UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE INGENIERIA ECONOMICA Y CIENCIAS SOCIALES

INFORME DE SUFICIENCIA

PROPUESTA DE SEGMENTACIÓN CON ANALISIS MULTIVARIADO DE CLIENTES PREFERENTES DE TELEFONICA DEL PERU

PARA OBTENER EL TÍTULO PROFESIONAL DE LICENCIADO EN ESTADISTICA

LEGUIA LOAYZA CESAR AUGUSTO

Lima- Perú

2004

A mi esposa, Carolina, por su apoyo incondicional, por su cariño, respeto y colaboración. Por compartir las alegrías, las penas y por tener siempre palabras de aliento. Porque todo junto a ella se hace más fácil.
A mis hijas, Alejandra y Valeria, por darme todo su cariño y por contagiarme sus alegrías.

A mis padres, por hacerme conocer a Dios, y por enseñarme que debemos ser perseverantes, no importa las circunstancias que nos depare la vida.

UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE INGENIERÍA ECONÓMICA Y CCSS

ESCUELA PROFESIONAL DE ESTADÍSTICA

PROPUESTA DE SEGMENTACIÓN CON ANÁLISIS MULTIVARIADO DE CLIENTES PREFERENTES DE TELEFÓNICA DEL PERÚ

Alumno: César Augusto Leguía Loayza

Asesor: Víctor Sánchez

En los últimos años, se ha hablado de un rápido y permanente cambio que vive nuestra sociedad, situación en la que estamos inmersos y no podemos escapar. Dicho cambio es ciertamente una evolución dinámica en todos los ámbitos de nuestra realidad mundial, cuya expresión más evidente son los innumerables avances tecnológicos en el campo de las denominadas tecnologías de información, las cuales se van a convertir de hecho, en motor del crecimiento y la trasformación económica, además de configurarse como uno de los soportes básicos de la sociedad del siglo XXI.

Acorde con estos tiempos de cambio, las telecomunicaciones en los últimos años han experimentado en nuestro país un desarrollo muy considerable y unos esfuerzos de inversión muy importantes, permitiendo una mayor competencia y como consecuencia una progresiva oferta de productos y servicios diferenciados. Vivimos hoy una competencia global. No solamente con relación al abarcamiento geográfico, sino también a la vigilia de la competencia. Ante esta situación el conocimiento individual del cliente es fundamental. La mejor manera de adecuarse a las preferencias individuales es identificando grupos de clientes con preferencias similares. Este proceso de identificación de grupos de consumidores con las mismas preferencias se conoce como segmentación de mercados. Las segmentaciones de mercado permiten llevar a cabo programas de marketing ventajosos, tal y como el diseño de productos específicos para distintos segmentos, es decir, ofrecer servicios diferenciados a los clientes de acuerdo al nivel de rentabilidad que tienen.

El Objetivo de este estudio es proponer una metodología para la segmentación de clientes en general y su respectiva aplicación práctica a los mejores clientes residenciales de Telefónica del Perú (Clientes Preferentes). Dicho se a de paso, estos clientes para mantener la condición de preferentes tienen que cumplir una serie de condiciones, entre ellas tenemos el nivel de facturación en la suma de sus servicios de Telefonía Local, Larga Distancia, Televisión por Cable, Telefonía Móvil y Terra (Acceso a Internet). Así, el presente informe pretende utilizar herramientas de Análisis Multivariado en la búsqueda de una propuesta de clasificación de los Clientes Preferentes, así como un modelo predictivo de clasificación. Para este análisis utilizamos todas las variables corporativas disponibles que se consideraron importantes para el análisis. Sin embargo, existe una especial consideración por los ratios de rentabilidad, a través de los cuales podemos diferenciar rápidamente a los clientes.

INDICE

Introducción .....

Resumen Ejecutivo .....

CAPÍTILO I

1. Población objetivo 2
2. Justificación 3
3. Objetivo del análisis 3
4. Variables incluidas en el estudio 4
5. Medotología 5
   - 5.1. Población 5
   - 5.2. Marco teórico 5
   - 5.3. Soporte de la investigación 6
     - 5.3.1. Análisis previo de los datos 6
       - 5.3.1.1. Datos perdidos 7
       - 5.3.1.2. Casos atípicos 7
     - 5.3.2. Supuestos del Análisis Multivariante 8
     - 5.3.3. Pruebas de Normalidad 8
       - 5.3.3.1. Análisis gráfico de la Normalidad 9
       - 5.3.3.2. Test estadístico de la Normalidad 9

CAPITULO II

   - 5.4. Análisis Factorial (AF) 10
     - 5.4.1. Supuestos del AF 11
     - 5.4.2. Aplicabilidad del AF 11
     - 5.4.3. Estimación de los factores y valoración del ajuste final 13
     - 5.4.4. Criterio para el cálculo del número de factores 13
       - 5.4.4.1. Criterio de la Raíz Latente o Autovalores 13
       - 5.4.4.2. Criterio a priori 14
       - 5.4.4.3. Criterio del Porcentaje de la Varianza 14

5.4.5. Interpretación de los factores....15
5.4.6. Rotación de factores....15
5.4.7. Criterio para la significación de las cargas factoriales....17
5.4.8. Uso de las combinaciones lineales....18

CAPITULO III

5.5. Análisis Cluster (AC)....20
5.5.1. Representación Gráfica....20
5.5.2. Objetivos....20
5.5.3. Selección de las variables del AC....21
5.5.4. Diseño de investigación para el AC....21
5.5.5. Detección de Datos Atípicos....21
5.5.6. Medidas de similitud y estandarización de variables....21
5.5.7. Supuestos del AC....22
5.5.7.1. Representatividad de la muestra....22
5.5.7.2. Multicolinealidad en el AC....22
5.5.8. Proceso de segmentación....23
5.5.9. Interpretación de los conglomerados y perfil del cliente....25-29

CAPÍTULO IV

5.6. Análisis Discriminante(AD)....30
5.6.1. Representación gráfica del AD....30
5.6.2. Objetivos....31
5.6.3. Selección de las variables dependientes e independientes....31
5.6.4. Tamaño muestral....31
5.6.5. Supuestos del AD....32
5.6.6. Estimación del modelo y ajuste global....33-36
5.6.6.1. Test que validan la aplicabilidad del análisis....
5.6.6.2. Test de igualdad de medias....
5.6.7. Cálculo de las puntuaciones Z discriminantes....36

5.6.8. Cálculo de las probabilidades posteriores 40
5.6.9. Conclusiones. 40

CAPITULO V

5.7. Análisis de Correspondencia Múltiple (ACM) 40
5.7.1. Objetivos del ACM 41
5.7.2. Supuestos del ACM 41
5.7.3. Análisis de la matriz de indicadores.....43
5.7.4. Mapas perceptuales.....44
5.7.5. Obtención de resultados 45

CAPITULO VI

6. Conclusiones y recomendaciones 45-46
Bibliografía .....
Anexos .....
Matriz de Correlaciones MC.....
Sintaxis de las corridas en SPSS.....

INTRODUCCIÓN

En los últimos años, se ha hablado de un rápido y permanente cambio que vive nuestra sociedad, situación en la que estamos inmersos y no podemos escapar. Dicho cambio es ciertamente una evolución dinámica en todos los ámbitos de nuestra realidad mundial, cuya expresión más evidente son los innumerables avances tecnológicos en el campo de las denominadas tecnologías de información, las cuales se van a convertir de hecho, en motor del crecimiento y la trasformación económica, además de configurarse como uno de los soportes básicos de la sociedad del siglo XXI.

Acorde con estos tiempos de cambio, las telecomunicaciones en los últimos años han experimentado en nuestro país un desarrollo muy considerable y unos esfuerzos de inversión muy importantes, permitiendo una mayor competencia y como consecuencia una progresiva oferta de productos y servicios diferenciados. Vivimos hoy una competencia global. No solamente con relación al abarcamiento geográfico, sino también a la vigilia de la competencia. Ante esta situación el conocimiento individual del cliente es fundamental. La mejor manera de adecuarse a las preferencias individuales es identificando grupos de clientes con preferencias similares. Este proceso de identificación de grupos de consumidores con las mismas preferencias se conoce como segmentación de mercados. Las segmentaciones de mercado permiten llevar a cabo programas de marketing ventajosos, tal y como el diseño de productos específicos para distintos segmentos, es decir, ofrecer servicios diferenciados a los clientes de acuerdo al nivel de rentabilidad que tienen.

Existen muchas técnicas univariadas para segmentar a los clientes; ya sea por rangos de facturación, por rentabilidad o tenencia de servicios, etc. En el mejor de los casos utiliza un Clustering. Sin embargo, la cuestión práctica nos indica que son muchas las variables que contribuyen al proceso de segmentación. Ante esta situación, nos inclinamos a usar los Análisis Multivariados como una alternativa sólida y de fácil uso en todo proceso de segmentación.

Finalmente, quiero adelantarme que a lo largo de este informe encontrarán una forma peculiar de afrontar el análisis, intentando siempre ser lo más pragmático posible, sin dejar de lado el sustento teórico, razón de muchos despropósitos que espero puedan comprender.

RESUMEN EJECUTIVO

El desarrollo económico, el crecimiento de la competencia, hace que hoy las empresas tengan la necesidad de conocer más a sus clientes, es decir, conocer más de cerca sus necesidades, sus aspiraciones, sus costumbres, hábitos de consumo, metas, etc. Consecuencia de ello, la tendencia actual es identificar a sus mejores clientes, es decir, a los clientes más rentables. Desde otro punto de vista, serían los primeros que la competencia quisiera tenerlos como clientes. Ante esta amenaza, el conocimiento individual del cliente es fundamental. Dado lo difícil y costoso que sería obtener tal información, se intentará formar grupos con comportamientos muy parecidos que nos permitan gestionar de manera semejante dichos grupos. Este proceso de identificación de grupos de consumidores con las mismas preferencias se conoce como segmentación. La segmentación permite llevar a cabo programas de marketing focalizados, ahorrando costos, evitando la comunicación masiva, etc. Esto significa programas de marketing ventajosos, la atención diferenciada de acuerdo por ejemplo al nivel de rentabilidad.

Existen muchas técnicas univariadas para segmentar a los clientes utilizando una única variable. Como ejemplo tenemos, rangos de facturación, rentabilidad por cliente, tenencia de servicios, en el mejor de los casos Clustering. Sin embargo, la cuestión práctica nos indica que son muchas las variables que contribuyen al proceso de segmentación. Ante esta situación, nos inclinamos a usar los Análisis Multivariados.

El Objetivo de este estudio es proponer una metodología para la segmentación de clientes en general y su respectiva aplicación práctica a los mejores clientes residenciales de Telefónica del Perú (Clientes Preferentes). Dicho se a de paso, estos clientes para mantener la condición de preferentes tienen que cumplir una serie de condiciones, entre ellas tenemos el nivel de facturación en la suma de sus servicios de Telefonía Local, Larga Distancia, Televisión por Cable, Telefonía Móvil y Terra (Acceso a Internet). Así, el presente informe pretende utilizar herramientas de Análisis Multivariado en la búsqueda de una propuesta de clasificación de los Clientes Preferentes, así como

un modelo predictivo de clasificación. Para este análisis utilizamos todas las variables corporativas disponibles que se consideraron importantes para el análisis. Sin embargo, existe una especial consideración por los ratios de rentabilidad, a través de los cuales podemos diferenciar rápidamente a los clientes.

El informe comprende seis capítulos muy bien definidos. El primero, orientado al análisis exploratorio de la información, es decir, búsqueda de datos atípicos, datos perdidos, estandarización de la información así como pruebas de Normalidad. El segundo capítulo, está reservado para el Análisis Factorial que nos permitió reducir la dimensión de las variables originales. En tercer lugar, se encuentra el Análisis Cluster que utilizando los factores hallados en el Análisis Factorial nos consintió formar grupos homogéneos al interior del grupo y heterogéneos entre ellos. Cuarto capítulo, una vez clasificados los clientes diseñamos un modelo predictivo de clasificación, a través del cual hallamos la función discriminante que nos acceda identificar el error cometido a través del Análisis Cluster y clasificar a los clientes a posteriori. En el quinto capítulo, complementaremos el análisis con la búsqueda del perfil de los clientes a través del Análisis de Correspondencia Múltiple. En el sexto capítulo, las conclusiones y recomendaciones.

En la segmentación obtenida de todo el estudio (Tesis) podemos destacar que en la primera agrupación (Cluster1) el 10% de los mejores clientes residenciales que representan el 25% de la rentabilidad, 21% de la facturación total y 41% de la facturación en Telefonía Móvil de este segmento. De la misma manera podríamos detallar que el segundo segmento (Cluster2) esta compuesto por el 17% de los clientes, representan el 22% de la rentabilidad, el 26% de facturación total y alrededor del 30% de la facturación en el consumo de Larga Distancia de este segmento. El tercer segmento (Cluster3) esta compuesto por el 25% de los Clientes Preferentes, representan el 25% de la rentabilidad total, 20% de la facturación total y 42% de la facturación en Televisión por Cable. Por último, el cuarto y último segmento (Cluster4) esta compuesto por los clientes menos

rentables, en porcentaje representa el 48% de los clientes que son la mayoría, 29% de la rentabilidad total y 29 %de la facturación total.

La propuesta de segmentación funcionó con la aplicación de una serie de Análisis Multivariado de manera secuencial, la cual adquirió una mayor eficacia a diferencia de cualquier metodología de segmentación, inclusive si comparamos con la aplicación de manera separada. Por otro lado esta metodología también puede ser aplicada en diferentes empresas del mercado, especialmente las vinculadas al servicio.

CAPÍTULO I

I.1. Población Objetivo

Nuestra población objetivo son los mejores clientes residenciales de Telefónica del Perú (Clientes Preferentes) Creado en el año de 1999 con alrededor de 15,000 clientes importantes, cifra que terminó incrementándose luego de identificar aquellos clientes que cumplían con todas las condiciones. Estos clientes son aquellos que en la suma de sus servicios en Telefonía Fija, Telefonía Móvil, Televisión por Cable e Internet tienen una facturación promedio mayor a $125 durante los últimos 12 meses, en uno o más servicios residenciales. Actualmente estos clientes representan el 5% de la facturación total y el 25% de la rentabilidad de los ingresos del mercado residencial de Telefónica del Perú.

Veamos algunas cifras que nos puedan hacer conocer mejor a este segmento.

Facturación total por Unidad de Negocio

Facturación Total. **28,243**	*Facturación T. Local.*	*68%*
	Facturación Larga Dist.	*10%*
	Facturación Cable	*13%*
	Facturación Móvil	*8%*
Cantidades en miles de soles Octubre 2003	*Total*	*100%*

Facturación promedio por cliente

Facturación Total por cliente **449,41**	*Facturación T. Local*	*69%*
	Facturación Larga Dist.	*10%*
	Facturación Cable	*12%*
	Facturación Móvil	*8%*
	Facturación Terra	*1%*
Cantidades en soles Octubre 2003	*Total*	*100%*

I.2. Justificación

Telefónica del Perú tiene alrededor de 2 millones de líneas, permitiendo mejor servicio y mayor cobertura. Esto significa enormes esfuerzos de inversión muy importantes. En este sentido, se ha pasado de un monopolio a un estado de libre mercado, lo que significa que los nuevos competidores del mercado están al acecho de los mejores clientes, es decir, los que consumen más, los que pagan mejor, en otras palabras los más rentables. Es importante entonces el trabajo de investigación, el cual nos permitirá conocer mejor a los clientes y sobre todo proteger a los más importantes. Además, este conocimiento contribuye a una mejor planificación de las campañas de marketing, campañas de fidelización, acciones de retención de clientes, planificación estratégica de la empresa en general.

I.3. Objetivos

Los objetivos que se pretende alcanzar con este trabajo de tesis, son:

Objetivos principales

- Segmentar a los mejores clientes residenciales de Telefónica del Perú (Clientes Preferentes) en grupos o segmentos con características homogéneas.

Objetivos secundarios

- Diseñar un modelo predictivo que clasifique a los clientes nuevos en el segmento apropiado.
- Determinación de las variables más relevantes por segmento.

Identificar a los clientes candidatos a ser dados de baja.

I.4. Variables incluidas en el estudio

Las variables que se consideraron para el estudio son tipo corporativas. Estas variables tienen que ver con el uso (Tráfico), antigüedad, tenencia, baja del servicio y el consumo de los servicios de

Telefonía Fija, Telefonía Móvil, Televisión por Cable, Larga Distancia e Internet. No sé consideraron otro tipo de variables debido a que no se contaba con ellas.

El total de las variables vamos a agruparlas de acuerdo a la unidad de negocio por cuestiones prácticas:

I. Servicios y consumo de Telefonía Fija

1. QSERV_BA Cantidad de servicios de Telefonía Fija.
2. QSVA_BAS Cantidad de SVA de Telefonía Fija.
3. ANTIGUED Antigüedad de Telefonía Fija.
4. BASICA B Cantidad de bajas de Telefonía Fija.
5. RFAC BAS Rentabilidad para Telefonía Ffija.
6. FACT_BAS Facturación en Telefonía Fija.

II. Servicios y consumo de Telefonía Móvil

7. QSERV_MO Móviles cantidad de servicios móviles.
8. ANTI MOV Antigüedad de móviles.
9. MOV BAJA Cantidad de bajas de móviles.
10. RFAC MOV Rentabilidad para móviles.
11. FACT_MOV Facturación en móviles.

III. Servicios y consumo de Televisión por Cable

12. QSERV_CA Cantidad de servicios de Cable.
13. QSVA_CAB Cantidad de SVA de Cable.
14. ANTI CAB Antigüedad de Cable.
15. CABLE_BA Cantidad de bajas de Cable.
16. RFAC CAB Rentabilidad para Cable.
17. FACT_CAB Facturación en Cable.

IV. Servicios y consumo de Larga Distancia

18. LDISTANC Cantidad de Bajas de Larga Distancia.

19. RFAC LDI	Rentabilidad para Larga Distancia Internacional.
20. FACT LDI	Facturación en Larga Distancia Internacional
21. RFAC_LDN	Rentabilidad para Larga Distancia Nacional.
22. FACT_LDN	Facturación en Larga distancia Nacional.

I.5. Metodología

En esta parte del informe intentaremos explicar de manera clara los procedimientos que contiene el análisis, así como las herramientas que se utilizarán para conseguir estos objetivos.

I.5.1. Población

En el quehacer diario siempre se está buscando información para tomar decisiones acertadas. Esta búsqueda puede tener innumerables problemas, como que la información que se requiere generalmente está referido a un número grande de unidades y si se quiere hacer un análisis exhaustivo, tendremos problemas de costo y tiempo, con el riesgo de equivocarnos. Ante esta situación, podríamos imaginarnos que el muestreo es la única salida. Sin embargo, como en nuestro caso estamos tomando una parte de la población que vendría ser los mejores clientes residenciales de Telefónica del Perú. Esta división se da por cuestiones estrictamente comerciales y a las cuales nos tenemos que ajustar.

I.5.2.Marco teórico

Existen diferentes métodos de segmentación y sin duda su aportación es importante para las diferentes áreas del marketing. Inicialmente la clasificación de los clientes se realizaba por rangos de facturación, la cual fue importante en su momento porque te permitía diferenciar a los clientes que más facturan. Sin embargo, al considerar una sola variable traía como consecuencia algunos errores de clasificación porque todo el análisis estaba centrada en una sola variable. Posteriormente y con la intención de considerar más de una variable se utilizaron los conceptos de Scoring que nos permitía incluir en la formula muchas más variables, ponderadas por la importancia para el investigador. De la misma

manera, los conceptos de rentabilidad ayudaron a mejor la clasificación de los clientes (Segmentación de Valor) Finalmente, se ve la necesidad de clasificar a los clientes por otros criterios que no tienen directamente que ver con la facturación o rentabilidad (Segmentación Comportamental) sino por variables como Tráfico, Deuda, antigüedad, tenencia de servicios, etc. Es allí donde la calidad de las técnicas multivariadas cobran mayor importancia, todo esto respaldado por supuesto gracias al avance del uso de las computadoras.

I.5.3. Soporte teórico de la investigación

Nuestro principal soporte teórico está basado en las Técnicas Multivariadas, de mucha utilidad en las áreas de Marketing y Ventas. Estas técnicas serán aplicadas en el orden siguiente:

1. Análisis exploratorio de la información
2. Análisis Factorial.
3. Análisis Cluster.
4. Análisis Discriminante.
5. Análisis de Correspondencia Múltiple.

Óptica de la investigación

La realización de este proyecto será teórico práctico, buscando afianzar el uso de las Técnicas Multivariadas para la segmentación de clientes. Realizar esta investigación es todo un reto del cual estoy seguro salir airoso. Siempre estuve convencido que el trabajo de sustentación de la tesis elegida tendría que ser un problema real, es decir, que este trabajo sirva como base para nuevas metodologías prácticas para la toma de decisiones y que signifique no solamente un provecho personal sino una contribución a la "Minería de Datos"

I.5.3.1. Análisis previos de los datos

El análisis previo de los datos, también denominado Análisis Exploratorio de la Información, es un paso necesario, imprescindible, conduce a una mejor

predicción y una mejor evaluación de la dimensionalidad. En esta primera etapa pusimos el mayor cuidado, habitualmente esta primera etapa es descuidada por los investigadores.

Existen diferentes métodos, uno de ellos es el método gráfico que nos permitió comprender las características de los datos y sus relaciones subyacentes de la variable.

I.5.3.1.1. Datos perdidos

A pesar de todos los controles de calidad de información que podamos poner en la data no podríamos evitar la presencia de datos perdidos o datos ausentes. Por esta razón, el reto consistió en que la ausencia de esta información no afecten a los resultados del análisis.

Lo primero que tuvimos que hacer es encontrar razones por lo que estos datos estaban ausentes y ver si son relevantes a la hora de tomar decisiones. Particularmente tuvimos muy pocos datos perdidos, lo que no nos permitió generar reglas de mejoras o algún patrón en los datos ausentes que nos pueda ayudar a entender y evitar dicho proceso.

Para solucionar el tema de los datos perdidos o datos ausentes usamos una metodología que es muy práctica y dado que la cantidad de datos perdidos es mínima, optamos por reemplazarlos por el promedio del resto de las observaciones que si tienen correctamente esta información. Como ejemplo podemos decir que, si un cliente no tiene la antigüedad de su servicio lo reemplazaremos por el promedio de la antigüedad del resto de las observaciones. De la misma manera, si un cliente "Activo" no tiene facturación, lo reemplazaremos con el promedio de facturación del resto de las observaciones.

I.5.3.1.2. Casos atípicos

Los casos atípicos son observaciones con características identificables que las diferencia claramente de las otras observaciones. Como en nuestro caso no surgieron por error de un determinado procedimiento, entonces no lo excluimos

del análisis. En el caso que así fuere, evaluaríamos la permanencia en el análisis. Si las observaciones no tienen explicación, intentaríamos omitirlos a menos que se considere que representen a un segmento de la población.

Como ejemplo hemos comparado dos variables: Facturación promedio de seis meses en Telefonía Local y número de Teléfonos Fijos que posee el cliente. Podemos apreciar gráficamente que existen datos atípicos por cliente, es decir, altos promedios de facturación y alta tenencia de Telefonía Fija en comparación con el resto de observaciones como se muestra encerrado en los círculos verdes.

Distribución normal Fact_Local Fijos

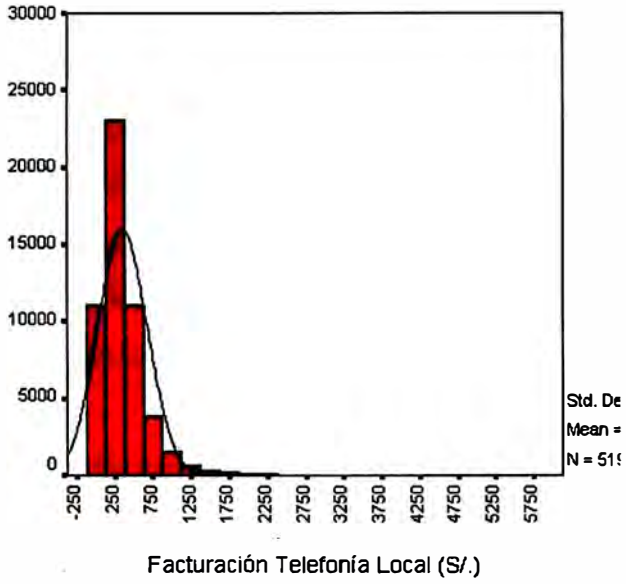


Gráfico 1

Fact_Local–Número Teléfonos

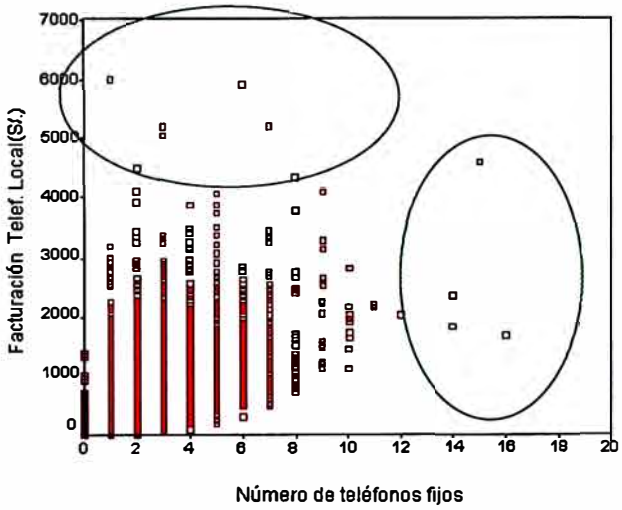


Gráfico 2

I.5.3.2. Supuestos del Análisis Multivariante

La necesidad de asegurarnos mejores resultados nos permite comprobar los supuestos de Normalidad que tienen que ver con la propia complejidad de las relaciones, y debido fundamentalmente a la gran cantidad de variables que estamos usando en este análisis.

5.3.3. Pruebas de Normalidad

El supuesto más importante del Análisis Multivariante es la Normalidad de los datos, en referencia a la distribución de los datos para una única variable métrica y su correspondencia con una Distribución Normal. Tanto los métodos estadísticos univariantes como los multivariantes se basan en el supuesto de la Normalidad Univariante. Si una variable es una Normal Multivariante, entonces sus componentes siguen una distribución Normal. Sin embargo, lo contrario no es necesariamente cierto.

I.5.3.3.1 Análisis gráfico de la Normalidad

El método más simple para diagnosticar la Normalidad de la variable es una combinación visual de histogramas que compara los valores de los datos observados con una Distribución Normal (Para muestras grandes). La Distribución es Normal si la línea que representan la distribución real de los valores sigue de cerca la diagonal.

Veamos como ejemplo la variable Antiguedad_Ba que mide la antigüedad del servicio de Telefonía Fija.

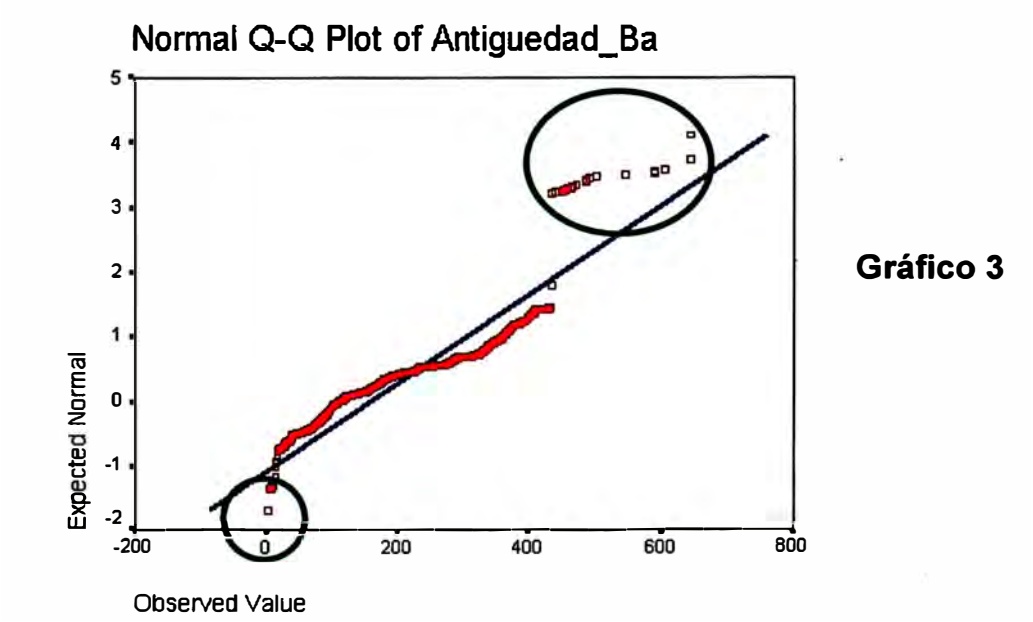


Gráfico 3

Como vemos en el gráfico adjunto, la variable no sigue una Distribución Normal del todo en un primer momento, algunas observaciones encerradas en los círculos verdes son nada menos que los datos atípicos, los cuales dicho sea de paso serán analizados por separado. Análogamente, podríamos analizar toda las variables pero que por temas de espacio solo analizaremos una variable a manera de ejemplo.

I.5.3.3.2. Test estadístico de normalidad

De la misma manera que en el Análisis Gráfico de Normalidad escogimos a manera de ejemplo cuatro variables para evaluar la Normalidad a través del test estadístico de Kolgomorov- Smirnov.

Tests of Normality

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Cuentas_Basica	.317	51991	.000
Fact_Bas	.139	51991	.000
Cuentas_Cab	.341	51991	.000
Fact_Cab	.234	51991	.000

a. Lilliefors Significance Correction

Tabla 1

Se comprueba en el Gráfico 3 que, en efecto, podemos suponer que la distribución de estos datos es aproximadamente Normal, ya que los puntos se aproximan bastante a la línea roja. Por otro lado, en el Tabla 1 podemos ver el Test Kolgomorov- Smirnov que nos afirma que las variables elegida siguen una Distribución Normal (p-valor < NS). Por cuestiones prácticas no estamos presentando la prueba por cada una de las variables comprometidas sino solamente de cuatro variables o que no significa que la prueba no se haya echo para el resto de variables.

II.5.4. ANÁLISIS FACTORIAL(AF)

Esta técnica ha experimentado una utilización creciente durante la última década en todas las áreas de la investigación, especialmente en el área empresarial. El propósito general de esta técnica analítica es encontrar una manera de condensar la información contenida en una serie de variables originales en una serie de dimensiones más pequeñas denominadas valores teóricos o factores, con una mínima pérdida de información.

El Análisis Factorial encuentra factores mucho más pequeños en número a las variables originales, las cuales forman los factores a través de combinaciones lineales que pueden utilizarse en Análisis Multivariantes posteriores. El propósito es retener la naturaleza y el carácter de las variables originales con una mínima pérdida de información.

Representación de los factores:

$$
\begin{aligned}
F_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
F_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \qquad i = 1,\ldots,m \quad j = 1,\ldots,p \\
&\vdots \\
F_m &= a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mp}X_p
\end{aligned}
$$

Donde:

F_i = Son los factores que representan a las variables originales

X_j = Son las variables originales que intervienen en el estudio.

a_{ij} = Pesos de la combinación lineal.

II.5.4.1. Supuestos en el Análisis Factorial

Si bien es cierto, los supuestos del Análisis Multivariante anteriormente descritos son importantes para el análisis, estos supuestos son más de tipo conceptual que estadístico y el obviar cualquiera de ellos produce una disminución en las correlaciones observadas. En realidad, solo es necesario la Normalidad cuando se aplica una prueba estadística, es más, es deseable que haya cierto grado de multicolinealidad, dado que el objetivo es identificar serie de variables correlacionadas.

II.5.4.2. Aplicabilidad del análisis

Test que evalúan la aplicabilidad del modelo.

- **Matriz de Correlaciones**

  Ver Anexo 1 **(Tabla 2)**

- **Test de esfericidad de Bartlett.**

  Es una prueba que verifica si existe correlación lineal entre las variables. Esta verificación se hace utilizando el valor obtenido por el estadístico de Bartlett y luego contrastándolo con el valor de la Chi-cuadrado.

- **Índice KMO (Káiser-Meyer-Olkin) de adecuación de la muestra.**

  Es una medida que nos indica el grado de correlación entre las variables y la conveniencia de utilizar el Análisis Factorial en el estudio. Este índice tiene la siguiente regla:

  - Si **KMO**<0.5 no resultaría aceptable para hacer un Análisis Factorial.
  - Si 0.5<**KMO**<0.6 grado de correlación medio, entonces habría aceptación media.
  - Si **KMO**>0.6 indica alta correlación y por lo tanto es conveniente realizar un Análisis Factorial.

Test de KMO y Bartlett

Indice KMO de Medida de adecuación muestral		.597
Test de Esfericidad de Bartlett	Approx. Chi-Cuadrado	349640.3
	gl	136
	Sig.	.000

Tabla 3

De estos resultados se puede concluir:

- Un valor alto de la estadística de prueba de Bartlett favorecerá el rechazo del la hipótesis nula (Las variables no están correlacionadas) Si esta hipótesis no puede rechazarse, deberá ponerse en duda lo adecuado del Análisis Factorial.
- Como, $0.5 \prec KMO = 0.597 \prec 0.6$, indica correlación media, por lo tanto, abría una aceptación media del uso del Análisis Factorial en el estudio.

II.5.4.3. Estimación de los factores y la valoración del ajuste general

Para realizar esta operación, tomamos la decisión de utilizar el Análisis de Componentes Principales por ser este método el que mejor se acomoda para reducir la dimensión de las variables de estudio. De esta manera desestimamos el Análisis Factorial Común (AFC).

II.5.4.4. Criterio para el cálculo del número de factores a ser extraídos

Para decidir cuantos factores debemos extraer, tenemos que empezar con algún criterio determinado. Hemos escogido 3 métodos para el cálculo que compararemos . Son el porcentaje de varianza total, el criterio de raíz latente o el criterio a priori.

II.5.5.4.1. Criterio de la raíz latente o autovalores

Es una de las técnicas que más se utiliza y muy sencilla de aplicar. La lógica que se usa en esta técnica, es que cualquier factor individual debería explicar la varianza de por lo menos una única variable. Cada variable contribuye con un valor de uno para el autovalor total. Por tanto, solo se consideran los factores que tienen raíces latentes o autovalores mayores que uno, los factores con raíces latentes menores que uno explican menos de una variable, no son significativas y se desestima la incorporación al análisis.

El criterio que se usa para el cálculo del número de factores a ser extraídos es el "Criterio de la raíz latente". En forma práctica, este criterio consiste en encontrar el punto de corte, en el cual la curva cambia de sentido, es decir de cóncava a convexa o viceversa. Además, para dicho criterio sólo se consideran los factores que tienen raíces latentes o autovalores mayores que 1. Por otro lado, el investigador puede elegir el número de factores de acuerdo a la exigencia del análisis. En la figura se observa que el punto de corte se ubica en el **factor 5**, el criterio del gráfico de sedimentación sugiere utilizar 5 componentes.

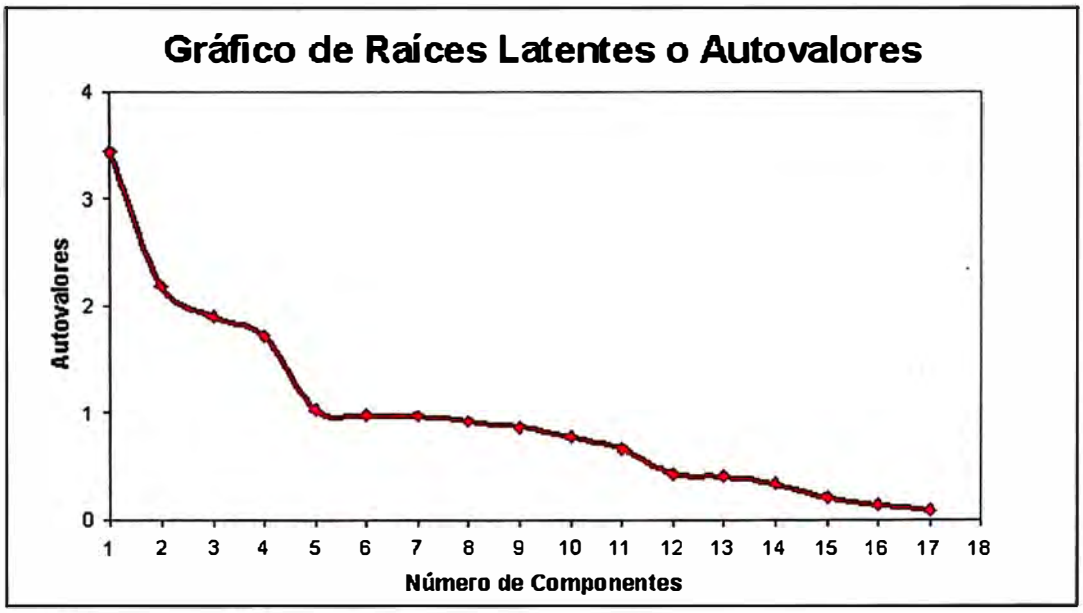


Gráfico 4

Para definir el número de factores a elegir no se consideraron las cinco variables de rentabilidad por cada unidad de negocio.

II.5.4.4.2. Criterio a Priori

En este criterio simplemente se decide el número de factores que se quiere obtener. Esta técnica sirve para replicar un trabajo de investigación donde se tiene que elegir una cantidad de factores usada en otra investigación. Adicionalmente, la decisión de cuantos factores utilizar tiene que ver con el criterio del investigador.

II.5.4.4.3. Criterio de porcentaje de la varianza

Esta metodología utiliza la varianza total extraída como parámetro de determinación. El propósito es asegurar una significación práctica de los factores derivados, asegurando que expliquen un porcentaje de la varianza. No se tiene un solo criterio absoluto de hasta cuanta varianza se puede explicar, la selección tiene que ver con el criterio del investigador y la importancia de la investigación.

La tabla 5 adjunta muestra la aplicación del Análisis de Componentes Principales, la cual podemos observar que el primer factor (Factor 1) presenta una varianza de 20.22% de la varianza total. De modelo similar, el segundo factor (Factor 2) representa el 12.8% de la varianza total y así sucesivamente hasta tomar en cuenta el criterio del valor propio (>1) que sugiere utilizar 5 componentes. Adicionalmente, el criterio de la varianza explicada (> 60%) nos confirma utilizar 5 factores.

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	3.437	20.220	20.220	3.437	20.220	20.220	2.839	16.697	16.697
2	2.179	12.819	33.039	2.179	12.819	33.039	2.396	14.094	30.791
3	1.900	11.178	44.216	1.900	11.178	44.216	1.861	10.948	41.740
4	1.719	10.110	54.327	1.719	10.110	54.327	1.642	9.661	51.401
5	1.024	6.025	60.352	1.024	6.025	60.352	1.522	8.951	60.352
6	.977	5.745	66.097						
7	.972	5.718	71.815						
8	.916	5.389	77.204						
9	.864	5.081	82.285						
10	.772	4.540	86.824						
11	.659	3.876	90.701						
12	.428	2.519	93.220						
13	.401	2.361	95.581						
14	.333	1.958	97.538						
15	.204	1.201	98.739						
16	.135	.792	99.531						
17	7.966E-02	.469	100.000						

Extraction Method: Principal Component Analysis.

Tabla 5

II.5.4.5. Interpretación de los factores

Para interpretar los factores y seleccionar la solución factorial, inicialmente calculamos la matriz de factores no rotados, la cual nos de una idea preliminar acerca del número de factores a extraer y contiene las Cargas Factoriales. En este primer paso estamos simplemente interesados en encontrar la mejor combinación lineal de las variables que cuentan con el mayor porcentaje de varianza.

Component Matrix [a]

	Component				
	1	2	3	4	5
Cuentas_Basica	.51	-.15	.56	.34	-.22
Fact_Bas	.43	-.17	.50	.50	.02
Cuentas_Cab	.76	-.31	-.23	-.17	.04
Fact_Cab	.80	-.34	-.25	-.25	.12
Fact_Ldn	.06	-.08	.20	.37	.52
Fact_Ldi	.20	-.12	.16	.36	.36
Cuentas_Mov	.49	.75	-.17	.14	-.04
Fact_Mov	.45	.73	-.15	.15	.02
Antiguedad_Ba	.32	-.26	.13	.26	-.59
Antiguedad_Ca	.69	-.23	-.16	-.18	-.08
Antiguedad_Mo	.45	.68	-.15	.12	-.11
Bajas_Ba	.10	.21	.58	-.63	.12
SVA_Ba	.21	-.03	.16	.27	.21
Bajas_Ca	-.12	.22	.31	-.11	.11
SVA_Ca	.58	-.26	-.23	-.27	.18
Bajas_Mo	.20	.19	.12	-.16	.28
Bajas_LD	.24	.15	.73	-.50	-.13

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Tabla 6

El objetivo de las soluciones factoriales no rotadas es la reducción de los datos, pero tenemos que percatarnos si esta solución no rotada facilita la interpretación más adecuada de las variables examinadas. Por experiencia se podría afirmar que la mayor parte de veces las soluciones factoriales se tiene que rotar. La rotación es deseable porque simplifica la estructura de los factores y habitualmente es difícil determinar si los factores no rotados serán significativos.

En segunda instancia, hacer uso de cualquier método de rotación implica lograr soluciones factoriales más simples, fáciles de interpretar y más significativas. Dependiendo del grado de aportación de la rotación se puede replantear el modelo eliminando variables.

II.5.4.6. Rotación de Factores

Herramienta de mucha importancia a la hora de interpretar los factores. Veamos la parte práctica, se gira cierto ángulo en el origen de los ejes de referencia hasta alcanzar una determinada posición, si este ángulo alcanza 90 grados se denomina rotación ortogonal, si es oblicua la rotación puede adquirir varios métodos de rotación:

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Cuentas_Basica	.131	.071	.205	.739	.382
Fact_Bas	.068	.041	.089	.582	.599
Cuentas_Cab	.849	.085	-.039	.135	.073
Fact_Cab	.940	.067	.004	.059	.085
Fact_Ldn	-.055	-.045	-.008	-.096	.663
Fact_Ldi	.066	-.004	-.049	.043	.576
Cuentas_Mov	.086	.922	.037	.012	.028
Fact_Mov	.059	.875	.046	-.030	.073
Antiguedad_Ba	.146	-.014	-.169	.736	-.091
Antiguedad_Ca	.733	.118	.010	.206	-.028
Antiguedad_Mo	.076	.840	.029	.068	-.028
Bajas_Ba	.052	-.019	.886	-.076	-.095
SVA_Ba	.051	.075	.009	.102	.406
Bajas_Ca	-.222	.041	.346	-.067	.059
SVA_Ca	.744	.017	.028	-.069	.058
Bajas_Mo	.137	.179	.296	-.172	.159
Bajas_LD	.060	.010	.883	.284	-.098

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

a. Rotation converged in 6 iterations.

Tabla 7

El método de rotación Varimax es el que usé porque generalmente es el de mayor uso por los investigadores. Este método lo que hace es minimizar las variables con cargas altas en un factor, mejorando la interpretación de los factores.

Component Transformation Matrix

Component	1	2	3	4	5
1	.778	.446	.131	.348	.240
2	-.394	.849	.249	-.219	-.120
3	-.315	-.199	.738	.449	.338
4	-.355	.189	-.595	.370	.589
5	.119	-.071	.145	-.702	.683

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

Tabla 8

En la tabla 8 presentamos la matriz de trasformación que nos permite rotar y tener una mejor perspectiva para la interpretación de los resultados.

II.5.4.7. Criterio para la significación de las Cargas Factoriales

Para interpretar los factores, se escogió uno de los tantos métodos prácticos que en torno a las Cargas Factoriales existen. La siguiente exposición considera diversos aspectos relativos a la significación práctica y estadística, además del número de variables, que afectan a la interpretación de las cargas factoriales.

- Las cargas mayores a +/- 0.3 están en el nivel mínimo.
- Las cargas de +/- 0.4 se consideran más importantes
- Las cargas de +/- 0.5 o mayores, se consideran significativas.

Cuanto mayor sea el tamaño absoluto de las Cargas Factoriales, más importante resulta la carga al interpretar la matriz factorial. La importancia de las cargas la hemos diferenciado de acuerdo a los colores, es decir, las cargas factoriales de color negro no son significativas, las de color azul se consideran importantes y las de color rojo son prácticamente significativas.

Component Score Coefficient Matrix

	Component				
	1	2	3	4	5
Cuentas_Basica	-.047	.003	.069	**.421**	.116
Fact_Bas	-.056	-.009	.021	.273	**.312**
Cuentas_Cab	**.306**	-.020	-.030	-.013	.001
Fact_Cab	**.350**	-.036	.000	-.084	.027
Fact_Ldn	-.022	-.039	.017	-.212	**.516**
Fact_Ldi	.008	-.024	-.018	-.103	**.415**
Cuentas_Mov	-.030	**.393**	-.015	-.007	-.015
Fact_Mov	-.034	**.373**	-.006	-.046	.031
Antiguedad_Ba	-.025	-.004	-.142	**.547**	-.239
Antiguedad_Ca	.253	.003	-.012	.072	-.087
Antiguedad_Mo	-.034	**.360**	-.022	.047	-.066
Bajas_Ba	.034	-.043	**.489**	-.097	-.036
SVA_Ba	-.006	.015	.003	-.024	.274
Bajas_Ca	-.082	.017	.193	-.059	.069
SVA_Ca	.293	-.045	.022	-.150	.042
Bajas_Mo	.058	.051	.170	-.195	.153
Bajas_LD	-.006	-.030	**.463**	.163	-.120

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

Tabla 9

II.5.4.8. Uso de las combinaciones lineales

Para identificar combinaciones lógicas de variables, entonces basta con la interpretación de los factores, pero como nuestro objetivo va más allá de esto, identificaremos variables apropiadas para aplicaciones subsiguientes de otras técnicas estadísticas. Para esta nueva alternativa examinaremos la matriz factorial y seleccionaremos las variables con la mayor Carga Factorial sobre cada factor para que actúe como variable suplente del factor. Esto es fácilmente identificable cuando algunas de las cargas destacan considerablemente sobre otras, en caso contrario pueden ser mucho más difíciles. Aquí es donde entra la experiencia del investigador donde a pesar que una variable tiene mayor Carga Factorial se tenga que obviarlo y si considerar una variable con menor carga factorial.

Cada uno de los factores representa una **combinación lineal** de las variables que han intervenido en el estudio de acuerdo a las cargas factoriales.

Posteriormente estos factores representativos los emplearemos en el Análisis Cluster.

I \ J	1	2	3	4	5
1	-0.0467	0.0029	0.0692	0.4207	0.1157
2	-0.0562	-0.0092	0.0214	0.2732	0.3115
3	0.3056	-0.0202	-0.0301	-0.0128	0.0012
4	0.3503	-0.0364	-0.0004	-0.0840	0.0269
5	-0.0220	-0.0387	0.0166	-0.2119	0.5156
6	0.0080	-0.0239	-0.0181	-0.1030	0.4150
7	-0.0303	0.3930	-0.0155	-0.0072	-0.0148
8	-0.0337	0.3727	-0.0055	-0.0457	0.0309
9	-0.0249	-0.0044	-0.1417	0.5475	-0.2394
10	0.2525	0.0035	-0.0117	0.0716	-0.0873
11	-0.0338	0.3604	-0.0218	0.0473	-0.0656
12	0.0340	-0.0433	0.4889	-0.0971	-0.0364
13	-0.0058	0.0153	0.0033	-0.0238	0.2743
14	-0.0816	0.0169	0.1928	-0.0586	0.0692
15	0.2931	-0.0455	0.0219	-0.1499	0.0422
16	0.0579	0.0514	0.1699	-0.1951	0.1533
17	-0.0057	-0.0299	0.4625	0.1630	-0.1204

Tabla 10

Cargas de los factores (P_{ij})

Cálculo de factores

$$F_j = \sum_{i=1}^{17} P_{ij} * \frac{(X_i - \mu_i)}{\sigma_i}$$

Donde j =1, 2, 3, 4, 5

Conclusiones y recomendaciones del Análisis Factorial

- Las 17 variables originales se han transformado en 5 factores. Estos factores explican el 60% de la Varianza Total.
- Para el Análisis de Clasificación utilizaremos como variables los factores hallados en el Análisis Factorial a pesar que esto puede interpretarse como una pérdida de información.
- Dado que no existen muchas variables con Cargas Factoriales significativas decidimos utilizar la totalidad de los factores con su respectiva carga factorial.

CAPÍTULO III

III.5.5. ANÁLISIS CLUSTER

El Análisis Cluster es una técnica que agrupa a los individuos o clientes en conglomerados de tal forma que los individuos dentro de cada grupo poseen características similares, es decir, alta homogeneidad dentro del Cluster y características diferentes a la de los otros grupos, alta heterogeneidad entre clusters.

III.5.5.1 Representación Gráfica

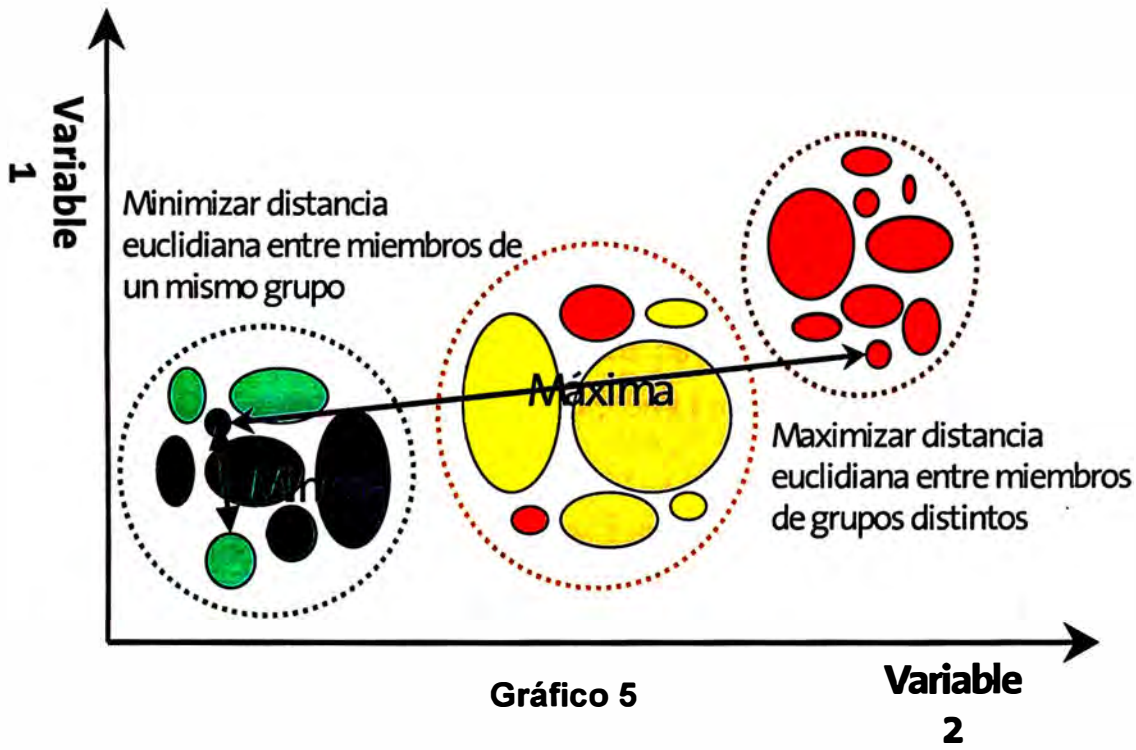


Gráfico 5

III.5.5.2 Objetivos

El objetivo más importante del Análisis Cluster es la segmentación de un grupo de individuos en dos o más segmentos de acuerdo a su similitud, es decir, características específicas similares.

III.5.5.3 Selección de variables del Análisis Cluster

La selección de las variables debe tomar en cuenta consideraciones teóricas, y prácticas, tomando en cuenta que las variables seleccionadas caractericen los objetos que estamos agrupando y que cumplan con el objetivo planteado para incluirlas, dado que esta técnica no tiene manera de diferenciar las variables relevantes de las irrelevantes.

Al tomar los factores del Análisis Factorial como variables seleccionadas para el Análisis Cluster, nos evitamos la eventualidad de seleccionar variables inapropiadas, es decir, variables redundantes (multicolinealidad) que podrían afectar drásticamente los resultados.

III.5.5.4 Diseño de investigación para el Análisis Cluster

Antes de todo proceso de clasificación tenemos que detectar los errores atípicos, definir las medidas de similitud de las observaciones y estandarizar las variables. Con esto estamos buscando una estructura de datos que se ajuste a la metodología seleccionada.

III.5.5.5 Detección de datos atípicos

Los datos atípicos como sabemos son observaciones totalmente disímiles con el resto de la población. La detección de este tipo de datos ya lo presentamos ampliamente como parte de la preparación de la información del Análisis Multivariado, por consiguiente ya no nos ocuparemos en esta parte.

III.5.5.6.1 Medidas de similitud

La medida de similitud entre objetos es una medida de correspondencia, o parecida, entre objetos que van a ser agrupados. Así como en otros Análisis

Multivariados utilizamos la Matriz de Correlaciones como medida de similitud o correspondencia.

Esta similitud para el caso del Análisis Cluster puede medirse de varias formas, siendo las medidas de Distancia Euclidiana la más utilizada. Esta distancia se utiliza para calcular medidas específicas y es la medida de distancia recomendada para los métodos de Análisis Cluster del Centroide y Ward.

III.5.5.6.2 Estandarización de variables

La forma más común de estandarización es la conversión de cada variable a una variable Z , restando la medida y dividiendo por la desviación típica de cada variable. Esta trasformación, a cambio, elimina el sesgo introducido por las diferentes unidades o mediciones de varios atributos de las variables utilizadas en el análisis.

III.5.6.7 Supuestos del Análisis Cluster

Las exigencias de Normalidad, Linealidad y Homocedasticidad que son tan importantes en otras técnicas multivariadas realmente tienen poco peso en el Análisis Cluster. Como tal, tiene fuertes propiedades matemáticas pero no fundamentos estadísticos. Sin embargo, debemos concentrarnos en dos asuntos críticos: La representatividad de la muestra y la multicolinealidad.

III.5.6.7.1 Representatividad de la muestra

Hablamos de representatividad cuando se trata de una muestra. En nuestro caso no tiene sentido este fundamento porque estamos trabajando sobre el total de población que contiene a todas las observaciones del estudio. Esto nos quita el peso de pensar en una muestra representativa.

III.5.6.7.2 Multicolinealidad en el Análisis Cluster

La Multicolinealidad es importante e imprescindible en algunas técnicas multivariantes a causa de la dificultad que podría ocasionar en el resultado final del análisis. Pero en el Análisis Cluster la Multicolinealidad no asume el

protagonismo que en otras técnicas multivariadas e inclusive nos arriesgamos a comentar que aquellas variables que son multicolineales solamente estarían ponderadas con más fuerza.

III.5.5.7 Proceso de segmentación

Test de Linealidad

Es el análisis de varianza (ANOVA) que proporciona información acerca de las variables y cuales son significativas, es decir, sabremos que variables contribuyen de un modo significativo al proceso de agrupación.

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
FAC1_1 * CLUSTER4	Between Groups	Linearity	2371.428	1	2371	2885	.000
FAC2_1 * CLUSTER4	Between Groups	Linearity	11201.0	1	11201	14740	.000
FAC3_1 * CLUSTER4	Between Groups	Linearity	343.419	1	343	346	.000
FAC4_1 * CLUSTER4	Between Groups	Linearity	5609.427	1	5609	6571	.000
FAC5_1 * CLUSTER4	Between Groups	Linearity	9149.064	1	9149	11546	.000

ANOVA (Test de Linealidad)

Observando el cuadro ANOVA podemos apreciar que todas las variables son significativas para el proceso de segmentación.

5.5.8. ¿Deben utilizarse los métodos jerárquicos o no jerárquicos?

Una de las dificultades que tuve era decidir si aplicaba el método jerárquico o no jerárquico: Finalmente tomé la decisión de utilizar ambos métodos para obtener los beneficios de cada uno. Veamos algunas consideraciones que se tomó en cuenta para elegir uno u otro método.

- Con la técnica jerárquica podemos establecer el número de Cluster

- El método jerárquico tiene la ventaja de ser más rápido, debido a que emplea menos recursos informáticos.
- Los métodos jerárquicos inicialmente pueden darnos una idea equivocada y llevar resultados artificiales.
- Los métodos jerárquicos tienen dificultades al aplicar muestras grandes.
- Los métodos no jerarquizados tienen varias ventajas sobre las técnicas jerárquicas. Los resultados son menos susceptibles a los datos atípicos, y a medida de la distancia utilizada.

5.5.8 ¿Cuantos grupos deben formarse?

No existe procedimiento o criterio estadístico objetivo estándar para saber la cantidad de grupos a formarse, es decir, no existe test de significación estadística como en otras técnicas multivariantes. Pero definitivamente, el mayor peso para decidir de cuantos segmentos tener es saber: ¿Que pretendemos hacer con ellos?, ¿Queremos tener una atención diferenciada por segmentos?, ¿Podemos ofrecer promociones y campañas diferenciadas?. Si no tenemos claro que es lo que vamos a gestionar con estas clientes mejor sería solo diferenciar a los mejores clientes del resto o simplemente no segmentar.

III.5.5.9 Interpretación de los conglomerados

La interpretación de los conglomerados implica la exploración de cada conglomerado en términos de cualidades características, es decir, el perfil de cada Cluster o asignar una etiqueta precisa que describa la naturaleza de los conglomerados.

CLUSTER4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	5260	10.1	10.1	10.1
	2	8579	16.5	16.5	26.6
	3	13163	25.3	25.3	51.9
	4	24987	48.1	48.1	100.0
	Total	51989	100.0	100.0	

Tabla 12

Perfiles de los clientes por Cluster dependiendo de las variables de facturación

CLUSTER	Estad.	Fact_ Bas	%	Fact_M ov	%	Fact_C ab	%	Fact_ Ldi	%	Fact_L dn	%	Fact_ Total	%	Rent_ Total	%
1	Suma	4039449	22%	834766	41%	603272	15%	438062	30%	201598	32%	5713923	21%	1627377	25%
	Media	768		159		115		83		38		1086		309	
	Clientes	5260	10%	5260	10%	5260	10%	5260	10%	5260	10%	5260	10%	5260	10%
2	Suma	3791412	21%	517334	25%	702034	17%	364454	25%	181354	29%	7003601	26%	1426203	22%
	Media	442		60		82		42		21		816		166	
	Clientes	8579	17%	8579	17%	8579	17%	8579	17%	8579	17%	8579	17%	8579	17%
3	Suma	4335903	24%	529879	26%	1735358	42%	294231	20%	95705	15%	5497770	20%	1640033	25%
	Media	329		40		132		22		7		418		125	
	Clientes	13163	25%	13163	25%	13163	25%	13163	25%	13163	25%	13163	25%	13163	25%
4	Suma	6047020	33%	154139	8%	1053151	26%	374685	25%	156622	25%	8958485	33%	1905436	29%
	Media	242		6		42		15		6		359		76	
	Clientes	24987	48%	24987	48%	24987	48%	24987	48%	24987	48%	24987	48%	24987	48%
Total	Suma	18213783	100%	2036118	100%	4093816	100%	1471432	100%	635278	100%	27173779	100%	6599049	100%
	Media	350		39		79		28		12		523		127	
	Clientes	51989	100%	51989	100%	51989	100%	51989	100%	51989	100%	51989	100%	51989	100%

Table 13

La Tabla 13 nos muestra que:

- Existe un orden descendente de algunas variables relevantes por segmento, lo que puede comprobar en primera instancia la validez del proceso de segmentación. Esto lo podemos apreciar por la comparación de las facturaciones promedio.

- El Cluster1 o segmento superior, está conformado por el 10% de los clientes, representan 25% de la rentabilidad, 21% de la facturación total, el 41% de la facturación en Telefonía Móvil y alrededor del 30% de la facturación en Larga Distancia.

- El Cluster2 o segundo segmento, está conformado por el 17% de los clientes. Dentro de sus características más importantes podemos mencionar que representan el 22% de la rentabilidad, el 26% de la facturación total, más del 25% de la facturación en Larga Distancia y Telefonía Móvil.

- El Cluster 3 o tercer segmento, está conformado por el 25% de los clientes, representan el 25% de la rentabilidad, el 42% de la facturación de la Televisión por Cable, 26% de la facturación en Telefonía Móvil y más del 15% de la facturación en Larga Distancia.

- El Cluster4 o cuarto segmento tiene el 48% del total de clientes, representan el 29% de la rentabilidad, el 33% de la facturación Local y total, alrededor de 25% de la facturación de la Televisión por Cable, Larga Distancia y Telefonía Móvil.

- En resumen, podemos apreciar que existe una variable determinante por cada segmento, la facturación en Telefonía Móvil es determinante en el segmento superior. En el Cluster2, el comportamiento de las variables es mucho más homogéneo que el resto de segmentos. En el tercer segmento, la facturación en Televisión por Cable es sobresaliente, lo mismo que en el último Cluster la Telefonía Local es determinante.

Perfiles de los clientes por Cluster dependiendo de las variables cantidad de servicios

CLUSTER	Estad.	Qserv_Bas	%	Qserv_Cab	%	Qserv_Ld	%	Qserv_Mov	%	QSVA_Bas	%	QSVA_Cab	%
1	Suma	14383	19%	4603	15%	14383	19%	5358	38%	6007	22%	1907	12%
	Media	2,7		0,9		2,7		1,0		1,1		0,4	
	Clientes	5260	10%	5260	10%	5260	10%	5260	10%	5260	10%	5260	10%
2	Suma	14247	19%	5342	18%	14247	19%	3903	28%	9297	34%	2734	17%
	Media	1,7		0,6		1,7		0,5		1,1		0,3	
	Clientes	8579	17%	8579	17%	8579	17%	8579	17%	8579	17%	8579	17%
3	Suma	19155	26%	11321	37%	19155	26%	3495	25%	5449	20%	9125	56%
	Media	1,5		0,9		1,5		0,3		0,4		0,7	
	Clientes	13163	25%	13163	25%	13163	25%	13163	25%	13163	25%	13163	25%
4	Suma	26733	36%	9111	30%	26733	36%	1205	9%	6805	25%	2476	15%
	Media	1,1		0,4		1,1		0,0		0,3		0,1	
	Clientes	24987,0	48%	24987,0	48%	24987,0	48%	24987,0	48%	24987,0	48%	24987,0	48%
Total	Suma	74518	100%	30377	100%	74518	100%	13961	100%	27558	100%	16242	100%
	Media	1,4		0,6		1,4		0,3		0,5		0,3	
	Clientes	51989	100%	51989	100%	51989	100%	51989	100%	51989	100%	51989	100%

Tabla 14

De la Tabla 14 se observa lo siguiente:

- El mayor promedio en la cantidad de servicios de Telefonía Fija Local pertenecen al Cluster 1 (2,7 servicios), el cual representa el 19% del total de la cantidad de servicios de Telefonía Fija Local. De igual manera, los servicios de valor agregado de Telefonía Local tienen un promedio de tenencia superior al resto (1,1 por servicio). También podemos afirmar que tienen un teléfono Móvil por clientes y alrededor de 0.9 servicios de Televisión por Cable.

- El Cluster2 tiene el 19% de los servicios de Telefonía Fija local y son aproximadamente el 1,7 servicios por cliente de este segmento, de los cuales un 1,1 aproximadamente tiene servicios de valor agregado.

- El Cluster 3 tiene el 26% de los servicios de telefonía fija local, 1.3 servicios en promedio por cliente, de los cuales 0.4 tiene servicios de valor agregado. De la misma manera, podemos afirmar que este segmento tiene aproximadamente 0.9 servicios de Televisión por Cable y 0.3 celulares por cliente.
- El Cluster 4 tiene el 48% de los servicios de Telefonía Fija Local, sin embargo el promedio de servicios por cliente es de 1,1 servicios por cliente, de los cuales 0,3 tienen algún Servicio de Valor Agregado. En cuanto a la tenencia de los servicios de Televisión por Cable es de aproximadamente 0,3 servicios por cliente y no cuentan con algún servicio de Telefonía Móvil.

El Perfiles de los clientes por Cluster dependiendo de las variables de antigüedad

CLUSTER	Estad.	Ant_Bas	%	Ant_Cab	Ant_Mov	Ant_LD
1	Media	21,0		3,5	3,3	21,0
	Clientes	5.260	10%	5.260	5.260	5.260
2	Media	11,3		24,5	15,0	136,2
	Clientes	8579	17%	8579	8579	8579
3	Media	18,2		3,3	0,8	18,2
	Clientes	13.163	25%	13.163	13.163	13.163
4	Media	10,1		1,1	0,1	10,1
	Clientes	24.987	48%	24.987	24.987	24.987
Total	Media	13,5		2,1	0,8	13,5
	Clientes	51989	100%	51989	51989	51989

Tabla 15

De la Tabla 15 se observa lo siguiente:

- El mayor promedio en antigüedad de los servicios de Telefonía Local Fija se encuentra en el Cluster 1, seguida por el resto de los segmentos en orden descendente. Esto nos permite afirmar que los clientes más antiguos son los que más consumen y los que mantienen una regularidad en su facturación.

- Luego de la antigüedad de los servicios de Telefonía Fija Local que finalmente es la antigüedad de los clientes. De la misma manera, los servicios que le siguen es la Televisión por Cable con 3,5 años en promedio en los tres primeros segmentos y de 1 años en el último segmento. En cuanto a los servicios de Telefonía móvil a excepción del primer segmento que en promedio tiene más de tres años, en el resto es de segmentos el promedio de antigüedad es menor a un año.

Perfiles de los clientes por Cluster dependiendo de las variables baja y tenencia

CLUSTER	Estad.	Bajas_Bas	%	Bajas_Ca	%	Bajas_Mo	%	Bajas_LD	%
1	Suma	2271	11%	509	10%	1998	16%	6697	16%
	Media	0,4		0,1		0,4		1,3	
	Clientes	5262	10%	5262	10%	5262	10%	5262	10%
2	Suma	3764	18%	1097	21%	2521	20%	7336	18%
	Media	0,4		0,1		0,3		0,9	
	Clientes	8579	17%	8579	17%	8579	17%	8579	17%
3	Suma	5386	25%	1006	19%	3256	26%	10573	26%
	Media	0,4		0,1		0,2		0,8	
	Clientes	13163	25%	13163	25%	13163	25%	13163	25%
4	Suma	9843	46%	2621	50%	4523	37%	16298	40%
	Media	0,4		0,1		0,2		0,7	
	Clientes	24987,0	48%	24987,0	48%	24987,0	48%	24987,0	48%
Total	Suma	21264	100%	5233	100%	12298	100%	40904	100%
	Media	0,4		0,1		0,2		0,8	
	Clientes	51991	100%	51991	100%	51991	100%	51991	100%

De la Tabla 16 se observa lo siguiente:

- El promedio de bajas en el servicio de Telefonía Fija (0,4 por cliente), Televisión por Cable (0,1 por cliente) en los cuatro segmentos, es aproximadamente igual, es decir, la variable "baja" no es un elemento diferenciador en los segmentos.
- Las bajas en Telefonía Móvil y Larga Distancia si cumplen el papel de variables determinantes de los segmentos, siendo los segmentos superiores los que tienen mas bajas.

Validación y perfil de los grupos

Luego de haber explicado el perfil por cada segmento, presentamos los gráficos que también nos ayudan a encontrar el perfil y las características que explican de mejor manera cada segmento.

Como podemos apreciar la facturación de Telefonía Local (línea roja) es importante en la determinación de los segmentos, siendo más importante en el segmento superior, de igual manera la facturación en telefonía móvil es importante para este segmento (línea celeste). Por otro lado, vemos que la facturación por Televisión por Cable es importante para el Cluster 3.

Facturación en soles

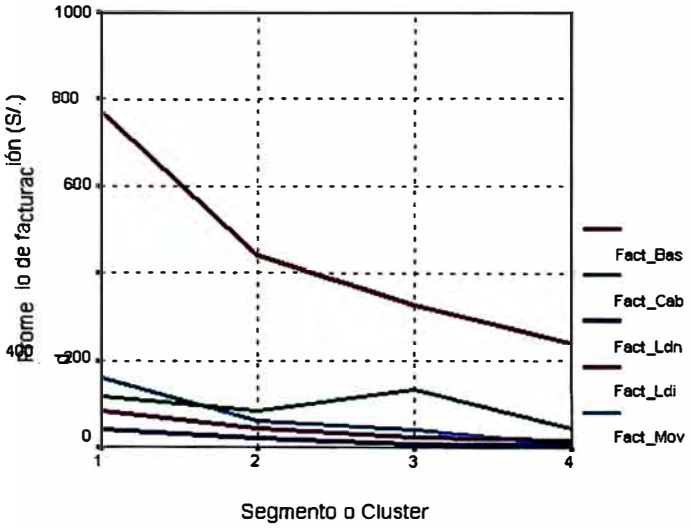


Gráfico 16

Tenencia de servicios

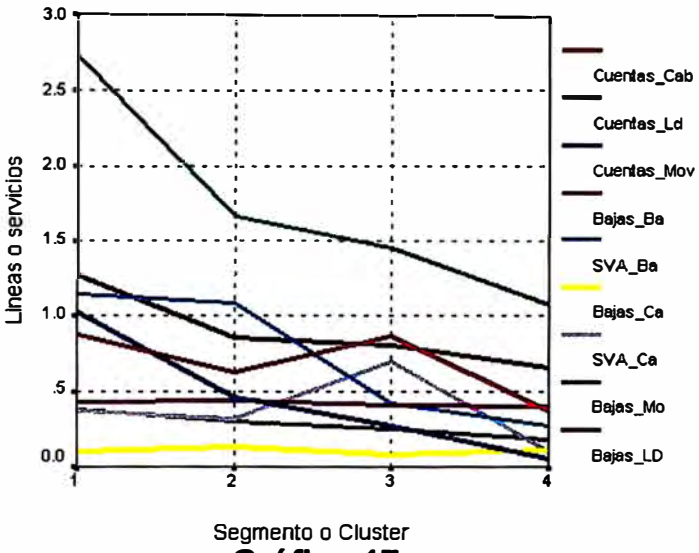


Gráfico 17

CAPITULO IV

IV.5.6. ANÁLISIS DISCRIMINANTE (AD)

Después de clasificar a los clientes, es necesario comprobar si esta clasificación fue adecuada, para lo cual realizamos el Análisis Discriminante que nos permite resolver los problemas de asignación a grupos, es decir, dado un

nuevo elemento del que se conocen las variables métricas, predecir en que segmento deben estar sin la necesidad de volver a realizar un Análisis Cluster, de tal manera que puedan ser clasificados en el Cluster que más se acomode a sus características. La gran diferencia con el Análisis Cluster es que mientras en éste la clasificación es el resultado del análisis, en el Discriminante los grupos son previamente definidos.

IV.5.6.1 Representación gráfica del Análisis Discriminante para dos poblaciones

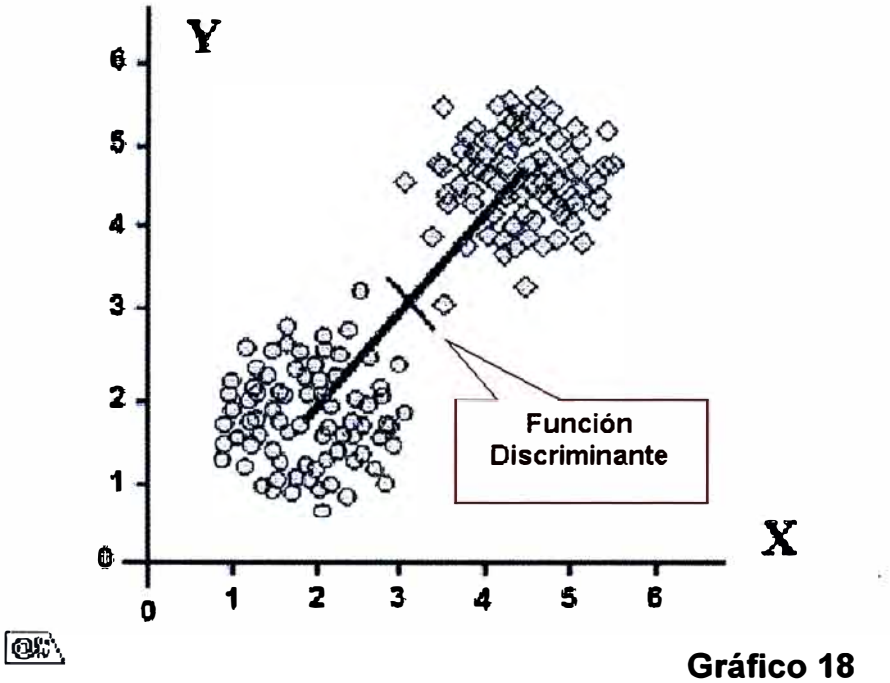


Gráfico 18

FUENTE: //www.estadistico.com, Análisis Discriminante - José E. Gondar Flores

Objetivo

1. Determinar si existen diferencias estadísticamente significativas entre los grupos previamente definidos en el Análisis Cluster.
2. Identificar las variables independientes que cuantifican mejor las diferencias entre Cluster.

3. Hallar la función discriminante para clasificar a los clientes dentro de cada grupo, en base a la identificación de las variables relevantes.
4. Clasificar a los nuevos Clientes Preferentes (Incorporación individual o masiva)
5. Identificar a los clientes candidatos a bajas (Inactivación del servicio)

Si bien es cierto, el Análisis Discriminante es útil para comprender las diferencias entre los grupos o para clasificar correctamente a los clientes en grupos o clases. Sin embargo, la mayor contribución práctica tiene que ver con obtener la función discriminante, la cual nos permite clasificar a cualquier cliente a posteriori, previa validación con la función discriminante. De esta forma, el Análisis Discriminante puede utilizarse para clasificar otras observaciones, en nuestro caso clientes, dentro de grupos definidos.

IV.5.6.3 Selección de las variables dependientes e independientes

Para aplicar el Análisis Discriminante tenemos que identificar primero si las variables son independientes o dependientes, métricas o categóricas. Por otro lado, podríamos utilizar las variables originales pero utilizaremos los factores hallados en el Análisis Factorial que representen al total de las variables, a pesar que esto podría significar una pérdida de información y esto aumentaría el error de clasificación.

IV.5.6.4. Tamaño muestral

En cuanto al tamaño de muestra el Análisis Discriminante es bastante sensible a la relación entre el tamaño muestral y el número variables predictoras. En algunos textos sugieren 20 observaciones por cada variable predictora, menos observaciones por variable podrían llegar a ser inestables. En nuestro caso, trabajamos con el íntegro de la población, teóricamente no deberíamos tener ningún problema de este tipo.

IV.5.6.5 Supuestos del Análisis Discriminante

Las condiciones para la correcta aplicación del Análisis Discriminante.

1. Normalidad Multivariante de las variables independientes. La ausencia de esto supuesto puede causar problemas en la estimación de la función discriminante.
2. Las matrices de covarianza son iguales, en caso contrario pueden afectar considerablemente al proceso de clasificación.
3. Multicolinealidad entre las variables independientes. Al igual que con algunas técnicas multivariantes, un supuesto implícito es que todas las relaciones sean lineales. Las funciones no lineales no están reflejadas en la función discriminante a menos que se realicen trasformaciones específicas de las variables.

IV.5.6.6 Estimación del modelo y ajuste global

Podemos optar por dos métodos los más conocidos de estimación para la obtención de la función discriminante, puede ser en forma simultánea (Incluye todas las variables explicativas). Este método se utiliza cuando se valora la precisión de la clasificación. Por otro lado, tenemos el método paso a paso cuando lo primordial es explicar la pertenencia a los grupos (Incluye las variables independientes que tengan cierto poder de explicación).

Dado que nos interesa obtener una función discriminante con un buen poder de clasificación y las variables que explican esta clasificación, usaremos el método paso a paso.

IV.5.6.6.1 Test que validan la aplicabilidad del análisis

Para que sea efectivo este análisis se debe tener en cuenta las siguientes pruebas:

1. **Test de Box**

Nos permite probar si las variaciones entre las variables de cada grupo son diferentes.

Box's M		8419,177
F	Approx.	133,606
	gl1	63
	gl2	3,3E+09
	Sig.	,000

Tabla 17

Observamos que la significación del valor de F es menor que 0.05. Por tanto, se concluye que las variaciones de las variables en cada grupo son diferentes. Por lo tanto un Análisis Discriminante es adecuado.

IV.5.6.2. Test de Igualdad de Medias

Permite verificar si las medias de las variables involucradas en el análisis son diferentes entre los grupos por medio del **Estadístico Wilks' Lambda.**

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
FAC1_1	.822	3750.761	3	51987	.000
FAC2_1	.760	5476.879	3	51987	.000
FAC3_1	.991	153.969	3	51987	.000
FAC4_1	.854	2971.346	3	51987	.000
FAC5_1	.792	4541.194	3	51987	.000

Tabla 18

Veamos, el cuadro nos muestra los valores del estadístico Wilks' Lambda asociados a sus respectivos valores de F. Además, nos muestra que el nivel de significación es menor que 0.05 (**p<0.05**). Por lo tanto, se concluye que las medias de los factores son diferentes en todos los Cluster. Por otro lado, el factor que tiene un valor de Wilks' Lambda más pequeño (F más grande) es el Fact2_1 que es la que entrará en primer lugar, previamente hemos fijado el valor F mínimo

para entrar o el valor de F máximo para salir. Nosotros tomaremos los valores por defecto del programa SPSS (3,84 y 2,71 respectivamente). Como su F supera el valor mínimo, entonces puede entrar.

El método paso a paso fija un nivel de tolerancia a través de una medida de asociación entre las variables independientes ($1-r^2$, donde r^2 es el coeficiente de determinación). Cuando la tolerancia de la variable i es muy pequeña significa que dicha variable esta muy correlacionada con el resto de las variables explicativas, lo que puede crear algunas dificultades en la estimación. El programa establece un mínimo de tolerancia que es de 0.001.

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Wilks' Lambda
0	FAC1_1	1.000	1.000	3750.761	.822
	FAC2_1	1.000	1.000	5476.879	.760
	FAC3_1	1.000	1.000	153.969	.991
	FAC4_1	1.000	1.000	2971.346	.854
	FAC5_1	1.000	1.000	4541.194	.792
1	FAC1_1	.990	.990	3971.226	.618
	FAC3_1	.998	.998	192.640	.751
	FAC4_1	.952	.952	3993.633	.618
	FAC5_1	.926	.926	6291.332	.557
2	FAC1_1	.985	.914	4060.977	.452
	FAC3_1	.994	.922	262.293	.549
	FAC4_1	.897	.854	5291.329	.427
3	FAC1_1	.960	.833	4622.186	.337
	FAC3_1	.991	.849	314.640	.419
4	FAC3_1	.991	.828	320.548	.331

Tabla 19

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	FAC2_1	1.000	5476.879	
2	FAC2_1	.926	7301.879	.792
	FAC5_1	.926	6291.332	.760
3	FAC2_1	.854	9367.018	.658
	FAC5_1	.873	7728.831	.618
	FAC4_1	.897	5291.329	.557
4	FAC2_1	.833	10035.207	.532
	FAC5_1	.864	7992.832	.493
	FAC4_1	.874	5884.801	.452
	FAC1_1	.960	4622.186	.427
5	FAC2_1	.828	10198.327	.526
	FAC5_1	.859	8138.424	.486
	FAC4_1	.872	5959.548	.445
	FAC1_1	.960	4629.455	.419
	FAC3_1	.991	320.548	.337

Tabla 20

Evaluamos las variables y comprobamos que todas las variables (Factores) superan el F mínimo y el orden de ingreso lo dará el valor del estadístico Wilks más bajo (F más alto), también todos cumplen el requisito de la tolerancia

Ahora analizaremos las variables que deben salir, lo que significa que deben superar al F máximo para ser excluidas (El programa por defecto toma el 2,71).

La Tabla 21 resume las variables que se incorporan a la función discriminante.

Wilks' Lambda

Step	Number of Variables	Lambda	df1	df2	df3	Exact F				Approximate F			
						Static.	df1	df2	Sig.	Statistic	df1	df2	Sig.
1	1	.760	1	3	51987	5477	3	51987	.000				
2	2	.557	2	3	51987	5880	6	103972	.000				
3	3	.427	3	3	51987					5883.173	9	126518	.000
4	4	.337	4	3	51987					5825.432	12	137537	.000
5	5	.331	5	3	51987					4712.465	15	143503	.000

Tabla 21

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	.331	57476.584	15	.000
2 through 3	.816	10554.406	8	.000
3	.984	820.361	3	.000

Tabla 22

Una vez determinado las variables que participan en el modelo y calculado la función discriminante, se determina si es globalmente significativa. Se plantea la hipótesis nula de si las medias poblacionales difieren significativamente en los dos grupos considerados. En el caso contrario, no sería adecuado el análisis, ya que las variables rele

Como se comprueba en la Tabla 22, el nivel de significancia es cero, lo que permite rechazar la hipótesis nula y afirmar que la función discriminante es significativa.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.466[a]	86.9	86.9	.771
2	.206[a]	12.2	99.1	.413
3	.016[a]	.9	100.0	.125

a. First 3 canonical discriminant functions were used in the analysis.

Tabla 23

Como se ve en la Tabla 23 la primera función explica por si misma un 86.9 % de la varianza. Se han empleado las 3 primeras funciones discriminantes canónicas en el análisis.

IV.5.6.7. Cálculo de las puntuaciones Z discriminantes

La puntuación Z discriminante de cualquier función discriminante puede calcularse para cada observación mediante la siguiente fórmula.

$$Z_{jk} \equiv a + W_1 X1k + W_2 X_{2k} + ... + W_n X_{nk}$$

Donde

z_{jk} = Puntuación Z discriminante de la función discriminante j para el objeto k.

a = Constante

w_i = Ponderación discriminante para la variable independiente i

x_{ik} = Variable Independiente i para el objeto k.

Esta puntuación nos permite medidas directas para comprobar observaciones para cada función. Los datos con puntuaciones Z muy similares son más parecidas que aquellas con puntuaciones dispares. En consecuencia, la observación se clasifica dependiendo de la puntuación de clasificación. Utilizamos la función discriminante para luego de validar cada observación clasificada.

Pesos de las variables en las funciones discriminantes

Classification Function Coefficients

	CLUSTER4			
	1	2	3	4
FAC1_1	.912	.099	.804	-.650
FAC2_1	3.128	.670	.007	-.892
FAC3_1	.512	.184	-.076	-.131
FAC4_1	2.372	.258	.218	-.703
FAC5_1	2.676	.843	-.248	-.722
(Constant)	-6.219	-1.686	-1.696	-1.859

Fisher's linear discriminant functions

Tabla 24

En el cuadro se observan las variables asociadas a sus pesos en cada una de las respectivas funciones. Luego procedemos el cálculo de las funciones discriminantes, las cuales se muestran en el siguiente cuadro:

$$D1 = 0.9117 \times Factor1 + 3.1278 \times Factor2 + 0.5119 \times Factor3 + 2.3715 \times Factor4 + 2.6758 \times Factor5$$

$$D2 = 0.0990 \times Factor1 + 0.6699 \times Factor2 + 0.1843 \times Factor3 + 0.2576 \times Factor4 + 0.8432 Factor5$$

$$D3 = +0.8043 \times Factor1 + 0.0067 \times Factor2 - 0.0764 \times Factor3 + 0.2184 \times Factor4 - 0.2483 \times Factor5.$$

$$D4 = -0.6497 \times Factor1 - 0.8922 \times Factor2 - 1308 \times Factor3 - 0.7029 \times Factor4 - 0.7221 \times Factor5.$$

Cluster vs. Su Predicción

Classification Results[a]

			Predicted Group Membership				
		CLUSTER4	1	2	3	4	Total
Original	Count	1	3970	835	455	0	5260
		2	976	4436	1521	1646	8579
		3	570	2104	9420	1069	13163
		4	80	1391	3718	19798	24987
	%	1	75.5	15.9	8.7	.0	100.0
		2	11.4	51.7	17.7	19.2	100.0
		3	4.3	16.0	71.6	8.1	100.0
		4	.3	5.6	14.9	79.2	100.0

a. 72.4% of original grouped cases correctly classified.

Tabla 25

Del cuadro se observa lo siguiente:

- Hay clientes que han sido clasificados en su Cluster real y hay otros tantos que fueron clasificados en otros cluster. De los cuales se obtiene el siguiente resultado:

  7.659%+8.482%+%+18.120%+38.005%=**72.4%**

  Esto nos dice que en la predicción el 72.4% de nuestros clientes son clasificados en su Cluster real.

- Por consiguiente diremos que solo el 27.6% de los clientes fueron erróneamente clasificados.

Una vez validado todo el modelo el siguiente paso es la interpretación de las funciones discriminantes. Utilizando el criterio de los coeficientes estandarizados y la matriz de estructuras, establecemos la importancia relativa de cada variable independiente a la hora de discriminar las variables.

La Tabla 26 muestra los coeficientes estandarizados de los 5 factores que entraron en la función

Standardized Canonical Discriminant Function Coefficients

	Function		
	1	2	3
FAC1_1	.379	.881	.350
FAC2_1	.865	-.119	-.105
FAC3_1	.161	-.122	.155
FAC4_1	.689	.160	-.667
FAC5_1	.758	-.397	.493

Tabla 26

Otro método que más se utiliza últimamente para interpretar estos resultados debido a las deficiencias encontradas en la metodología anterior son las puntuaciones discriminantes, las cuales miden la correlación simple entre cada variable independiente y la función discriminante.

En la Tabla 27 se mide las correlaciones intra grupos combinadas entre las variables discriminantes y las funciones discriminantes cónicas tipificadas. Las variables están ordenadas por el tamaño de la correlación con la función.

Structure Matrix

	Function		
	1	2	3
FAC2_1	.462*	-.130	-.137
FAC1_1	.187	.888*	.419
FAC4_1	.327	.156	-.769*
FAC5_1	.388	-.416	.613*
FAC3_1	.066	-.102	.153*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

Tabla 27

IV.5.6.9 Cálculo de las probabilidades posteriores

Una vez hallada la función discriminante el siguiente paso es establecer la capacidad predictiva del análisis.

$P\left(Gi/_X\right)$: Probabilidad de que el cliente pertenezca al Cluster i dado sus características Donde $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{19} \\ X_{20} \end{bmatrix}$

$$P\left(G_2/_X\right) = \frac{\exp(D_2)}{\sum_{i=1}^{4}\exp(D_i)}$$

$$P\left(G_2/_X\right) = \frac{\exp(D_2)}{\sum_{i=1}^{4}\exp(D_i)}$$

$$P\left(G_3/_X\right) = \frac{\exp(D_3)}{\sum_{i=1}^{4}\exp(D_i)}$$

$$P\left(G_4/_X\right) = \frac{\exp(D_4)}{\sum_{i=1}^{4}\exp(D_i)}$$

El cliente pertenece al Cluster con mayor probabilidad.

CAPITULO V

V.5.7. ANÁLIIS DE CORRESPONDENCIA MÚLTIPLE

Es una técnica de interdependencia cuyo aporte más importante es la elaboración de mapas perceptuales de variables categóricas, es decir, nos

permite de manera gráfica, mostrar los perfiles de las observaciones previamente clasificadas por el Análisis Cluster y validadas por el Análisis Discriminante.

V.5.7.1. Objetivos del Análisis de Correspondencia

El objetivo del Análisis de Correspondencia es entender la relación entre las observaciones, hallar los perfiles por Cluster, es decir, características muy particulares de cada grupo que nos permita conocer al interior del grupo. .

V.5.7.2. Supuestos del Análisis de Correspondencia

En cuanto a los supuestos en el caso del Análisis de Correspondencia, existe una relativa libertad en cuanto a supuestos básicos.

V.5.7.3. Cálculo de la medida de asociación

El análisis de correspondencia utiliza el análisis Chi-cuadrado para estandarizar los valores de frecuencia

Análisis de la matriz de Indicadores

Axis	Inertia	Proportion	Cumulative	Histogram
1	0.4608	0.2304	0.2304	**********
2	0.2767	0.1384	0.3688	**********
3	0.1854	0.0927	0.4615	**********
4	0.1664	0.0832	0.5447	**********
5	0.1531	0.0765	0.6212	*********
6	0.1376	0.0688	0.6900	********
7	0.1363	0.0682	0.7582	********
8	0.1321	0.0661	0.8242	********
9	0.0896	0.0448	0.8690	*****
10	0.0803	0.0401	0.9092	*****
11	0.0757	0.0379	0.9470	****
12	0.0619	0.0310	0.9780	****
13	0.0343	0.0172	0.9952	**
14	0.0097	0.0048	1.0000	
Total	2.0000			

Tabla 28

Reteniendo cinco factores se puede explicar más del 62% de la variabilidad total de los datos. De la misma manera, podemos ver que la inercia

de Factor1 es el más cercano a 1 y por consiguiente el que más explica la varianza total.

					----Component 1----			----Component 2---		
ID	**Name**	**Qual**	**Mass**	**Inert**	**Coord**	**Corr**	**Contr**	**Coord**	**Corr**	**Contr**
1	CLUSTER1	0.332	0.014	0.064	1.352	0.206	0.057	1.058	0.126	0.058
2	CLUSTER2	0.104	0.024	0.060	0.715	0.101	0.026	0.123	0.003	0.001
3	CLUSTER3	0.090	0.036	0.053	0.304	0.031	0.007	-0.417	0.059	0.023
4	CLUSTER4	0.442	0.069	0.037	-0.690	0.441	0.071	-0.045	0.002	0.001
5	FT BAJO	0.812	0.048	0.048	-1.150	0.661	0.137	0.550	0.152	0.052
6	FT MEDIO	0.772	0.048	0.048	0.074	0.003	0.001	-1.241	0.770	0.265
7	FT ALTO	0.816	0.048	0.048	1.075	0.578	0.120	0.690	0.238	0.082
8	RT BAJO	0.783	0.048	0.048	-1.134	0.643	0.133	0.528	0.139	0.048
9	RT MEDIO	0.778	0.048	0.048	0.067	0.002	0.000	-1.245	0.776	0.267
10	RT ALTO	0.827	0.048	0.048	1.067	0.570	0.118	0.718	0.258	0.089
11	FM MEDIO	0.208	0.108	0.018	-0.228	0.160	0.012	-0.125	0.048	0.006
12	FM ALTO	0.208	0.035	0.054	0.702	0.160	0.037	0.386	0.048	0.019
13	FC BAJO	0.290	0.060	0.042	-0.607	0.263	0.048	0.195	0.027	0.008
14	FC MEDIO	0.034	0.039	0.052	0.191	0.014	0.003	-0.230	0.020	0.008
15	FC ALTO	0.190	0.044	0.049	0.650	0.188	0.040	-0.057	0.001	0.001
16	FLDN BAJ	0.307	0.049	0.047	-0.707	0.261	0.053	0.299	0.047	0.016
17	FLDN MED	0.061	0.046	0.048	0.214	0.022	0.005	-0.286	0.039	0.014
18	FLDN ALT	0.135	0.048	0.048	0.519	0.135	0.028	-0.030	0.000	0.000
19	FLDI BAJ	0.344	0.050	0.046	-0.723	0.285	0.057	0.329	0.059	0.020
20	FLDI MED	0.073	0.045	0.049	0.105	0.005	0.001	-0.385	0.068	0.024
21	FLDI ALT	0.222	0.048	0.048	0.667	0.222	0.046	0.014	0.000	0.000

* NOTE * There are no supplementary points to plot

Para interpretar los resultaos podemos utilizar analíticamente las salidas que en este caso nos de el **Minitab** o los Mapas Perceptuales. En el caso de éste último, si bien es cierto no llega a la precisión de los estadísticos nos da una idea clara de las relaciones de las variables y el nivel de contribución que tiene cada una de ellas.

En el cuadro adjunto podemos ver que Mass es el estadístico que nos da la distribución de las variables tanto en filas como en columnas. De esta manera podemos ver que la distribución es relativamente igual para todas las variables excepto para las observaciones que estén alrededor del Cluster 4 y de la facturación media de telefonía móvil.

La Inercia en el Análisis de Correspondencia es utilizada por analogía como la definición utilizada en matemáticas aplicada como es "El momento de

inercia". Bajo estas premisas podemos afirmar que el peso de la inercia de las variables es aproximadamente igual para todas las variables.

En cuanto a las contribuciones para la componente 2 podemos ver que la facturación total media (FT_MEDIA) y la rentabilidad media (R_MEDIA) son las que más contribuyen con un 27% aproximadamente.

Por otro lado tenemos las coordenadas de las variables, las cuales servirán para apreciar gráficamente a través del Mapeo Perceptual.

V.5.7.4. Mapas Perceptuales

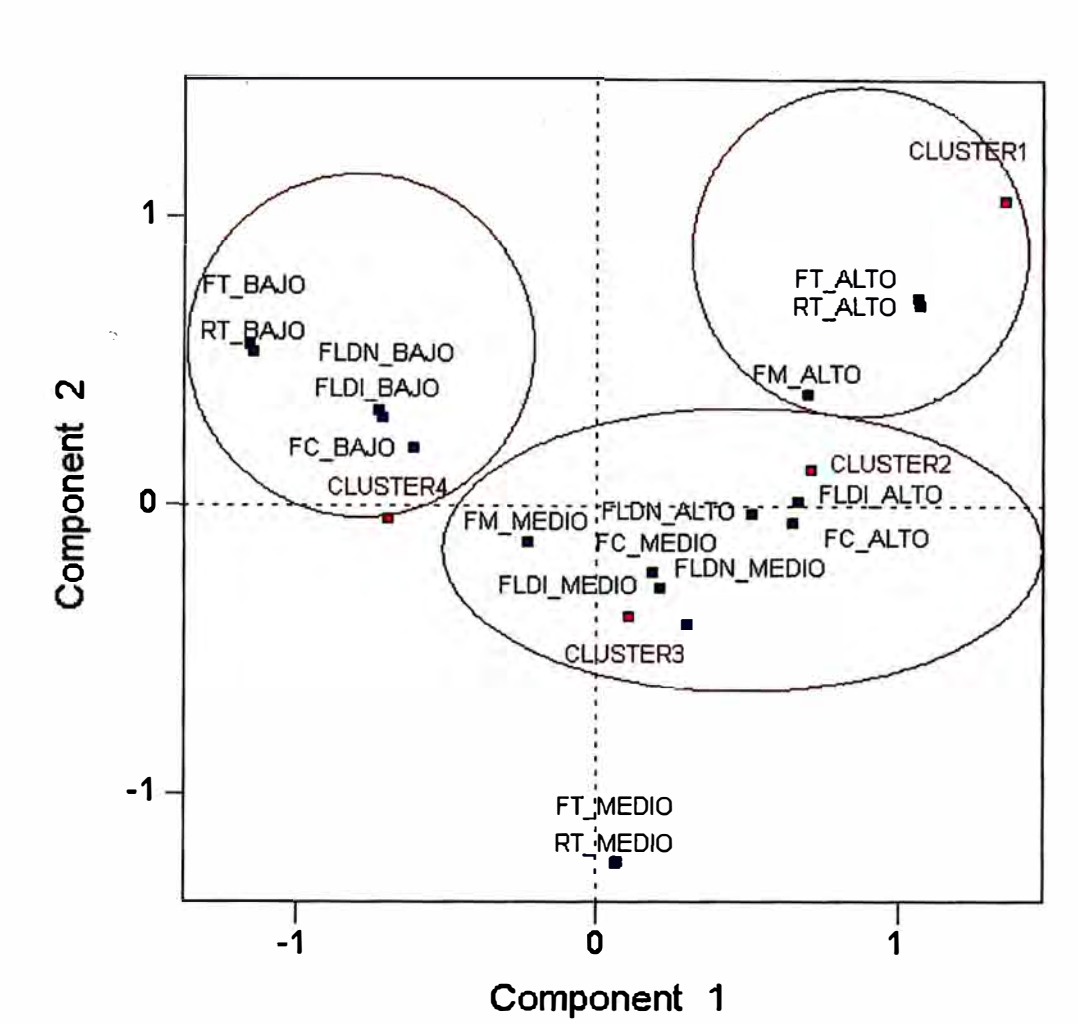
El análisis a través de Mapas Perceptuales nos permite observar de manera gráfica los factores, los atributos por segmento y la relación entre sí.

Un análisis de Mapas Perceptuales consta de dos elementos fundamentales:

- El espacio perceptual, que viene a ser la representación gráfica de los factores más relevantes que explican el perfil del cliente

- El Mapa Perceptual.- No hay posiciones buenas o malas, sino posiciones que nos indiquen un acercamiento a ciertos atributos que nos pueda explicar cualidades al interior del segmento.

Si queremos crear una estrategia a partir de la información que contamos, debemos combinar siempre el uso de los mapas perceptuales con las llamadas tablas de similitud, que mide el grado de asociación. Estos valores de similitud lo confrontamos a través de los valores de similitud (Chi-cuadrado), la cual ofrece una medida estandarizada de asociación que nos permite crear una medida de distancia métrica.



Para evaluar el ajuste conjunto debemos identificar en primer lugar el número apropiado de dimensiones y su importancia. El número máximo de dimensiones que puede ser estimado es uno menos el número más pequeño de filas y columnas.

V.5.7.5. Obtención de resultados

Para identificar la asociación entre categorías, dependerá de su proximidad, es decir, la distancia entre el segmento y sus atributos. Para ayudar a su mejor visualización encerramos con una línea roja el "radio de influencia" por cada segmento.

1. El Cluster1 o segmento superior, esta rodeado por variables asociados a la variable Facturación Total (FT_ALTA), mayor rentabilidad (RT_ALTA) y

Facturación Alta en Telefonía Móvil (FM_ALTA). Estos atributos nos indican que el segmento superior está conformado por aquellos que más facturan, los que generan más rentabilidad y con un poco menos de contribución que en los casos anteriores la mayor facturación en Telefonía Móvil. En este último atributo por temas de distancia, es decir, por "radio de influencia"

2. El Cluster 2 o segundo segmento, está conformado por atributos de alta facturación en Larga distancia y Telefonía Móvil. Asimismo, podemos apreciar que existen una serie de variables con facturación media que describen el perfil de este segmento.
3. El Cluster 3 o tercer segmento, está conformado por atributos de alta facturación en televisión por cable (FC-.ALTA), alta facturación en Telefonía Móvil (FM_ALTA) y Larga Distancia (FLDI_ALTA). Sin embargo, por distancias, estas variables están más alejadas para este segmento que en el Cluster 2.
4. El perfil del Cluster4 o cuarto segmento está explicada por menor facturación total (FT_MENOR), menor rentabilidad (RT_MENOR), menor facturación en Telefonía móvil (FM_MENOR) y Larga Distancia (FLDI_MENOR)

CAPÍTULO VI

VI.6. CONCLUSIONES Y RECOMENDACIONES

1. Existen diferentes formas de segmentar a los clientes, uno de ellos son los métodos multivariados que aplicados secuencialmente adquieren mucha más potencia que aplicándolos de manera independiente. Se recomienda no solo aplicarlos de manera secuencial sino también en otros rubros de la actividad comercial.

2. Se recomienda complementar este análisis con información cualitativa del mercado, lo que significaría trabajar no con la totalidad de la población sino con una muestra. La información cualitativa puede ser obtenida a través de encuestas por cliente.

3. Las variables que más resaltan como elementos diferenciadores por cada segmento, son:
    - La facturación en Telefonía Móvil, determinante en el segmento superior (Cluster 1)
    - En el Cluster 2, el comportamiento de las variables de consumo es mucho más homogéneo que el resto de los segmentos.
    - En el tercer segmento, la facturación de Televisión por Cable es sobresaliente.
    - En el último Cluster la Telefonía Local es determinante en la clasificación.

4. Los mejores clientes residenciales de Telefónica del Perú se distribuyen en cuatro segmentos de acuerdo a la rentabilidad:

    - El Cluster 1 →10% de los clientes →25% de la rentabilidad.
    - El Cluster 2→17% de los clientes→22% de la rentabilidad.
    - El Cluster 3-->-25% de los clientes-→25% de la rentabilidad.
    - El Cluster 4→48% de los clientes-→29% de la rentabilidad.

Como podemos apreciar a través de la segmentación, identificamos a los mejores y peores Clientes Preferentes. A los mejores clientes, se me ocurre hacerles "caricias[1]", promociones, campañas, etc.; de esta manera retenerlos. Con respecto a los clientes del segmento inferior serían los candidatos potenciales a ser dados de baja del programa de Cliente Preferente.

5. A pesar que en primera instancia puede significar una pérdida de precisión en la clasificación de las observaciones el incluir los factores y no las variables originales. Se tiene que evaluar si se puede arriesgar cambiar, menor precisión por menor dependencia de muchas variables. El error cometido en esta predicción es de 27.6%, es decir, 76.4% de nuestros clientes son clasificados en el Cluster adecuado.

[1] *Actividades que motiven fidelidad del cliente*

BIBLIOGRAFÍA

1.Drew, P: L`impotance du role de la telèmatique dans l"amènagement urbain et regional. Revue d`Ecoomie et Urbaine, nro3, 1989.

2.Iglesias María: Extensión del servicio telefónico básico: Reglamentación elaborada, Universidad de Valladolid , Departamento de Economía Aplicada.

3.FUNDESCO, Comunicaciones y desarrollo, Predicción y economía de las telecomunicaciones.

4.José A. Del Busto: La Tesis Universitaria.

5.Elena Abascal: Análisis Multivariante Aplicado el Marketing.

6.Hair Anderson- Tathan. Black: Análisis Multivariado

7.C.M. Cuadras: Métodos de análisis Multivariante

8.K.V. Mardia, J.T. Kent, J.M. Bibby: Multivariante análisis

9.Naresh K. Malhotra, Investigación de Mercados un enfoque práctico

10.Richard A. Johnson, Dean W. Wichern: Applied Multivariate Statistical

ANEXO 1

Sintaxis de las corridas realizadas en el SPSS

****

EXPLORATORIO Y LINEALIDAD

```
EXAMINE
  VARIABLES=cuentas_ fact_bas cuenta1 fact_cab cuenta2 fact_ldn fact_ldi
  cuenta3 fact_mov antigued antigu1 antigu2 antigu3 bajas_ba sva_ba bajas_ca
  sva_ca bajas_mo bajas_ld
  /PLOT  HISTOGRAM NPPLOT
  /MESTIMATORS  HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5)
TUKEY(4.685)
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

EXAMINE
  VARIABLES= fact_mov antigued antigu1 antigu2 antigu3 bajas_ba sva_ba
bajas_ca
  sva_ca bajas_mo bajas_ld
  /PLOT  HISTOGRAM NPPLOT
  /MESTIMATORS  HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5)
TUKEY(4.685)
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

EXAMINE
  VARIABLES=cuentas_ fact_bas cuenta1 fact  cab
  /PLOT  HISTOGRAM NPPLOT
  /MESTIMATORS  HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5)
TUKEY(4.685)
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

```
EXAMINE
  VARIABLES=cuenta2 fact_ldn fact_ldi cuenta3
  /PLOT  HISTOGRAM NPPLOT
  /MESTIMATORS  HUBER(1.339) ANDREW(1.34) HAMPEL(1.7,3.4,8.5)
TUKEY(4.685)
  /STATISTICS DESCRIPTIVES EXTREME
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

******************************************************************************
```

FACTORIAL

```
FACTOR/VARIABLES ncuentas nfact_ba ncuenta1 nfact_ca nfactldn nfactldi
ncuenta3 nfact_mo nantigue nantigu1 nantigu2 nbajas_b nsva_ba nbajas_c
nsva_ca nbajas_m nbajas_l  /MISSING LISTWISE /ANALYSIS ncuentas nfact_ba
ncuenta1 nfact_ca nfactldn nfactldi ncuenta3 nfact_mo nantigue nantigu1
nantigu2 nbajas_b nsva_ba nbajas_c nsva_ca nbajas_m nbajas_l
  /PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO EXTRACTION
ROTATION
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(100)
  /ROTATION VARIMAX
  /SAVE REG(ALL)
  /METHOD=CORRELATION


******************************************************************************
```

CLUSTER

```
QUICK CLUSTER
  fac1_1 fac2_1 fac3_1 fac4  1 fac5  1
  /MISSING=LISTWISE
  /CRITERIA= CLUSTER(4) MXITER(200) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER
  /PRINT INITIAL.
```

****

DISCRIMINANTE

```
DISCRIMINANT
 /GROUPS=cluster4(1 4)
 /VARIABLES=fac1_1 fac2_1 fac3_1 fac4_1 fac5_1
 /ANALYSIS ALL
 /SAVE=CLASS SCORES PROBS
 /PRIORS  EQUAL
 /STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR COV GCOV
TCOV TABLE
 /CLASSIFY=NONMISSING POOLED
```

****

PERFILES POR CLUSTER

```
SUMMARIZE
 /TABLES=fact_bas fact_cab fact_ldn fact_ldi fact_mov renta  BY cluster4
 /FORMAT=NOLIST TOTAL
 /TITLE='Case Summaries'
 /MISSING=VARIABLE
 /CELLS=SUM MEAN COUNT .

SUMMARIZE
 /TABLES=cuentas_ cuenta1 cuenta2 cuenta3 sva  ba sva  ca  BY cluster4
 /FORMAT=NOLIST TOTAL
 /TITLE='Case Summaries'
 /MISSING=VARIABLE
 /CELLS=SUM MEAN COUNT .

SUMMARIZE
 /TABLES=antigued antigu1 antigu2 antigu3  BY cluster4
 /FORMAT=NOLIST TOTAL
 /TITLE='Case Summaries'
 /MISSING=VARIABLE
 /CELLS=SUM MEAN COUNT .
```

ANEXO 2

Correlations

		Cuentas_Basica	Fact_Bas	Cuentas_Cab	Fact_Cab	Cuentas_Ld	Fact_Ldn	Fact_Ldi	Cuentas_Mov	Fact_Mov	Antiguedad_Ba	Antiguedad_Ca	Antiguedad_Mo	Antiguedad_LD	Bajas_Ba	SVA_Ba	Bajas_Ca	SVA_Ca	Bajas_Mo	Bajas_LD
Cuentas_Basica	Pearson Correlation	1.000	.627**	.236**	.220**	1.000**	.124**	.191**	.105**	.084**	.313**	.192**	.093**	.313**	-.021**	.175**	.013**	.120**	.062**	.381**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Fact_Bas	Pearson Correlation	.627**	1.000	.169**	.167**	.627**	.249**	.247**	.067**	.081**	.241**	.132**	.057**	.241**	-.005	.217**	-.013**	.086**	.034**	.141**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.245	.000	.003	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Cuentas_Cab	Pearson Correlation	.236**	.169**	1.000	.801**	.236**	.003	.076**	.156**	.120**	.163**	.657**	.129**	.163**	-.001	.115**	-.193**	.393**	.079**	.054**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.447	.000	.000	.000	.000	.000	.000	.000	.847	.000	.000	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Fact_Cab	Pearson Correlation	.220**	.167**	.801**	1.000	.220**	-.018**	.097**	.151**	.129**	.175**	.589**	.129**	.175**	.018**	.075**	-.103**	.748**	.084**	.065**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Cuentas_Ld	Pearson Correlation	1.000**	.627**	.236**	.220**	1.000	.124**	.191**	.105**	.084**	.313**	.192**	.093**	.313**	-.021**	.175**	.013**	.120**	.062**	.381**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Fact_Ldn	Pearson Correlation	.124**	.249**	.003	-.018**	.124**	1.000	.103**	-.005	.000	.021**	-.009*	-.012**	.021**	-.008	.044**	-.010*	-.023**	-.004	-.059**
	Sig. (2-tailed)	.000	.000	.447	.000	.000	.	.000	.235	.931	.000	.035	.008	.000	.073	.000	.020	.000	.381	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Fact_Ldi	Pearson Correlation	.191**	.247**	.076**	.097**	.191**	.103**	1.000	.017**	.044**	.085**	.049**	.018**	.085**	-.016**	.075**	-.005	.063**	.001	-.062**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.269	.000	.761	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Cuentas_Mov	Pearson Correlation	.105**	.067**	.156**	.151**	.105**	-.005	.017**	1.000	.767**	-.002	.149**	.697**	-.002	.032**	.071**	.009*	.097**	.129**	.062**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.235	.000	.	.000	.611	.000	.000	.611	.000	.000	.037	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Fact_Mov	Pearson Correlation	.084**	.081**	.120**	.129**	.084**	.000	.044**	.767**	1.000	-.022**	.108**	.574**	-.022**	.032**	.061**	.016**	.086**	.161**	.049**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.931	.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Antiguedad_Ba	Pearson Correlation	.313**	.241**	.163**	.175**	.313**	.021**	.085**	-.002	-.022**	1.000	.240**	.038**	1.000**	-.102**	.028**	-.031**	.101**	-.005	.021**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.611	.000	.	.000	.000	.000	.000	.000	.000	.000	.218	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Antiguedad_Ca	Pearson Correlation	.192**	.132**	.657**	.589**	.192**	-.009*	.049**	.149**	.108**	.240**	1.000	.185**	.240**	.051**	.081**	-.125**	.313**	.116**	.085**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.035	.000	.000	.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Antiguedad_Mo	Pearson Correlation	.093**	.057**	.129**	.129**	.093**	-.012**	.018**	.697**	.574**	.038**	.185**	1.000	.038**	.036**	.046**	.026**	.080**	.085**	.064**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.008	.000	.000	.000	.000	.000	.	.000	.000	.000	.000	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Antiguedad_LD	Pearson Correlation	.313**	.241**	.163**	.175**	.313**	.021**	.085**	-.002	-.022**	1.000**	.240**	.038**	1.000	-.102**	.028**	-.031**	.101**	-.005	.021**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.611	.000	.000	.000	.000	.	.000	.000	.000	.000	.218	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Bajas_Ba	Pearson Correlation	-.021**	-.005	-.001	.018**	-.021**	-.008	-.016**	.032**	.032**	-.102**	.051**	.036**	-.102**	1.000	-.021**	.152**	.026**	.122**	.726**
	Sig. (2-tailed)	.000	.245	.847	.000	.000	.073	.000	.000	.000	.000	.000	.000	.000	.	.000	.000	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
SVA_Ba	Pearson Correlation	.175**	.217**	.115**	.075**	.175**	.044**	.075**	.071**	.061**	.028**	.081**	.046**	.028**	-.021**	1.000	-.008	.002	.034**	.013**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.	.058	.691	.000	.004
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Bajas_Ca	Pearson Correlation	.013**	-.013**	-.193**	-.103**	.013**	-.010*	-.005	.009*	.016**	-.031**	-.125**	.026**	-.031**	.152**	-.008	1.000	-.076**	.076**	.122**
	Sig. (2-tailed)	.004	.003	.000	.000	.004	.020	.269	.037	.000	.000	.000	.000	.000	.000	.058	.	.000	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
SVA_Ca	Pearson Correlation	.120**	.086**	.393**	.748**	.120**	-.023**	.063**	.097**	.086**	.101**	.313**	.080**	.101**	.026**	.002	-.076**	1.000	.054**	.052**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.691	.000	.	.000	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Bajas_Mo	Pearson Correlation	.062**	.034**	.079**	.084**	.062**	-.004	.001	.129**	.161**	-.005	.116**	.085**	-.005	.122**	.034**	.076**	.054**	1.000	.116**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.381	.761	.000	.000	.218	.000	.000	.218	.000	.000	.000	.000	.	.000
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991
Bajas_LD	Pearson Correlation	.381**	.141**	.054**	.065**	.381**	-.059**	-.062**	.062**	.049**	.021**	.085**	.064**	.021**	.726**	.013**	.122**	.052**	.116**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.000	.000	.000	.
	N	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991	51991

**. Correlation is significant at the 0.01 level (2-tailed).