchi	8,844	3,080	46	11,970
Sierra	Cotopaxi	27,341	7,664	189	35,194
Sierra	Chimborazo	37,816	12,073	196	50,085
Sierra	Imbabura	28,540	11,899	282	40,721
Sierra	Loja	39,333	14,265	475	54,073
Sierra	Pichincha	146,375	35,801	1,445	183,621
Sierra	Tungurahua	33,477	10,985	240	44,702
Amazonía	Morona Santiago	6,857	1,182	38	8,077
Amazonía	Napo	11,629	1,660	56	13,345
Amazonía	Pastaza	4,248	720	53	5,021
Amazonía	Zamora Chinchipe	7,721	1,276	93	9,090
Amazonía	Sucumbíos	10,172	1,158	33	11,363
Amazonía	Promedio país	153	27	0	180

ANEXO 8: IMPORTAR DATOS DESDE EXCEL A SPSS

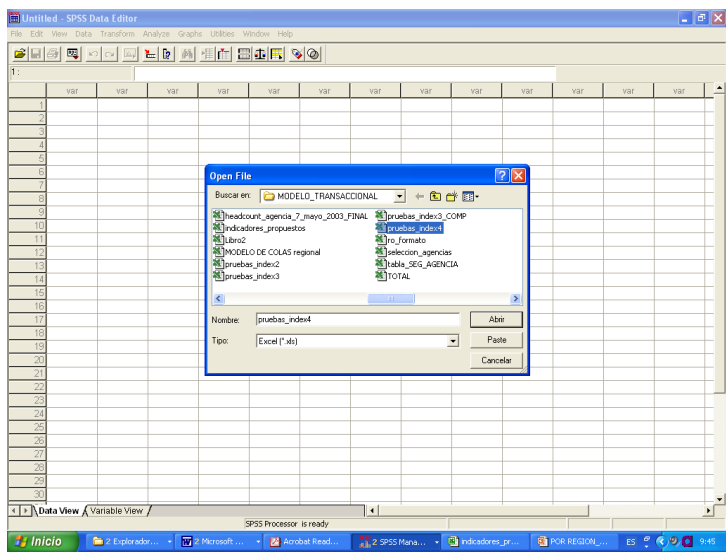
Una vez elaborada la tabla en EXCEL:

Figura 1:

	A	B	C	D	E	F	G	H	I	J	K	L
	COD_SEG	COD_AGE	AGENCIA	IPC	IF	IPC	IT	INC	IPAV	IC	IR	IPC
1												
2	6	39	AMBATO	604536.036	0.91119994	1465.5	0.69055649	1.90225387	0.04679107	0.02351334	0.92969956	505.29
3	6	43	CUMANDA	575484.896	0.78806107	1699.3125	0.94294247	0.48015679	0.08579656	0.05662188	0.85758157	325.63
4	6	67	PILLARO	413991.386	0.79172528	1379.8	1.00367505	0.60404657	0.06263625	0.01799747	0.89936629	394.50
5	6	129	SUR AMBATO	294698.046	0.89505722	1976.77778	0.71568091	0.39351453	0.08050674	0.02983245	0.88966081	271.89
6	5	22	CUENCA	2562971.31	0.6367054	3127.36364	0.61602838	1.95600493	0.09123937	0.03318476	0.87567586	809.06
7	5	32	HUAYNA CAPAC	424962.804	0.8515101	1864.47368	0.77049282	1.02440139	0.06056616	0.0236998	0.91573404	479.68
8	5	37	CENTRO CUENCA	387450.613	0.84174649	2305.13333	0.72586627	0.85897849	0.06620999	0.00896287	0.93482714	520.67
9	5	76	TOTORACOCOA	375995.263	0.89468136	2846.8	0.87136092	1.63910262	0.06849606	0.00808339	0.92342055	470.10
10	5	77	ORDOÑEZ LAZO	384560.071	1.05761714	2006.11111	0.93011983	0.7895728	0.04411765	0.00917141	0.94671094	702.67
11	5	78	REMIGIO CRESPO	580424.307	0.92571844	2819	0.56290319	0.70280882	0.07112223	0.02072909	0.90814868	699.50
12	5	79	AZOGLIES	564626.839	0.68193072	1803.90909	0.66524108	1.60783263	0.07948244	0.01579566	0.9047219	541.00
13	5	161	GUALACEO	380384.966	0.83954865	1997.33333	0.87339997	0.89231012	0.08442936	0.0093339	0.90623674	392.83
14	5	168	ARENAL	418945.708	0.74686407	2553.66667	0.85312601	0.38135693	0.09376694	0.01571816	0.89051491	307.50
15	5	167	SUCRE	519468.366	0.92755228	1923.2	0.57453163	0.60026627	0.0738255	0.03559965	0.89057485	342.70
16	2	13	ESMERALDAS	780332.337	0.86030488	2269.77419	0.89020644	0.78613494	0.05142433	0.02731429	0.92126138	825.52
17	2	87	ATACAMES	189762.819	1.32262714	1975.9	0.92571986	0.71043975	0.03444621	0.01563328	0.94992051	377.40
18	2	88	LAS PALMAS	322680.779	1.14869976	2641.25	0.9679799	0.75178246	0.05036232	0.02137681	0.92626087	345.00
19	2	93	GUININDÉ	222580.563	1.08959082	1962.63636	0.74375587	0.79087516	0.0244672	0.0122436	0.8632682	571.73
20	2	119	OLMEDO ESMERALD	201167.033	1.2893957	3254.5	0.78940752	0.47382463	0.04087658	0.01647346	0.94264796	327.80
21	3	11	GUAYASQUIL TORRE	3627412.6	0.7175442	1858.71053	1.08694872	1.20524672	0.1155232	0.0588062	0.83158856	436.89
22	3	19	CENYANARI	508359.183	1.02514321	1174.1	0.46993997	0.69186801	0.04183328	0.04890903	0.90925869	660.20
23	3	20	MERCADO CENTRAL	352166.13	1.1102526	1385.35714	0.98332901	0.44480036	0.04626175	0.02507088	0.92866736	478.64
24	3	26	AV. DE LAS AMERIC	542093.164	0.90788843	1692.30769	0.95191673	0.66793025	0.03963895	0.0320414	0.92831975	698.62
25	3	30	ALBAN BORJA	564202.205	0.87293663	1705.66667	0.85950897	0.77889797	0.05394345	0.04501488	0.90104167	672.00
26	3	44	LA ALBORADA	598663.706	0.99250115	1527.88462	0.594524	0.6450637	0.04930819	0.02889311	0.9217987	567.08
27	3	46	URDESA	721789.709	0.80249003	1455.10526	0.7813661	0.92166539	0.12866778	0.04596961	0.82436261	408.74
28	3	48	OLMEDO GUAYASQU	322586.93	1.17187178	2278.375	0.94613541	0.48670473	0.03311258	0.01903974	0.94784768	453.00
29	3	55	CENYANARI SUR	344369.854	1.2372903	1563.52381	1.05183492	0.64845868	0.02861534	0.01156323	0.95982143	650.67
30	3	56	AYACUCHO	362745.399	1.20349623	1495.27273	0.69661397	0.51944573	0.04004004	0.02127127	0.93868869	363.27
31	3	58	VIA DAULIE	387968.105	1.04367362	1354.08333	0.53286659	0.81148272	0.02296089	0.02856761	0.9484715	624.25
32	3	106	LIBERTAD	292034.374	1.19203541	1582.69231	0.89766625	0.53476055	0.02966163	0.02548596	0.94485241	534.23

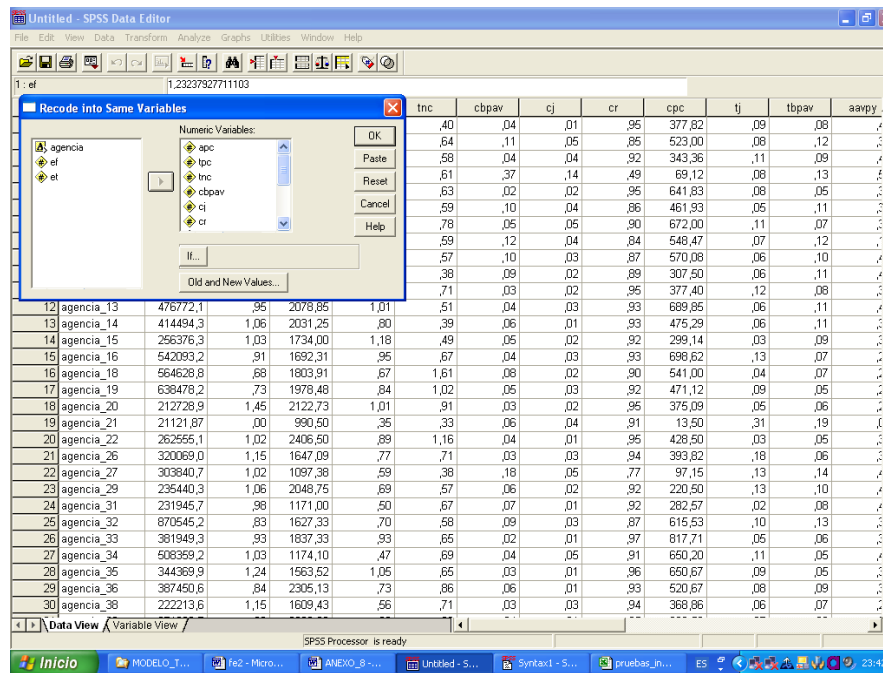
En el paquete SPSS versión 10 se abren las siguientes pantallas, en: File, Open, Data y se selecciona el tipo de extensión “.xls” y se escoge el documento, Abrir:

Figura 2:



Para el análisis se reemplazan los valores missing o faltantes a través del siguiente proceso: en la barra de herramientas seleccionar Transform, Recode, Info Same Variables y la opción Old and New Values, donde se transforman los missing en ceros.

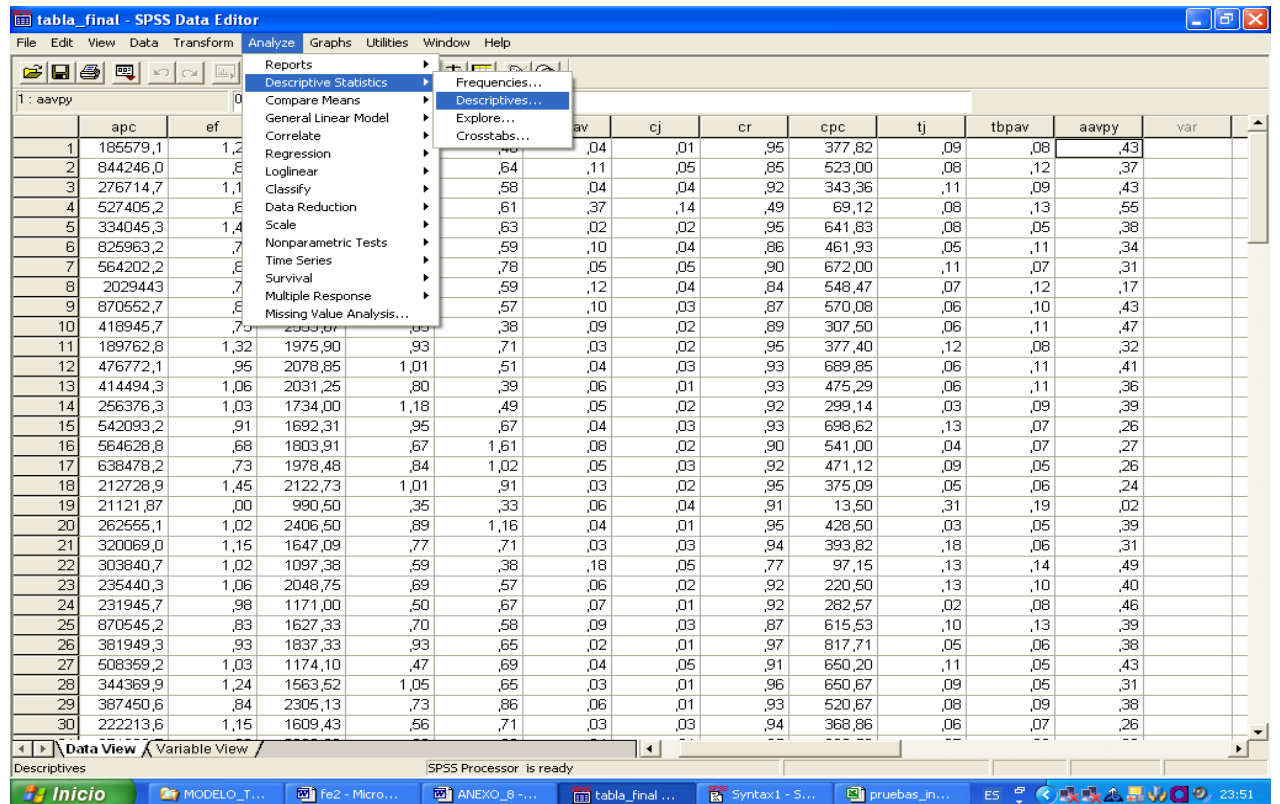
Figura 3:



Luego, se analiza la distribución de cada variable cuantitativa seleccionada en el estudio y se identifican los datos atípicos, a continuación se observan algunas pantallas:

En el menú Analyze, Descriptives statistics, Frecuencias obtienen la mediana, la media, la desviación estándar, el sesgo y la kurtosis, y se guardan las variables estandarizadas.

Figura 4:



El resultado es una tabla con el siguiente formato:

Figura 5:

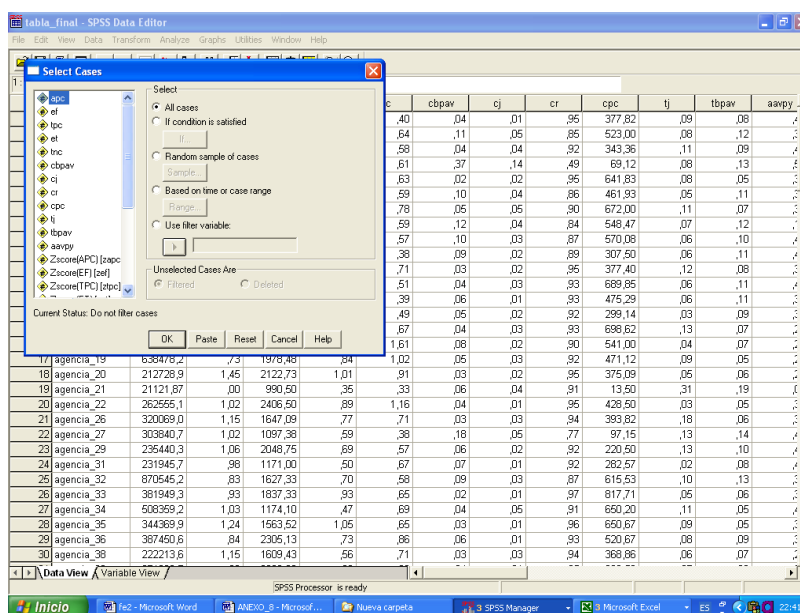
	APC	EF	TPC	ET	TNC	CBPAV	CJ	CR	CPC	TJ	TBPAV	AAVPHY
N	118	118	118	118	118	118	118	118	118	118	118	118
	0	0	0	0	0	0	0	0	0	0	0	0
Mean	445568,8	,9917	2007,9939	,8640	,7194	6,085E-02	2,369E-02	,9155	488,5008	7,681E-02	7,751E-02	,3508
Median	365082,1	,9940	1970,7000	,8536	,6572	4,984E-02	1,944E-02	,9265	471,1790	7,128E-02	7,327E-02	,3615
Std. Deviation	314786,5	,2956	575,5724	,2551	,3406	4,482E-02	1,650E-02	5,841E-02	266,7338	4,106E-02	2,922E-02	,1085
Skewness	2,326	-,599	1,257	2,023	2,253	3,543	3,254	-3,907	1,076	2,023	,860	-,495
Std. Error of Skewness	,223	,223	,223	,223	,223	,223	,223	,223	,223	,223	,223	,223
Kurtosis	8,474	3,611	3,211	10,955	7,305	20,051	20,329	24,477	3,279	8,344	1,561	1,331
Std. Error of Kurtosis	,442	,442	,442	,442	,442	,442	,442	,442	,442	,442	,442	,442
Minimum	,00	,00	990,50	,35	,29	,00	,00	,49	,06	,02	,02	,00
Maximum	2029443	1,96	4431,00	2,43	2,43	,37	,14	1,00	1595,40	,31	,19	,62

ANEXO 9: AGENCIAS ELIMINADAS SEGÙN CRITERIOS

Luego de identificar los límites en los cuales se eliminarán las agencias según cada variable. En el paquete SPSS se puede realizar este proceso a través de los siguientes pasos:

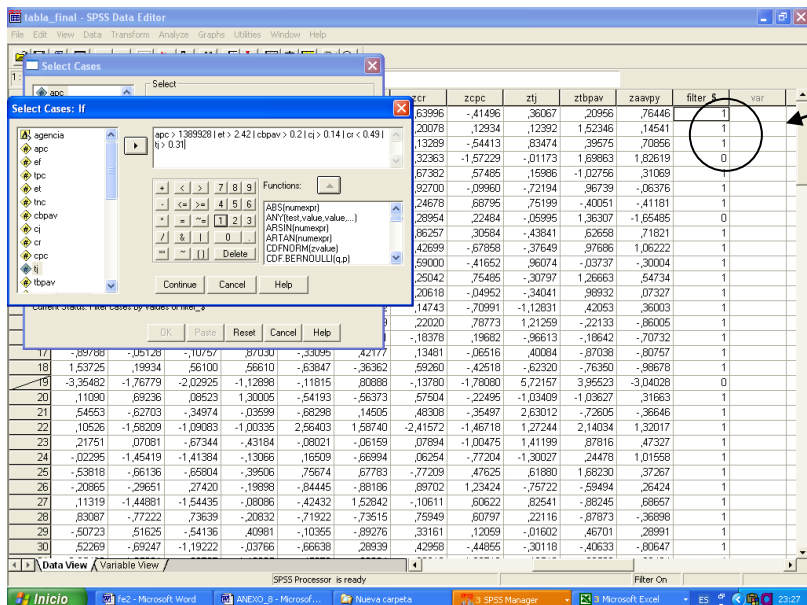
1. Menú Data
2. Select case
3. If condition is satisfied
4. if

Figura 6:



Cuando las restricciones son ingresadas se corre el filtro de selección de las agencias que deben ser eliminadas.

Figura 7:

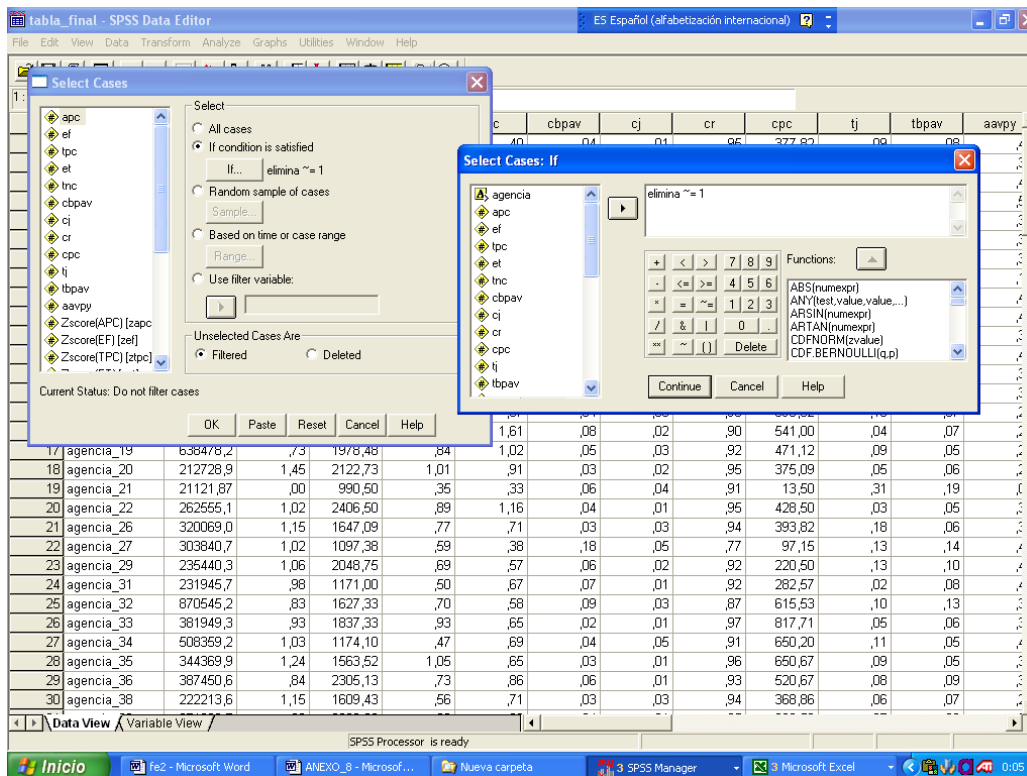


Se genera la variable Filter que contiene la identificación de

ANEXO 10: PARTICIÓN DE LA BASE DE DATOS

1. Eliminar las agencias que cumplieron las restricciones para datos atípicos.
Inicialmente abrir el menú DATA, Select Cases, If condition is satisfied y Select Case, if, y digitar la condición:

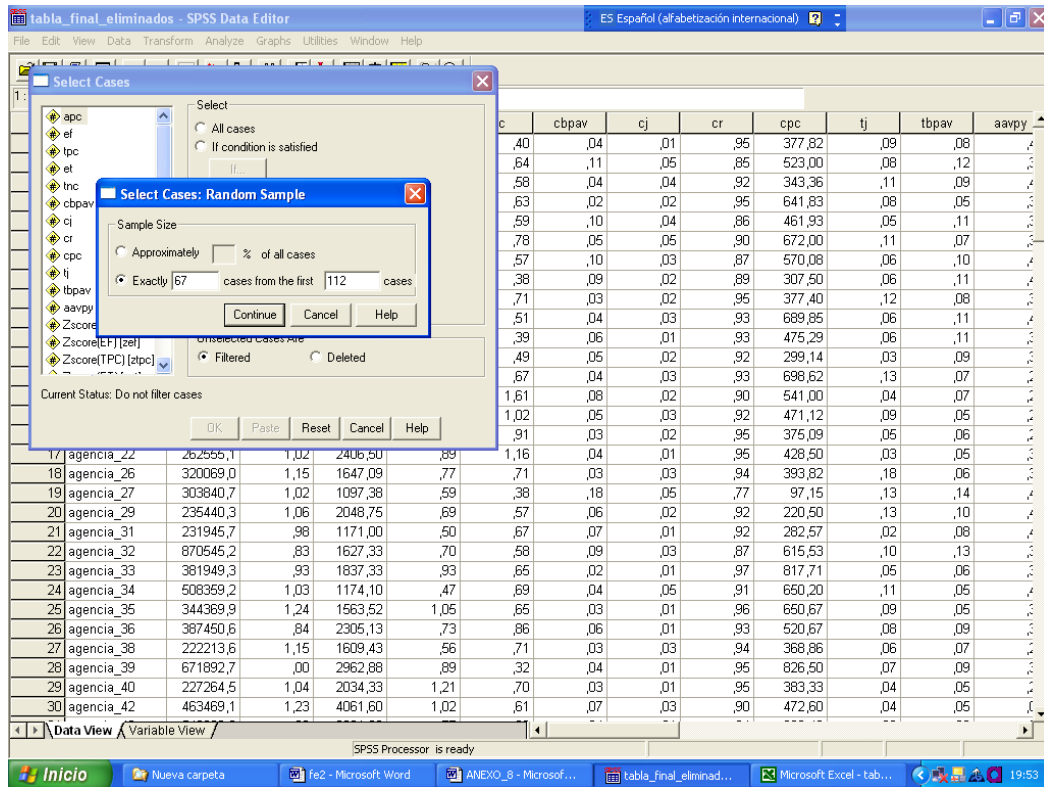
Figura 8:



Luego seleccionar Deleted, y OK.

2. Para formar los tres grupos, se utiliza el proceso: menú DATA, Select Cases, Random Sample Cases y Simple Size y en Exactly los 67 casos:

Figura 9:



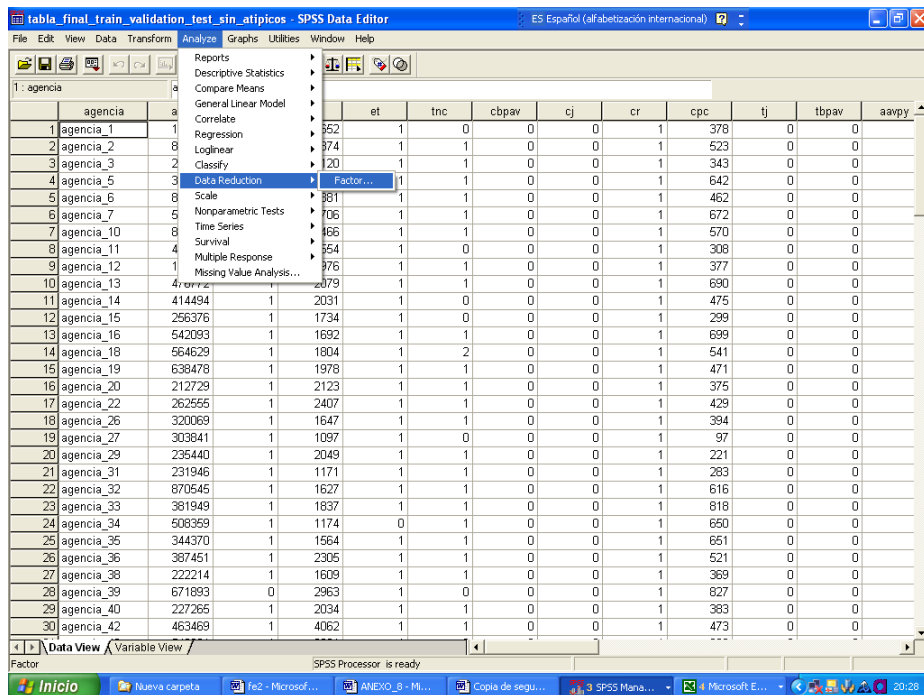
3. Luego de escoger los 67 casos se filtra la tabla y se seleccionan con el mismo proceso los grupos de validación y prueba o test.

ANEXO 11: ANÁLISIS DE COMPONENTES PRINCIPALES (ACP)

Con la tabla elaborada se realiza el siguiente proceso para realizar el ACP:

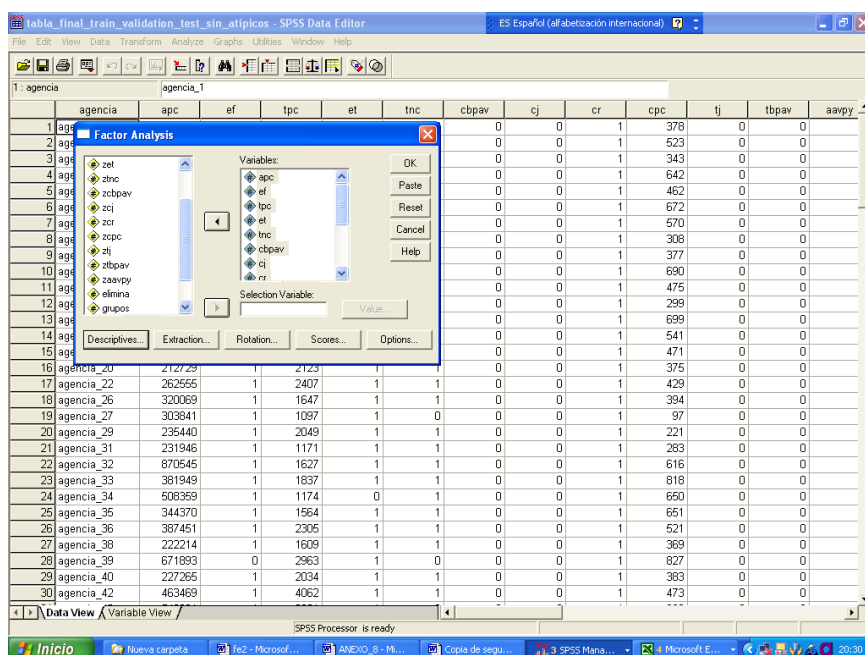
1. En el menú se selecciona Analyze, Data Reduction, Factor

Figura 10:



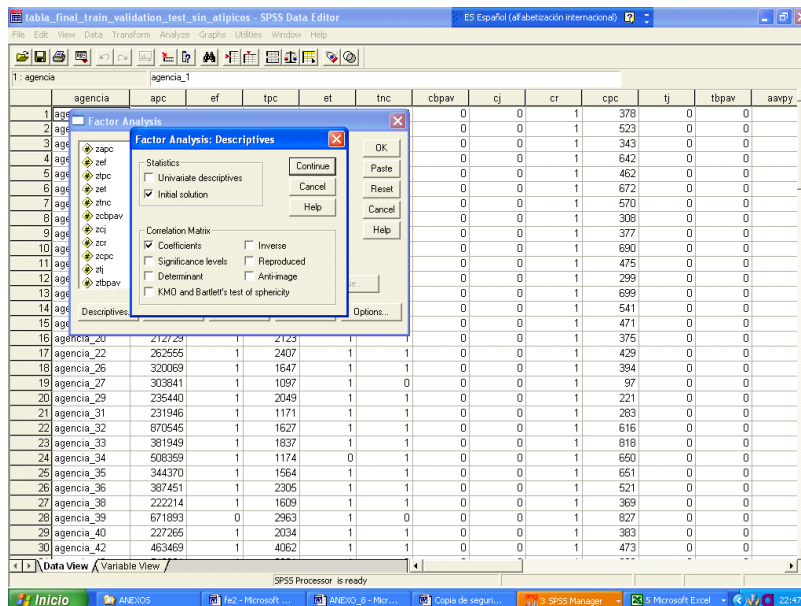
2. Luego, en la ventana se seleccionan las variables que serán analizadas y no importa si no están estandarizadas puesto el PAQUETE SPSS realiza este proceso automáticamente:

Figura 11:



3. En la misma ventana se abre la pestaña Descriptives donde se marca:
 - a. Statistic → Inicial Solution
 - b. Continue

Figura 12:



4. Una vez obtenida la matriz de correlaciones y la primera corrida del ACP, se analizan las comunalidades, los valores propios y la varianza explicada.

Figura 13:

Communalities

	Initial	Extraction
APC	1,000	,821
EF	1,000	,691
TPC	1,000	,573
ET	1,000	,761
TNC	1,000	,751
CBPAV	1,000	,819
CJ	1,000	,620
CR	1,000	,918
CPC	1,000	,772
TJ	1,000	,520
TBPAV	1,000	,716
AAVPY	1,000	,400

Extraction Method: Principal Component Analysis

- 4.1 Como se observa, las comunalidades son cercanas a 1 o superiores al 0.5, por lo tanto son adecuadas.

4.2 En la tabla están los valores propios superiores a 1 y una varianza explicada en el primer factor del 34% aproximadamente.

Figura 14:

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,053	33,771	33,771	4,053	33,771	33,771
2	1,777	14,806	48,577	1,777	14,806	48,577
3	1,413	11,774	60,350	1,413	11,774	60,350
4	1,121	9,340	69,690	1,121	9,340	69,690
5	,914	7,614	77,304			
6	,844	7,036	84,340			
7	,666	5,550	89,890			
8	,542	4,518	94,408			
9	,301	2,507	96,916			
10	,217	1,807	98,722			
11	,153	1,278	100,000			
12	5,372E-16	4,477E-15	100,000			

Extraction Method: Principal Component Analysis.

4.3 Por último están los coeficientes de cada variable respecto a los factores. La variable más importante para el primer factor es “Clientes Resto” con un coeficiente de -0,9, seguida por “Clientes Banca Privada y Alto Valor” con un coeficiente de 0,859. Para el segundo factor es la variable “Clientes Per cápita” con 0,853.

Figura 15:

Component Matrix^a

	Component			
	1	2	3	4
APC	,585	,681	-,109	5,428E-02
EF	-,327	-,462	,608	-1,01E-02
TPC	-,454	,137	-5,59E-02	,587
ET	-,388	-2,43E-02	,364	,691
TNC	-,335	-,306	-,726	,133
CBPAV	,859	-,239	-,101	,115
CJ	,717	-,262	-,135	,137
CR	-,901	,270	,122	-,133
CPC	-,205	,853	7,158E-03	-5,08E-02
TJ	,463	-1,01E-02	,512	-,209
TBPAV	,761	,222	,249	,161
AAVPY	,483	,102	9,052E-03	,395

Extraction Method: Principal Component Analysis.

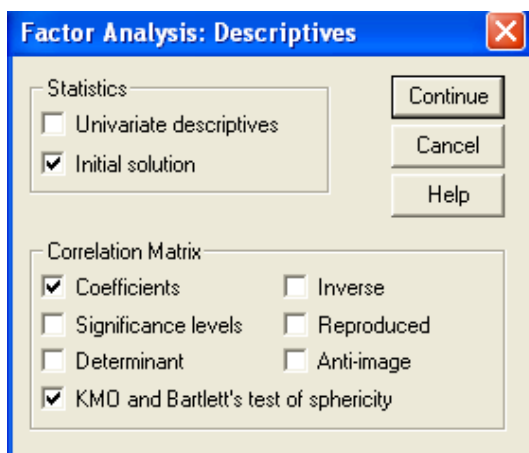
a. 4 components extracted.

4.4 En este primer ACP no se obtuvo la matriz de covarianza puesto que se genera una matriz no definida positiva, es decir que existe linealidad entre las variables estudiadas y se elimina la variable CR.

4.5 Una vez eliminada la variable CR, en la misma ventana se abre la pestaña Descriptives donde se marca:

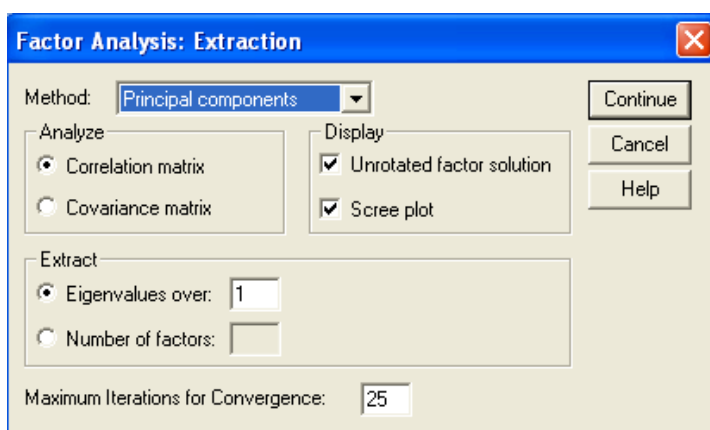
- a. Statistic → Inicial Solution
- b. Coefficient Matrix → Coeficientes, KMO and Barleth test o sphericity
- c. Continue

Figura 16:

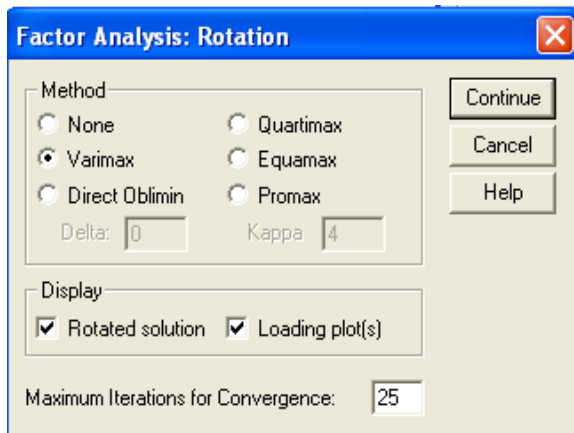


- d. En la Pestaña Extraction seleccionar Scree Plot → Continue

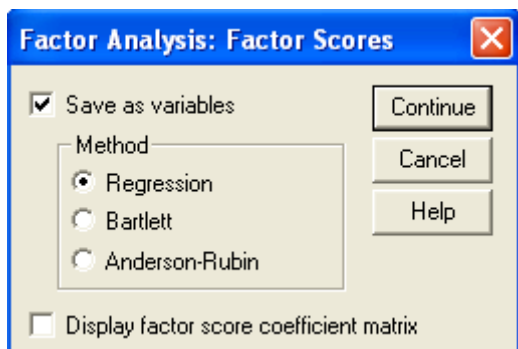
Figura 17:



- e. En la Pestaña Rotation seleccionar Varimax → loading plot → Continue, se obtiene la rotación ortogonal de los factores y un gráfico de la distribución de los individuos en los factores.

Figura 18:

- f. En la Pestaña Scores seleccionar Saves as variables → Method → Regression → Continue, con este proceso se generan las puntuajes de cada factor por el método de regresión.

Figura 19:

- g. Los resultados son analizados en el documento de esta tesis.

ANEXO 12: ANÁLISIS DE COMPONENTES PRINCIPALES PARA 4 Y 2 FACTORES

Siguiendo el proceso anterior y seleccionando en la pestaña Extraction el número de factores se obtienen los siguientes resultados:

Para cuatro factores: Factor Analysis

Figura 20:

Correlation Matrix

	CJ	TBPAV	AAVPY	EF	TPC	ET	TNC	CPC	TJ	APC	CR
Correlation CJ	1,000	,314	,275	-,157	-,244	-,169	,008	-,244	,390	,313	-,763
TBPAV	,314	1,000	,440	-,281	-,219	-,159	-,516	-,105	,320	,486	-,558
AAVPY	,275	,440	1,000	-,162	-,176	-,010	-,029	-,025	,196	,258	-,322
EF	-,157	-,281	-,162	1,000	,028	,290	-,103	-,142	,070	-,460	,174
TPC	-,244	-,219	-,176	,028	1,000	,333	,130	,164	-,256	-,181	,303
ET	-,169	-,159	-,010	,290	,333	1,000	,018	,030	-,122	-,165	,293
TNC	,008	-,516	-,029	-,103	,130	,018	1,000	-,112	-,337	-,262	,150
CPC	-,244	-,105	-,025	-,142	,164	,030	-,112	1,000	-,051	,477	,349
TJ	,390	,320	,196	,070	-,256	-,122	-,337	-,051	1,000	,191	-,276
APC	,313	,486	,258	-,460	-,181	-,165	-,262	,477	,191	1,000	-,401
CR	-,763	-,558	-,322	,174	,303	,293	,150	,349	-,276	-,401	1,000

Figura 21:

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,567
Bartlett's Test of Sphericity	Approx. Chi-Square	475,785
	df	55
	Sig.	,000

Figura 22:

Com m unalities

	Initial	Extraction
CJ	1,000	,690
TBPAV	1,000	,707
AAVPY	1,000	,488
EF	1,000	,701
TPC	1,000	,561
ET	1,000	,767
TNC	1,000	,787
CPC	1,000	,766
TJ	1,000	,520
APC	1,000	,818
CR	1,000	,808

Extraction Method: Principal Component Analysis.

Figura 23:

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,389	30,807	30,807	3,389	30,807	30,807	2,546	23,143	23,143
2	1,714	15,585	46,392	1,714	15,585	46,392	1,845	16,772	39,914
3	1,401	12,739	59,130	1,401	12,739	59,130	1,719	15,626	55,540
4	1,110	10,088	69,218	1,110	10,088	69,218	1,505	13,678	69,218
5	,855	7,771	76,989						
6	,822	7,476	84,465						
7	,627	5,702	90,166						
8	,517	4,702	94,869						
9	,271	2,465	97,333						
10	,178	1,615	98,948						
11	,116	1,052	100,000						

Extraction Method: Principal Component Analysis.

Figura 24:

Scree Plot

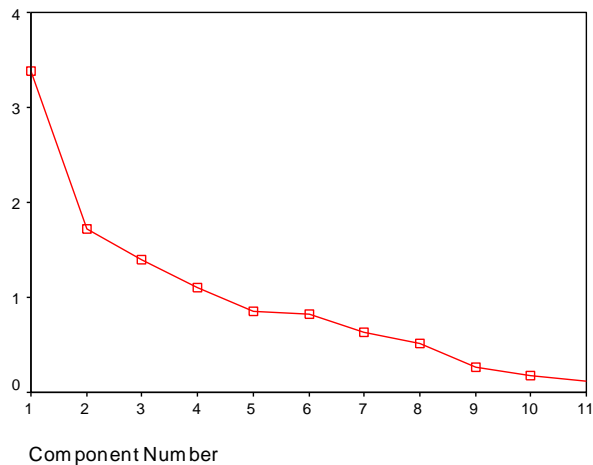


Figura 25:

Component Matrix

	Component			
	1	2	3	4
CJ	,703	-,340	-,194	,206
TBPAV	,778	,137	,247	,148
AAVPY	,515	2,838E-02	-1,64E-02	,471
EF	-,371	-,460	,593	-3,47E-02
TPC	-,477	,234	1,478E-02	,528
ET	-,391	8,199E-03	,400	,673
TNC	-,375	-,227	-,745	,201
CPC	-,113	,861	7,592E-02	-8,05E-02
TJ	,523	-,151	,450	-,147
APC	,639	,635	-7,32E-02	4,327E-02
CR	-,819	,308	,150	-,140

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Figura 26:

Rotated Component Matrix

	Component			
	1	2	3	4
CJ	,796	-4,71E-02	3,158E-02	-,232
TBPAV	,587	,302	,518	-6,26E-02
AAVPY	,633	,199	9,829E-02	,194
EF	-,271	-,683	,328	,232
TPC	-,176	8,422E-02	-,265	,673
ET	-3,70E-02	-,189	5,409E-02	,853
TNC	2,843E-02	-,129	-,877	7,001E-03
CPC	-,419	,737	,136	,169
TJ	,292	-9,84E-02	,622	-,195
APC	,327	,795	,257	-,113
CR	-,833	-3,66E-03	-,135	,311

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Figura 27:

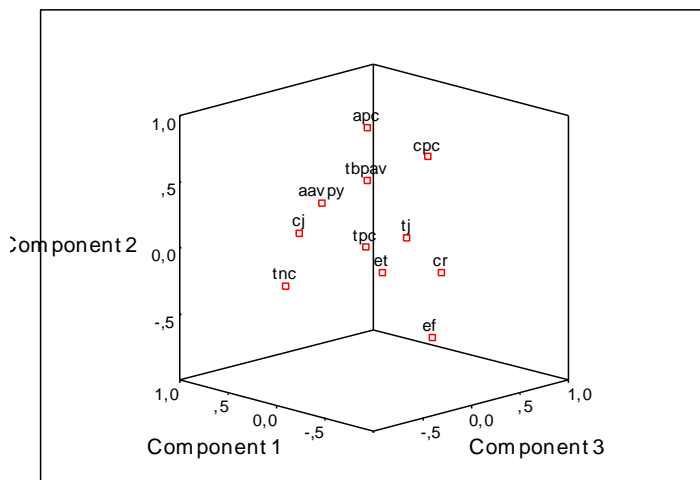
Component Transformation Matrix

Component	1	2	3	4
1	,773	,300	,405	-,384
2	-,321	,920	,114	,193
3	-,192	-,249	,882	,350
4	,512	,030	-,210	,832

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

Figura 28:

Component Plot in Rotated Space



En este gráfico está la nube de puntos de las variables proyectada sobre los tres primeros factores que explican el 56% de la varianza.

Para dos factores: Factor Analysis

Las comunalidades disminuyen para cada variable en especial para las variables eficiencia transaccional y transacciones no core.

Figura 29:

Com m unalities

	Initial	Extraction
CJ	1,000	,610
TBPAV	1,000	,624
AAVPY	1,000	,266
EF	1,000	,349
TPC	1,000	,282
ET	1,000	,153
TNC	1,000	,192
CPC	1,000	,754
TJ	1,000	,296
APC	1,000	,811
CR	1,000	,766

Extraction Method: Principal Component Analysis.

La representación gráfica de la varianza explicada por los valores propios se muestra a continuación:



Figura 30:

Component Matrix

	Component	
	1	2
CJ	,703	-,340
TBPAV	,778	,137
AAVPY	,515	2,838E-02
EF	-,371	-,460
TPC	-,477	,234
ET	-,391	8,199E-03
TNC	-,375	-,227
CPC	-,113	,861
TJ	,523	-,151
APC	,639	,635
CR	-,819	,308

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Figura 31:

Rotated Component Matrix

	Component	
	1	2
CJ	,780	4,038E-02
TBPAV	,617	,494
AAVPY	,438	,272
EF	-,104	-,581
TPC	-,530	-2,44E-02
ET	-,347	-,181
TNC	-,220	-,379
CPC	-,513	,701
TJ	,531	,119
APC	,255	,864
CR	-,867	-,124

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

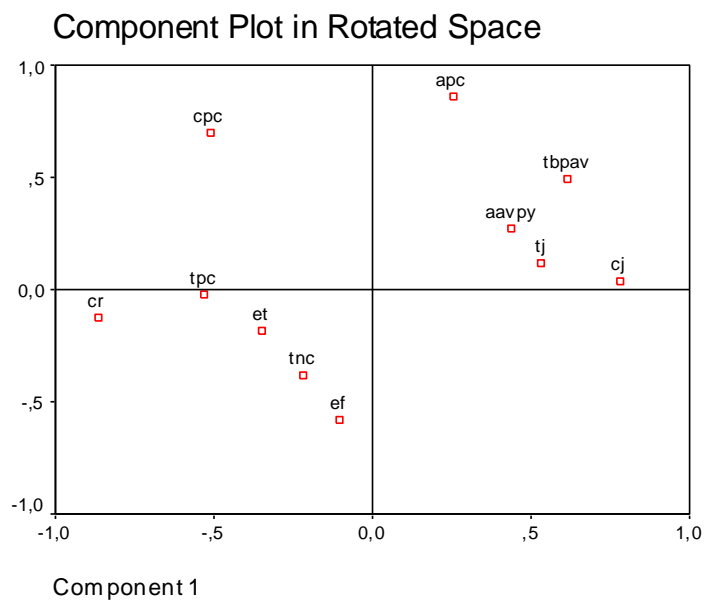
a. Rotation converged in 3 iterations.

El primer factor está caracterizado por un comportamiento del tipo de clientes que atienden las agencias, las transacciones y los montos (CJ, TBPAV, AAVPY, TPC, ET, TJ, CR) y el segundo factor está caracterizado por los per cápita de los clientes, los montos y eficiencia financiera (EF, TNC, CPC, APC).

Figura 32:**Component Transformation Matrix**

Component	1	2
1	,877	,481
2	-,481	,877

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.

Figura 33:

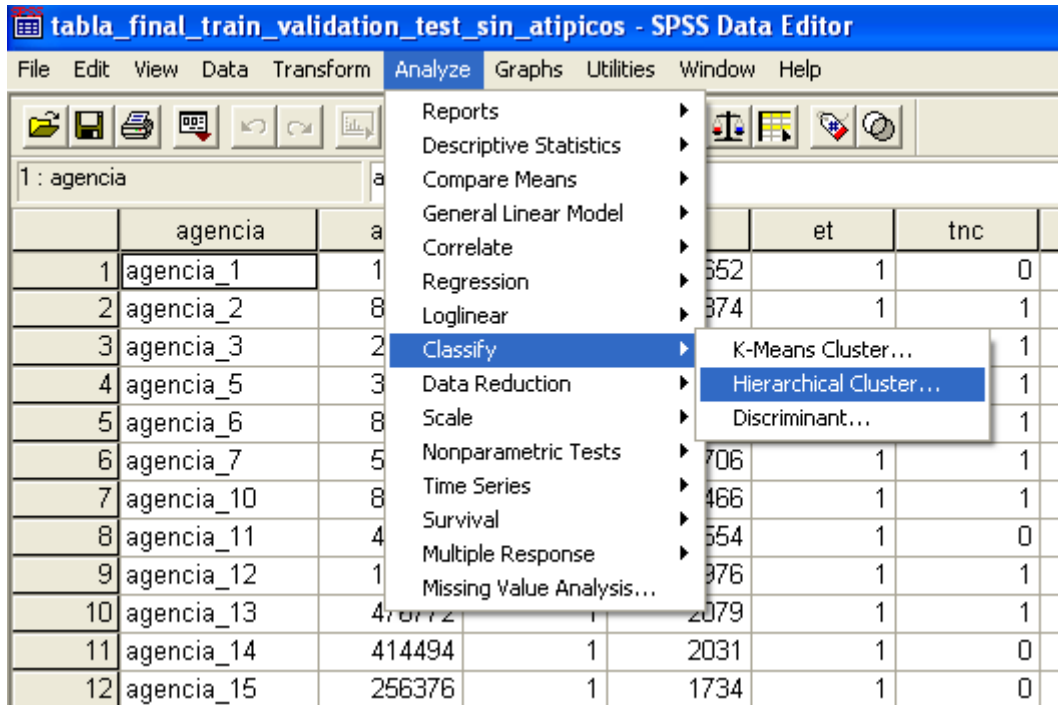
Como se puede observar en el gráfico existen: un grupo formado por las variables APC, TBPAY, AAVPY, CJ, TJ ; otro grupo formado por las variables CR, ET, TNC, EF, TPC y más alejada, al parecer con comportamiento dominante propio CPC.

Las diferencias que existen entre las diferentes corridas con diferente número de factores se observan en los valores de la comunalidad de cada variable respecto al número de factores seleccionado.

ANEXO 13: ANÁLISIS CLUSTER

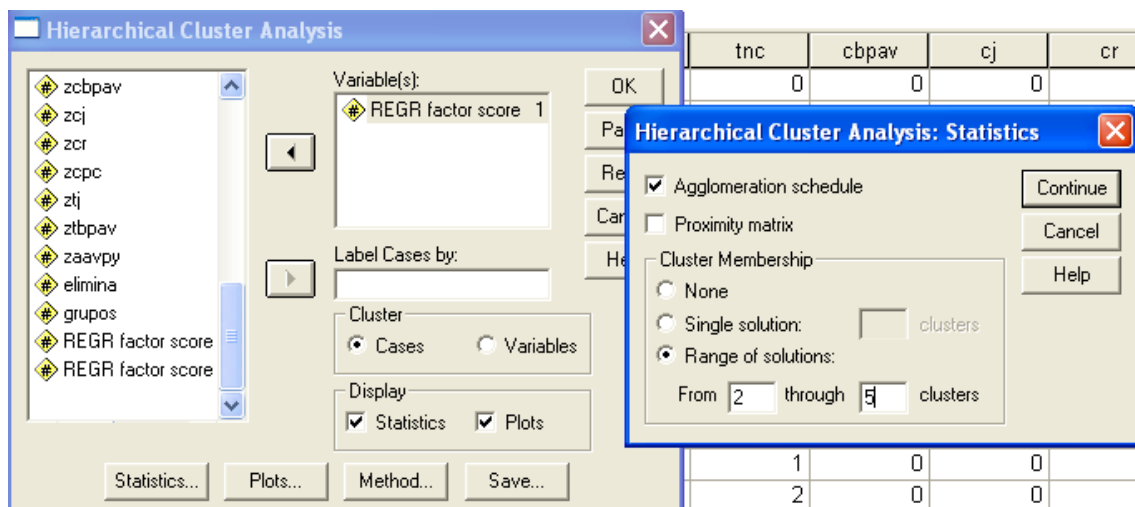
1. En el menú Analyze seleccionar: Classify→Hierarchical Cluster

Figura 34:



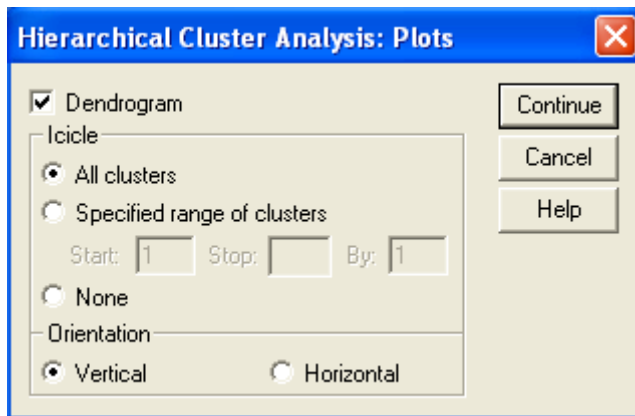
2. Luego, seleccionar la variable correspondiente al primer componente principal de Análisis de Componentes Principales. En la pestaña Statistics escoger el rango de cluster que se desea verificar:

Figura 35:



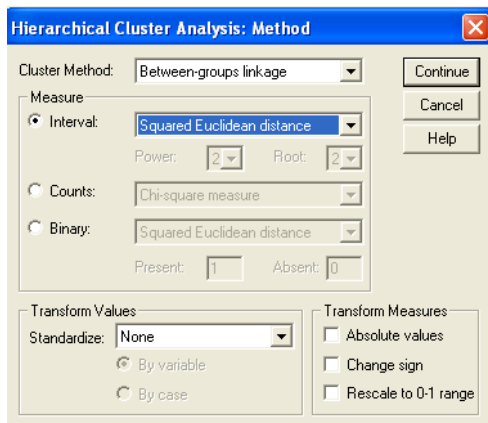
3. La mejor forma de ver cómo se llevó a cabo el proceso de agrupación es mediante la visualización de un gráfico dendograma, en la pestaña Plots:

Figura 36:



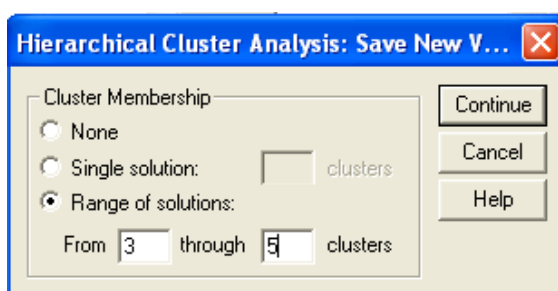
4. Para proceder a una medida de similitud se debe acudir a una consideración del nivel de medida de la variable, que en este caso es cuantitativa. Así pues, se utilizará la Medida de Similitud de Intervalo y el método de agrupación es la Relación entre Grupos:

Figura 37:



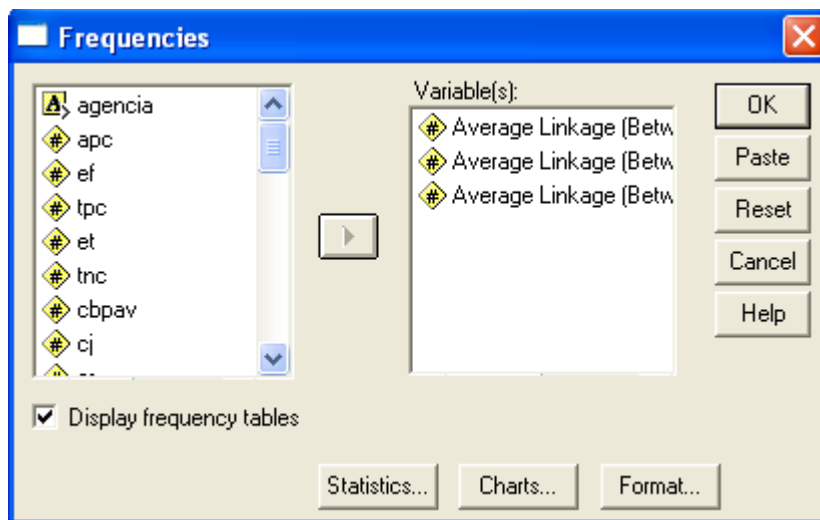
5. Para guardar las variables del proceso de agrupación:

Figura 38:



6. Finalmente Continue → OK
7. En el anexo 12.1 cluster jerárquico se pueden observar los resultados de la corrida del proceso mencionado.
8. Se realiza un análisis de frecuencias de las variables de agrupación:

Figura 39:



Los resultados son:

- Para 5 clusters:

Figura 40:

Average Linkage (Between Groups)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	72	64,3	64,3	64,3
	2	7	6,3	6,3	70,5
	3	26	23,2	23,2	93,8
	4	2	1,8	1,8	95,5
	5	5	4,5	4,5	100,0
Total		112	100,0	100,0	

- Para 4 clusters:

Figura 41:**Average Linkage (Between Groups)**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	72	64,3	64,3	64,3
2	9	8,0	8,0	72,3
3	26	23,2	23,2	95,5
4	5	4,5	4,5	100,0
Total	112	100,0	100,0	

- Para 3 clusters:

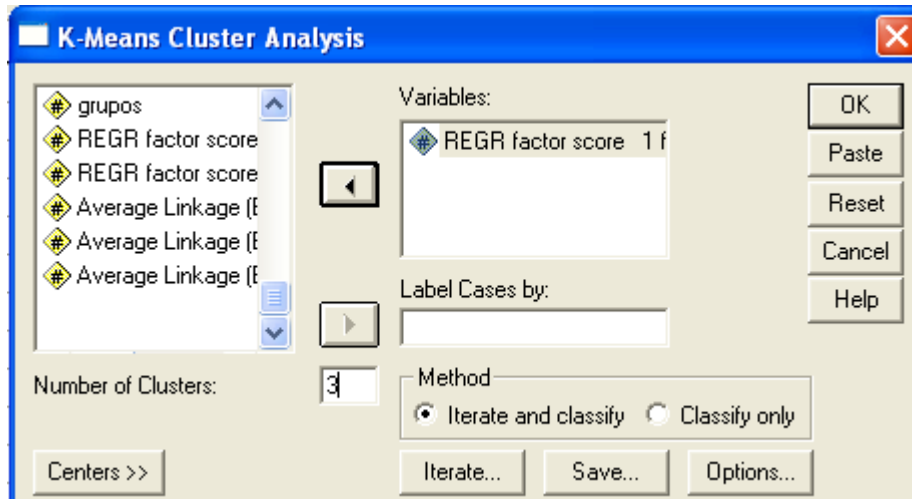
Figura 42:**Average Linkage (Between Groups)**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1	98	87,5	87,5	87,5
2	9	8,0	8,0	95,5
3	5	4,5	4,5	100,0
Total	112	100,0	100,0	

ANEXO 14: ANÁLISIS DE CLUSTER DE K-MEDIAS con 3 cluster.

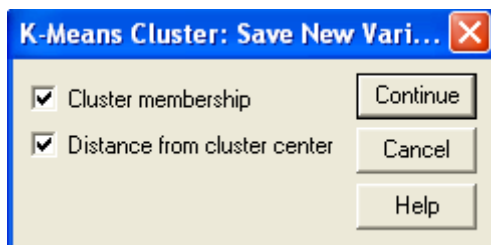
En el menú Analyze seleccionar: Classify→K-means Cluster

Figura 43:



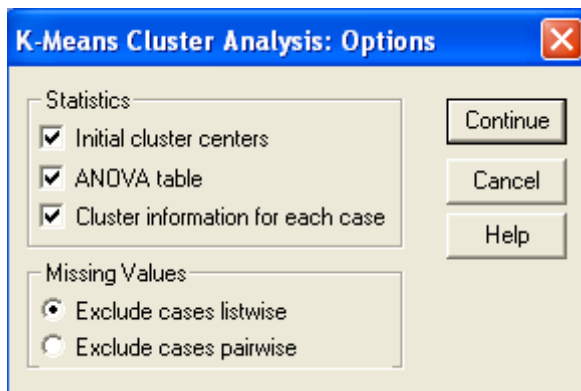
1. En la pestaña Save seleccionar Cluster membership y Distance from cluster center (pertenencia al cluster y distancia desde el centro del cluster, respectivamente)

Figura 44:



2. En Options,

Figura 45:



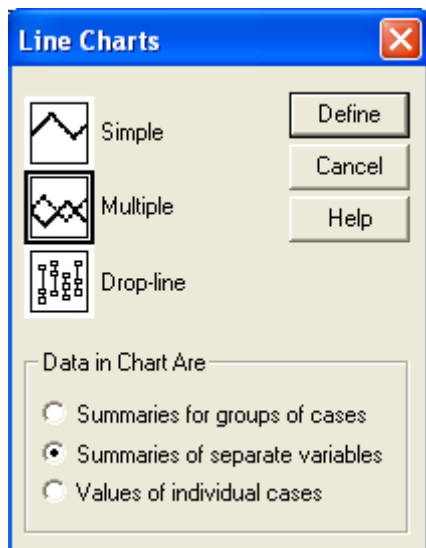
3. Finalmente OK.

ANEXO 15: CARACTERIZACIÓN DE LOS GRUPOS

15.1 GRÀFICO DE LÍNEAS:

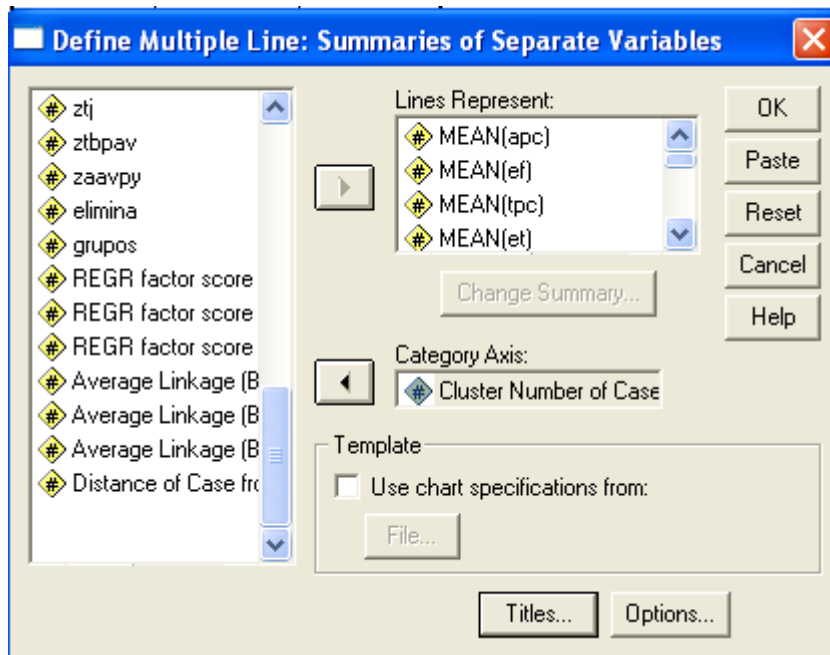
1. En Graphs→Line Charts→Multiple:

Figura 46:



2. Define→ se seleccionan todas las variables originales y la variable que identifica los grupos:

Figura 47:

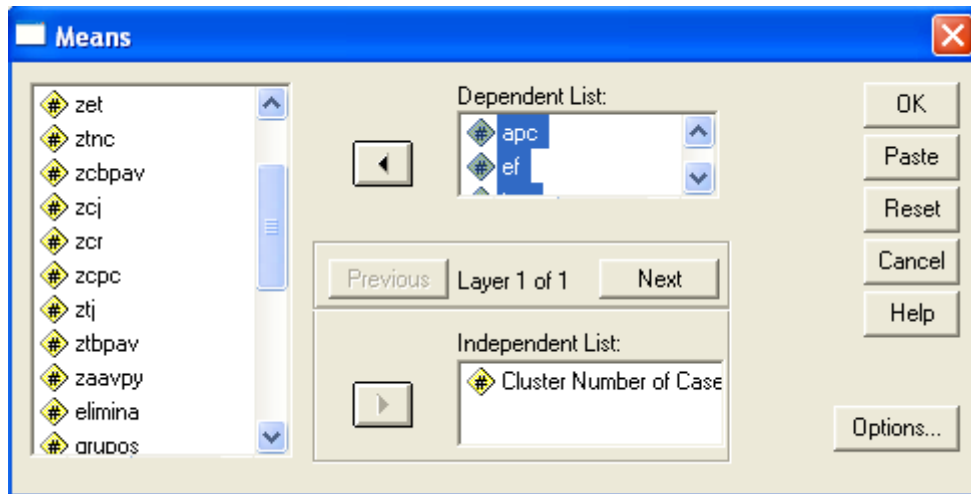


3. OK→ Analizar los resultados.

15.2 Medias

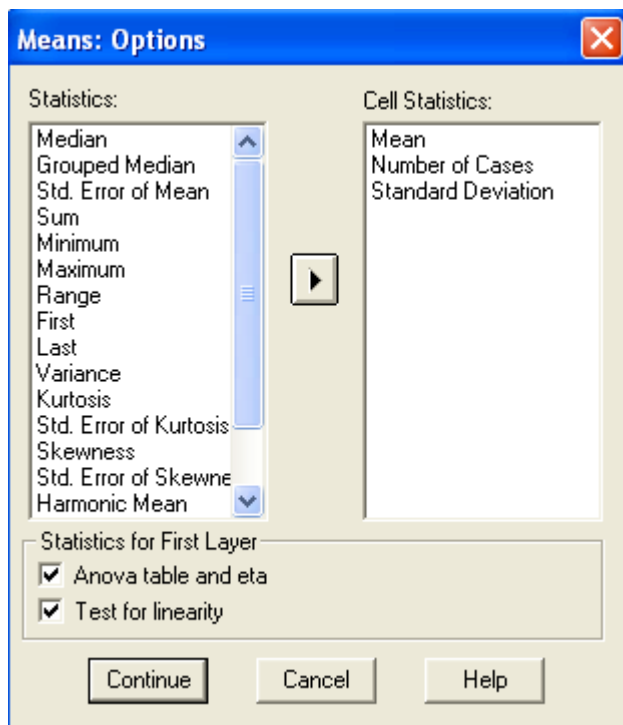
1. Seleccionar Analyze→Compare Jeans→Means

Figura 48:



2. Options:

Figura 49:

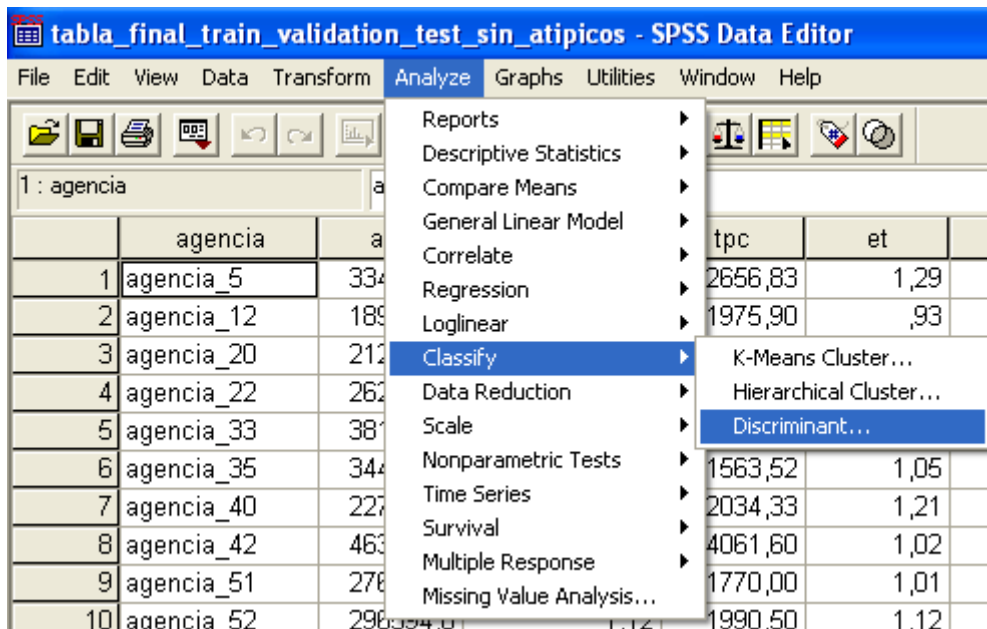


3. OK. Los resultados son: la matriz de medias, la tabla ANOVA.

ANEXO 16: ANÁLISIS DISCRIMINANTE

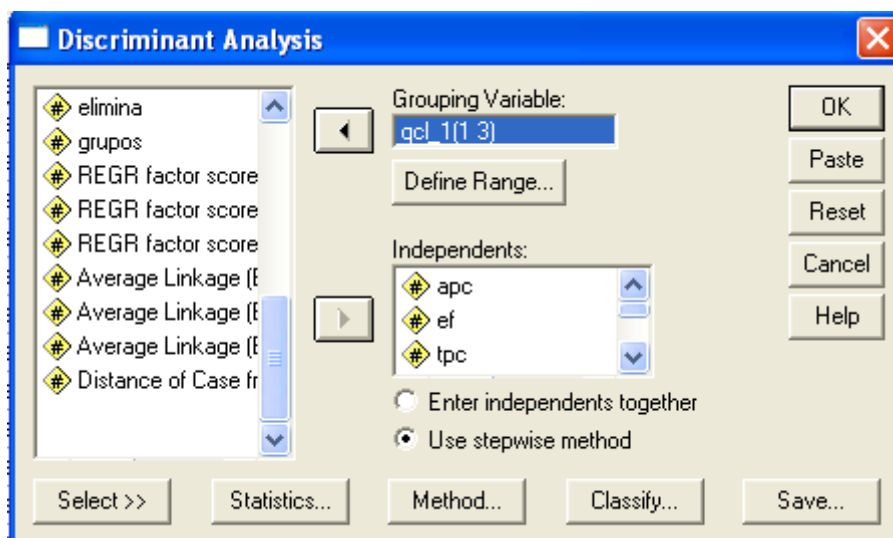
1. Seleccionar en la barra de menú: Analyze→Classify→Discriminant

Figura 50:



2. Aparece el cuadro de diálogo:

Figura 51:

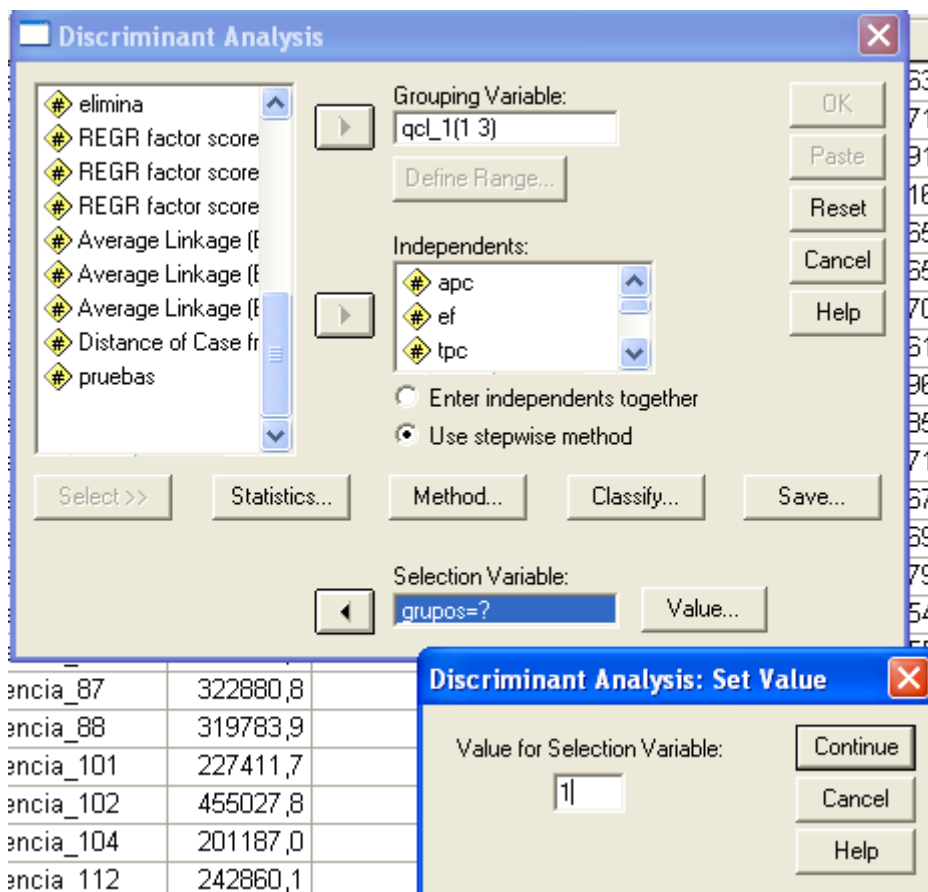


Donde: en Grouping Variable se muestra la variable de agrupamiento, en este caso obtenida en el análisis cluster y se define el rango o número de grupos; en Independents se muestran las variables independientes y debe ser mínimo una. La

lista independents con las variables independientes presenta dos opciones para definir el método de introducción y extracción de variables independientes:

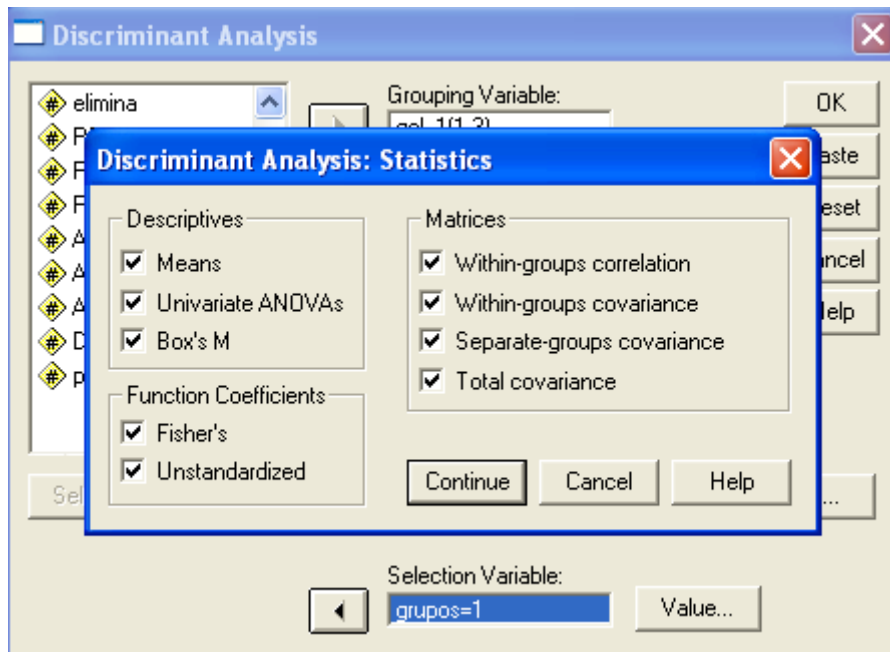
- Introducir todas las variables independientes juntas
 - Usar Stepwise o método paso a paso: se utiliza para controlar la introducción y extracción paulatinas de variables independientes. La introducción se realiza en función del nivel de significancia de la variable independiente.
3. Pulsar el botón select, donde aparece la opción Selection Variable, aquí se limita el Análisis Discriminante al subconjunto de casos que tengan un valor numérico específico para una variable de selección, puede ser cualitativa o cuantitativa. En este caso el subconjunto objetivo es el de entrenamiento, que se identifica en la variable grupo con el código 1.

Figura 52:



4. En el botón Statistics se observa:

Figura 53:



Estas opciones permiten obtener tablas de estadísticos descriptivos:

- Means (medias). Proporciona medias, desviación típica de cada variable discriminante por el total de los datos y para cada grupo de la variable dependiente.
- Univariate ANOVAs (ANOVAs Univariados). Realiza y muestra un ANOVA de una vía con cada una de las variables discriminantes, en relación con los grupos definidos por las variables independientes. Su función es contrastar la hipótesis nula de que las medias de cada una de las variables independientes son iguales para todos los grupos.
- Box's M (test de M de Box). Permite contrastar la H_0 de igualdad de Matrices entre grupos, es decir las matrices de covarianza para cada grupo son muestras extraídas de la misma población. Para muestras suficientemente grandes, un "p" no significativo indica que hay suficiente muestras suficientemente grandes, un "p" no significativo indica que hay suficiente evidencia de las matrices difieren. ES sensible a salidas de normalidad multivariada.

Además, los coeficientes de Función, que permite valorar la fuerza de la función de clasificación.

- Fisher's (coeficientes estandarizados de clasificación de Fisher). Pueden usarse directamente para la clasificación. Permite obtener un conjunto de coeficientes

para cada grupo. Y se asigna un caso al grupo que tiene la mayor puntuación discriminante.

- Unstandardized (coeficientes no estandarizados de clasificación). Estos coeficientes sólo pueden usarse para calcular las puntuaciones discriminantes no estandarizadas.

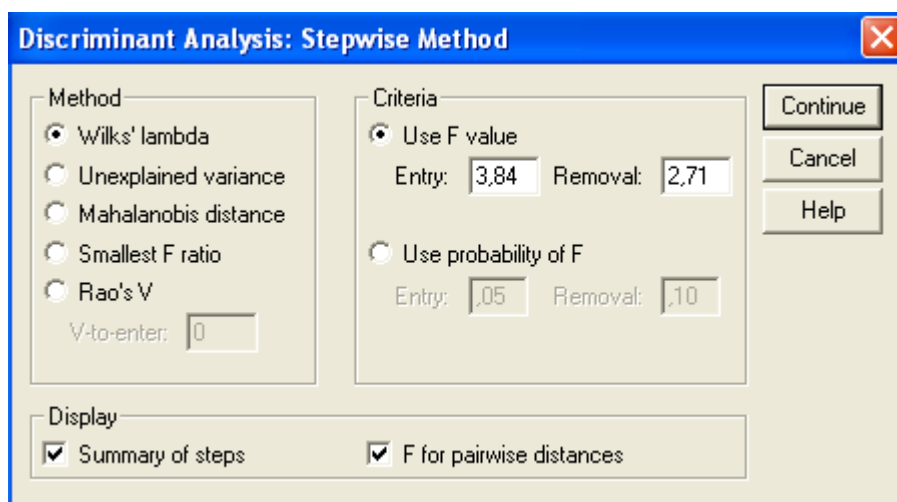
En Matrices, se muestra la información matricial de los coeficientes discriminantes para las variables independientes.

- Within Groups correlation (Matriz de correlaciones dentro de grupos). Muestra una matriz conjunta de correlaciones a través de todos los grupos. Se obtiene promediando las matrices separadas de covarianzas de todos los grupos antes de calcular las correlaciones.
- Within-groups covariance (Matriz de covarianza dentro grupos). Muestra una matriz conjunta a través de todos los grupos, que podría diferir de la matriz total de covarianzas. Se obtiene promediando las matrices separadas de covarianzas de todos los grupos.
- Separate-groups covariance (Matrices de covarianza de grupos separados). Muestra una matriz de covarianzas para cada grupo.
- Total covariance (Matriz total de covarianza). Muestra una matriz de covarianzas para todos los casos de la muestra como un todo, es decir, como si fueran una sola muestra.

Pulsar el botón Continue

5. Activada la opción Use Stepwise Method aparece habilitado el botón Method. Al activarlo aparece:

Figura 54:



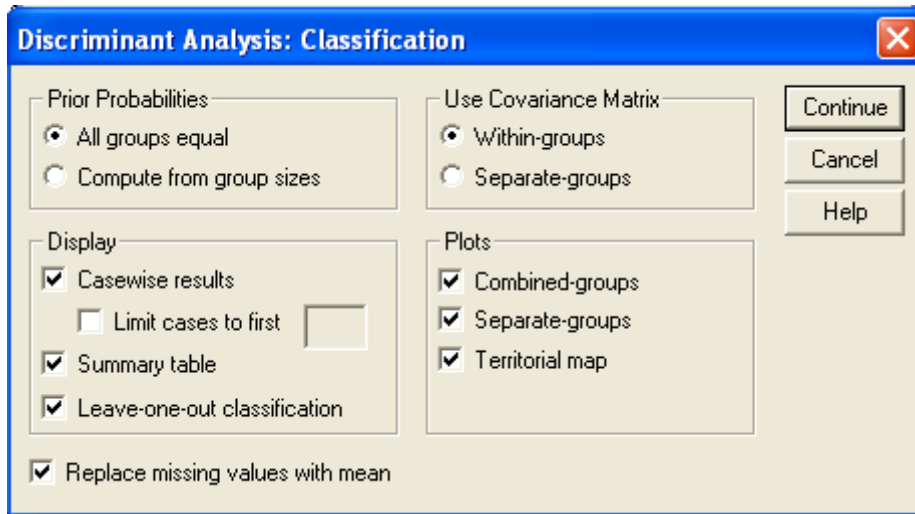
Method (Método). Permite especificar los criterios para la selección de las variables independientes.

- Wilks' Lambda (Lambda de Wilks). Es la opción por defecto. Se eligen para el análisis de las variables independientes que tengan el menor valor del estadístico Lambda de Wilks. En cada paso de este proceso, se introduce de este modo la variable independiente que minimiza el Lambda global de Wilks.
- Unexplained variance (Varianza no explicada). En cada paso, se introduce la variable independiente que minimiza la suma de la variación no explicada entre los grupos.
- Mahalanobis distance (Distancia de Mahalanobis). En cada paso, se introduce la variable independiente que maximiza la distancia de Mahalanobis para los grupos más cercanos. Ofrece una medida del grado de diferencia de los valores de un caso respecto de la media de todos los casos. Una distancia de Mahalanobis identifica un caso que tenga valores extremos sobre una o más variables independientes.
- Smallest F ratio (Mínima Razón de F). En cada paso, se introduce la variable independiente que maximiza la mínima razón de F entre cualquier par de grupos. Se calcula a partir de la distancia de Mahalanobis entre grupos.
- Rao's V (V de Rao). También llamada Traza de Lawley-Hotelling, mide las diferencias entre las medias de los grupos. En cada paso, se introduce la variable independiente que maximiza el incremento de V de Rao. Al activar esta opción se habilita el cuadro V-to-enter (Valor a introducir) que especifica el incremento mínimo de V para que una variable independiente se introduzca en el análisis. El valor p por defecto es cero.

Display (Mostrar). Permite especificar la visualización de la información adicional del análisis:

- Summary of steps (resumen de pasos). Seleccionada por defecto, proporciona tablas para la lambda de Wilks, las variables independientes introducidas/extraídas, las variables independientes del análisis y las variables independientes fuera del análisis. Muestra la tolerancia de todas las variables independientes y los estadísticos utilizados para la selección de las vi. Se muestran también los valores F, niveles de significancia y la tolerancia mínima.
 - F for pairwise distances (F para distancias emparejadas). Muestra una matriz de razones de la F emparejadas de cada par de grupos. Son tests de significación de las distancias Mahalanobis entre grupos. También es seleccionada.
6. Para la configuración de criterios de clasificación pulsar el botón Classify:

Figura 55:



Proporciona el control de los criterios usados en la clasificación de casos y de las salidas ofrecidas a partir de la fase de clasificación. Permite controlar el cálculo de probabilidades previas, obtener salidas de resumen de clasificación y controlar la clasificación de casos con valores perdidos.

En Prior Probabilities (Probabilidades previas). Establece las probabilidades previas: éstas se basan en la proporción muestral de casos de cada grupo.

- All groups equal (Todos grupos iguales). Considera iguales las probabilidades previas de pertenencia al grupo son iguales. Dejar activada sólo esta opción.
- Computes from group size (Calcular a partir de tamaños de grupos). Calcula las probabilidades previas a partir de la proporción muestral de casos de cada grupo. No utilizar esta opción si hubiesen fuertes diferencias en los tamaños muestrales.

En Display (Mostrar). Se dispone de las siguientes opciones:

- Casewise results (Resultados según casos). Muestra, para cada caso, códigos del grupo observado, el grupo predicho, las probabilidades posteriores y las puntuaciones discriminantes. Al activar esta opción se habilitan las siguientes:
 - Limit cases to first (Limitar casos a primero). Limita la salida a los primeros n caso. Se puede especificar un número personalizado de casos en el cuadro de texto adjunto.
- Summary table (Table resumen). También llamada matriz de confusión, muestra los casos correcta e incorrectamente clasificados para cada uno de los grupos basados en el Análisis Discriminante.

- Leave-one-out classification (Clasificación dejar-uno-fuera). También conocido como método U, clasifica cada caso por las funciones discriminantes derivadas de los demás casos.
- Se seleccionan todos los casos excepto limit cases to first.

Use Covarianza Matriz (utilizar la matriz de covarianzas). Permite usar la matriz de covarianzas del proceso.

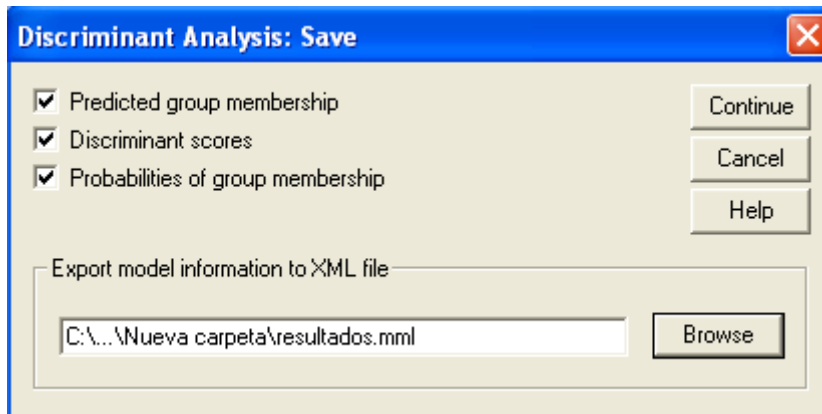
- Within-groups (Dentro de grupos). Utiliza matriz de covarianzas dentro de grupos para clasificar casos. Sólo se activa esta función
- Separate-groups (Grupos separados). Utiliza las matrices de covarianza de grupos separados para clasificar los casos. Como la clasificación se basa en las funciones discriminantes y no en las variables originales, esta opción no es siempre igual a la discriminación cuadrática.

Plots (Gráficos). Permite la presentación de los siguientes gráficos, se seleccionan todas las opciones:

- Combined-groups (Grupos combinados). Presenta para todos los grupos juntos un diagrama de dispersión para los valores de las dos primeras funciones discriminantes, o un histograma, si sólo hay una función discriminante.
- Separate-groups (grupos separados). Presenta diagramas de dispersión para cada grupo por separado, para los valores de las dos primeras funciones discriminantes.
- Territorial map (mapa territorial). Es un gráfico de fronteras utilizado para clasificar casos en grupos basados en los valores de la función discriminante.
- Pulsar continue.

7. La configuración de almacenamiento de nuevas variables aparece al pulsar el botón Save:

Figura 56:



Aparece el cuadro de diálogo: Discriminant Análisis: Save New Variables, permite guardar la información del proceso o de la puntuación discriminante para cada caso, como nuevas variables discriminantes.

- Predicted groups membership (Pertenencia al grupo predicho). Crea una sola variable discriminante, que indica la pertenencia predicha de cada caso a uno de los grupos de variable dependiente. Es decir, el grupo con la mayor probabilidad posterior, en base a las puntuaciones discriminantes.
- Discriminantes scores (Puntuaciones discriminantes). Crea una variable para cada función discriminante. Estas se calculan multiplicando los coeficientes discriminantes no estandarizados por los valores de las variables independientes, sumando estos productos y añadiendo después la constante. La puntuación media entre todos los casos es cero y la varianza conjunta dentro de los grupos es 1.
- Probabilities of group membership (Probabilidad de pertenencia al grupo). Crea una variable para cada uno de los grupos de la variable dependiente que indica la probabilidad con la que cada caso ha sido asignado a un grupo de la variable dependiente. Son probabilidades condicionadas. La primera variable contiene la probabilidad a posteriori de pertenencia al primer grupo, la segunda, la probabilidad de pertenencia al segundo, etc.
- Y se crea un archivo externo en Export model information xml file
- Pulsar el botón continue.

8. Resultados del análisis discriminante, ver el anexo 16.

9. Resultados del análisis discriminante de los grupos de entrenamiento y validación respecto al grupo de prueba ver el anexo 17.

10. Como resultado de la ejecución del análisis discriminante, se guardan una serie de variables nuevas dentro de la base de datos:

- Dis_1: identifica a cual de los 3 grupos pertenece cada individuo en función del modelo generado por la función discriminante.
- Dis1_1 y dis2_1 son, respectivamente, los valores para cada individuo de las puntuaciones discriminantes para las funciones 1 y 2.
- Dis1_2, dis2_2, dis3_2 estas 3 variables identifican, respectivamente, las probabilidades para cada individuo de ser asignado al grupo 1,2 y 3 en función del modelo generado por el análisis discriminante. Estas variables son complementarias entre sí, es decir, suman uno.

Figura 53:

The screenshot shows the SPSS Data Editor window titled 'tabla_final_train_validation_test_sin_atipicos - SPSS Data Editor'. The main data grid is visible, showing a table with 30 rows and 14 columns. The columns are labeled as follows: 'dis_1', 'dis1_1', 'dis2_1', 'dis1_2', 'dis2_2', 'dis3_2', 'dis_2', 'dis1_3', 'dis2_3', 'dis1_4', 'dis2_4', 'dis3_4', and 'var'. The 'var' column contains values ranging from 0.0000 to 0.0444. The rows represent individual cases, with the first row (case 1) having values: 1.00, 1.80648, 1.61793, .94062, .00000, .05938, 1.00, 2.55331, 1.72089, .99556, .00000, .00444. The table is displayed in 'Data View' mode.

	dis_1	dis1_1	dis2_1	dis1_2	dis2_2	dis3_2	dis_2	dis1_3	dis2_3	dis1_4	dis2_4	dis3_4	var
1	1,00	1,80648	1,61793	,94062	,00000	,05938	1,00	2,55331	1,72089	,99556	,00000	,00444	
2	1,00	1,18095	,69452	,72994	,00003	,27003	1,00	1,09228	,11766	,57315	,00000	,42685	
3	1,00	2,46610	1,33425	,98400	,00000	,01600	1,00	2,60836	,97496	,99374	,00000	,00626	
4	1,00	1,93971	-,38682	,90464	,00000	,09535	1,00	2,39993	-,34162	,97366	,00000	,02634	
5	1,00	2,03845	-,49394	,91885	,00000	,08115	1,00	2,18022	-,88129	,93343	,00000	,06657	
6	1,00	1,61484	,76968	,89001	,00000	,11999	1,00	2,38996	,66788	,98603	,00000	,01397	
7	1,00	2,00320	-,44732	,91428	,00000	,08572	1,00	3,10092	,01039	,99894	,00000	,00306	
8	3,00	,03860	1,28128	-,21050	,00439	,78511	1,00	,74870	1,82569	,61875	,00004	,38121	
9	3,00	,69166	-,27599	,38009	,00019	,61972	1,00	1,07148	,36588	,59940	,00000	,40059	
10	1,00	1,87712	-,10754	,90970	,00000	,09030	1,00	3,37192	,55256	,99899	,00000	,00101	
11	1,00	2,28835	-,106375	,93758	,00000	,06242	1,00	1,97593	-,171135	,82055	,00000	,17945	
12	1,00	1,88751	-,57522	,92568	,00000	,07432	1,00	2,58769	,47295	,99075	,00000	,00925	
13	1,00	1,58336	-,77890	,78529	,00000	,21471	1,00	2,77374	-,04267	,99219	,00000	,00781	
14	1,00	3,53681	-,57193	,99797	,00000	,00203	1,00	4,01015	,14737	,99977	,00000	,00023	
15	1,00	2,09949	-,28095	,93395	,00000	,06605	1,00	1,71226	-,89298	,79194	,00000	,20806	
16	1,00	1,68125	-,34583	,84423	,00000	,15577	1,00	2,88235	-,06361	,99413	,00000	,00587	
17	1,00	1,30389	,40301	,75966	,00002	,24032	1,00	1,24082	,18353	,67922	,00000	,32078	
18	1,00	1,73774	,15052	,88252	,00000	,11748	1,00	1,90499	-,29597	,90631	,00000	,09369	
19	1,00	3,06101	,75245	,99457	,00000	,00543	1,00	3,52617	,65937	,99939	,00000	,00061	
20	1,00	1,88032	,65270	,92672	,00000	,07328	1,00	2,27536	-,21505	,97434	,00000	,02566	
21	1,00	2,01494	,61932	,94397	,00000	,06503	1,00	1,30536	-,25705	,65352	,00000	,34648	
22	1,00	1,75695	1,66466	,93528	,00000	,06472	1,00	2,44686	1,82567	,99365	,00000	,00634	
23	1,00	1,12665	1,48956	,76751	,00005	,23244	1,00	2,67370	1,72470	,99581	,00000	,00419	
24	1,00	3,99899	,73802	,99923	,00000	,00077	1,00	4,05272	,41508	,99963	,00000	,00017	
25	1,00	2,54650	-,588363	,79945	,00000	,20055	1,00	2,45986	-,589658	,51561	,00000	,48439	
26	1,00	1,38818	1,05728	,83245	,00001	-,16754	1,00	2,75356	1,70483	,99742	,00000	,00258	
27	1,00	2,67686	-,21114	,98138	,00000	,01862	1,00	2,03877	-,103024	,89560	,00000	,10440	
28	3,00	-,19416	1,95016	,33089	,00332	,66580	1,00	2,75082	3,06452	,99895	,00000	,00105	
29	1,00	1,86728	3,49467	,97482	,00000	,02518	1,00	2,04127	2,64042	,99014	,00000	,00986	
30	1,00	2,89552	-,08043	,98907	,00000	,01093	1,00	3,49095	-,17387	,99882	,00000	,00118	

ANEXO 17: ENCUESTA

Cuánto tiempo tiene su o sus cuentas en el banco?

años meses

Usted considera que la atención de los asesores es:
 Usted considera que la atención de los cajeros es:
 Usted considera que la atención del banco en general es:

	muy mala	mala	buen	muy buena	excelente

Usted recomendaría la atención del Banco a un amigo?

.....

Nombre espontáneamente tres bancos donde le gustaría aperturar una cuenta

- 1
- 2
- 3

Cuáles son los principales problemas que ha observado en el banco, mencione 3, por favor?

- 1
- 2
- 3

EDAD
 SEXO
 SITUACIÓN LABORAL
 INGRESOS

menor a 500
 de 500 a 1000
 superior a 1000

NÚMERO DE FORMULARIO

REGIÓN DÓNDE SE LEVANTA LA INFORMACIÓN

NOMBRE DEL ENCUESTADOR