

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

DEPARTAMENTO DE ESTADÍSTICA E INFORMÁTICA



TRABAJO MONOGRÁFICO

**“DETERMINACIÓN DE PERFILES DE TURISTAS NACIONALES DE LOS NIVELES
SOCIOECONÓMICOS MEDIO Y ALTO MEDIANTE EL ANÁLISIS
CONGLOMERADO BIETÁPICO”**

**PRESENTADO PARA OPTAR AL TÍTULO DE INGENIERO ESTADÍSTICO E
INFORMÁTICO**

BEATRIZ ETA PORRAS HUAMÁN

MODALIDAD DE EXAMEN PROFESIONAL

LIMA - PERÚ

2014

ÍNDICE

I. INTRODUCCIÓN.....	1
II. EL PROBLEMA DE INVESTIGACIÓN.....	2
2.1 Fundamentación del Problema de Investigación.....	2
2.2 Formulación de las Preguntas de la Investigación.....	3
2.3 Objetivo de la Investigación.....	3
2.4 Justificación de la Investigación.....	4
III. MARCO TEÓRICO.....	4
3.1 Definición de Turismo.....	4
3.1.1 Tipo de turismo.....	6
3.1.2 Características generales del Turismo Interno.....	6
3.1.3 Modalidad de la actividad turística.....	7
3.1.4 Entorno económico.....	9
3.2 Conglomerado Bietápico o Conglomerados en Dos Fases.....	11
3.2.1 Pasos del algoritmo.....	11
3.2.2 Supuestos.....	12
3.2.3 Medidas de distancia.....	12
3.2.4 Selección del número de clusters.....	15
3.2.5 Importancia de las variables.....	16
3.2.6 Tratamiento de valores atípicos.....	17
3.2.7 Medida de Cohesión y Silueta.....	17
IV. Metodología de la Investigación.....	18
4.1 Tipo de Investigación.....	18
4.2 Formulación de la Hipótesis.....	18
4.3 Identificación y definición operacional de las variables.....	18
4.4 Diseño de la Investigación.....	20
4.5 Población y muestra.....	20

V. PROCEDIMIENTO Y ANÁLISIS DE DATOS.....	21
5.1 Análisis de Conglomerados Bietápico.....	21
5.2 Validación mediante la técnica CHAID	21
VI. RESULTADOS	22
6.1 Análisis inicial	22
6.2 Análisis definitivo	27
6.3 Validación de los conglomerados asignados	33
VII. CONCLUSIONES.....	38
VIII.RECOMENDACIONES	38
IX. BIBLIOGRAFÍA.....	39
X. ANEXOS.....	40

RESUMEN

El turismo en el Perú ha venido creciendo rápidamente en los últimos años (MINCETUR - Plan Estratégico Nacional de Turismo 2012 - 2021 – PENTUR, pág. 11). En el proceso de la investigación turística se recopila los datos, para ser analizados y efectuar una crítica, con fines de alcanzar elementos de trabajo hacia otras fases de desarrollo turístico, las mismas que permitirán obtener nuevos conocimientos que como aporte se utilizarán en la actividad turismo.

En este sentido el presente trabajo busca caracterizar o establecer cuáles son los perfiles de los turistas internos en el Perú de los niveles socioeconómicos medio y alto con la finalidad de diversificar la estrategia turística y de este modo poder administrar en forma sostenida recursos del sector público como el privado. Esta investigación utiliza una muestra de 2,400 turistas de los principales destinos turísticos del Perú.

El objetivo planteado para el desarrollo del presente trabajo es la determinación de los grupos de individuos observando sus características socio-demográficas, los hábitos de vida y las preferencias con respecto a viajes al interior del país. La metodología empleada corresponde a la técnica de agrupamiento Análisis de Conglomerados Bietápico. La pregunta de investigación planteada para el desarrollo del proyecto se orienta a determinar los diversos tipos de perfiles de turistas nacionales en los niveles socioeconómicos A, B y C.

Finalmente se determinó la homogeneidad de los grupos en función a la variabilidad intra-grupos, encontrándose dos grupos, descritos estos en función a los estadísticos encontrados dentro de cada grupo.

I. INTRODUCCIÓN

El sector turístico interno presenta una tendencia creciente en un periodo de 21 años (BADATUR – OTP), por lo cual MINCETUR en un análisis de la situación actual y futura con respecto al turismo en el Perú indica que el crecimiento turístico ha venido creciendo de manera sostenida y se perfila que en los siguientes años seguirá con un crecimiento sostenible.

Por otro lado, la segmentación de mercado es un proceso que consiste en dividir el mercado total de un bien o servicio en varios grupos más pequeños e internamente homogéneos. El segmento de mercado es un grupo homogéneo de consumidores que se pueden identificar dentro de un mercado, que tienen actitudes, hábitos y comportamientos, ubicación geográfica, características afines y que reaccionarían de modo parecido frente a algún estímulo comercial o de comunicación.

Con la técnica de segmentación denominada Análisis de Conglomerados Bietápico, se procederá a caracterizar los grupos hallados de acuerdo a las características y posiciones descritas por los entrevistados. Para alcanzar este objetivo se empleó información proporcionada por el Ministerio de Comercio Exterior y Turismo (MINCETUR), sobre un estudio de turismo emisor nacional llevado a cabo en las principales ciudades de Lima, Arequipa, Chiclayo, Huancayo y Trujillo.

II. EL PROBLEMA DE INVESTIGACIÓN

2.1 Fundamentación del Problema de Investigación

El Perú retrocedió cuatro lugares en el Índice de Competitividad en Viajes y Turismo del Foro Económico Mundial (WEF, por sus siglas en inglés) del 2012. Con este resultado, nuestro país se ubica en el puesto 73 de un total de 140 a nivel mundial y en la posición 13 de los 27 que forman el continente americano.

El informe detalla que el Perú es el tercer país en el mundo con la fauna más rica y con un buen número de lugares naturales y culturales considerados patrimonio de la humanidad. Además, reconoce la efectividad del marketing realizado por el Gobierno para promover el sector.

Sin embargo, también resalta que el Perú aún tiene tres problemas que detienen el crecimiento de la industria turística y explican la ligera caída. Estos son la inseguridad, la poca infraestructura turística y los altos impuestos para los vuelos comerciales.

Entre el 2002 y el 2012, el turismo interno, medido por la cantidad de viajes que se realizan en el país, crecería más de 4% en el año 2013 y se realizarán más de 36 millones de viajes al interior del Perú, estimó el Ministerio de Comercio Exterior y Turismo (PERÚ 21, abril del 2013).

Algunos de los estudios de PROMPERÚ, *Perfil del Vacacionista Nacional 2007* (que viaja por recreación), definen también con mayor exactitud cómo es el turista nacional. La mayoría de los viajeros suele buscar información durante el viaje; el viaje lo hacen principalmente por vía terrestre; predominan los viajes en grupo familiar (padres e hijos) y en grupo de amigos (sobre todo jóvenes). Los viajes son motivados por la búsqueda de descanso, relax y salir de la rutina, asimismo el clima y el paisaje natural influyen en la elección del destino.

Por último, para realizar este diagnóstico de demanda es imprescindible considerar el flujo de turistas a los distintos destinos turísticos del país. Ya que la estadística para medir la llegada de turistas hacia los recursos turísticos está incompleta, se ha analizado la llegada hacia las regiones a través de los reportes de pernoctación de los hoteles.

Se conoce los perfiles diferenciados en los turistas nacionales de acuerdo a las investigaciones preliminares. En el presente trabajo se buscarán encontrar estos segmentos en los niveles socioeconómicos medio y alto mediante la técnica estadística, segmentación Bietápica. Asimismo se busca brindar información a investigaciones futuras.

2.2 Formulación de las Preguntas de la Investigación

El problema de investigación se formuló a través de las siguientes preguntas:

- ¿Cómo son los perfiles de los potenciales turistas nacionales de los niveles socioeconómicos A, B y C?
- ¿La técnica Conglomerado Bietápico ayuda a identificar los perfiles de los potenciales turistas internos?

2.3 Objetivo de la Investigación

Objetivo general:

Identificar grupos de individuos en función a sus características sociodemográficas, hábitos de vida y preferencias con respecto a viajar al interior del país, en los niveles socioeconómicos A, B y C.

Objetivo específicos:

- Describir los grupos hallados de acuerdo a las variables utilizadas en la segmentación.
- Estimar la tasa correcta de clasificación para la segmentación Conglomerado Bietápico, mediante el algoritmo CHAID.

2.4 Justificación de la Investigación

Actualmente el turismo interno en el Perú es una de las actividades que contribuye principalmente en el incremento del PBI y la generación de empleo descentralizado, sin embargo, la inseguridad, la poca infraestructura turística y los altos impuestos para los vuelos comerciales, son los principales problemas que afectan al turista en el Perú, Foro Económico Mundial (WEF, por sus siglas en inglés).

Esta investigación se realiza con la finalidad de identificar perfiles de los turistas nacionales mediante el uso de variables demográficas, hábitos de vida y la intención con respecto a viajar al interior del país.

Los resultados de esta investigación permitirán dar a conocer a las entidades del Estado y demás instituciones interesadas en la actividad, los distintos grupos de turista nacional que actualmente existe en el Perú. También brindará información sobre la intención de realizar viajes al interior de país, el acceso a internet, el lugar ideal para realizar un viaje en familia etc.

Permitirá elaborar estrategias de marketing para cada grupo objetivo hallado, de esta manera se buscará difundir guías turísticas que permitan al turista nacional sentirse más seguro, además de alternativas de alojamiento confortables y medios de transporte, etc. Con ello habrá un mejor entendimiento del turista interno de los niveles socioeconómicos medio y alto, lo cual se reflejará en la satisfacción del viajero y el incremento de arribos en el interior de nuestro país.

III.MARCO TEÓRICO

3.1 Definición de Turismo

La definición de turismo realizado por la Organización Mundial de Turismo (OMT) es la siguiente:

“El turismo comprende las actividades que realizan las personas durante sus viajes y estancias en lugares distintos al de su residencia habitual por menos de un año y con fines de ocio, negocios, estudio, entre otros.”

(http://www.mincetur.gob.pe/TURISMO/Producto_turistico/Fit/fit/Guias/Amazonas.pdf, conceptos fundamentales de turismo).

Correspondiente a esta definición, se puede indicar que el turismo encuentra una relación estrecha entre el viaje por placer y una forma placentera de buscar recreación en el tiempo libre del cual dispone un individuo.

Las características vinculadas a este sector se puede catalogar como una actividad multisectorial muy compleja, debido a que involucra a un grupo de personas que no necesariamente se encuentran relacionadas directamente, ya sea en el ámbito social, económico, político, cultural y educativo.

Nuestro país posee grandes fortalezas para que el turismo nacional se desempeñe eficazmente. Se podría catalogar al Perú como un país que posee una gran biodiversidad, siendo uno de los más biodiversos del mundo por la variedad de climas, paisajes, flora y fauna que en el cual coexisten, esto se debe a que contamos con la Amazonía y muchos otros destinos históricos atractivos para los turistas internacionales y nacionales (turismo receptivo y turismo interno). Este perfil de turismo brinda grandes oportunidades. En la actualidad, el sector turismo brinda beneficios económicos a las poblaciones que viven cerca o aledaños a zonas turísticas, ya que los pobladores aprovechan este movimiento de personas para realizar diversos tipos de actividades no solo de intercambio comercial, sino también a nivel empresarial.

El gran movimiento de turistas y las distintas empresas que circulan alrededor de las zonas turísticas pueden producir que se descuide el medio ambiente y generar contaminación, ya sea esta por medio del suelo, aire o agua que lastimará indudablemente el ecosistema de las zonas periféricas.

3.1.1 TIPO DE TURISMO

- **Turismo Receptivo**

El turismo receptivo es también conocido como receptor, siendo el tipo de turismo que realizan los extranjeros o no residentes, los mismos que se movilizan de un país determinado a otro denominado destino. Este tipo de turismo trasciende directamente en la economía del país de destino, mediante la inversión, el empleo y la generación de divisas para el mismo.

- **Turismo Interno**

El turismo interno es también conocido como doméstico, es el turismo realizado en el mismo país que se encuentra el residente al interior del mismo. Este tipo de turismo estaría clasificado dentro del campo de operación del turismo educativo. En el presente trabajo se centrará en este tipo de turismo.

- **Turismo Egresivo**

El turismo egresivo es también conocido como emisor, es aquel tipo de turismo el cual realizan los nacionales o residentes de un país hacia los países extranjeros.

3.1.2 Características generales del Turismo Interno

En relación con el turismo en el Perú, MINCETUR (2008) afirma que:

Hoy, en el Perú, se presenta una "nueva promesa de desarrollo", que esta vez se fundamenta en su rico y abundante legado patrimonial, cultural y natural que a través de los años han comenzado a tomarse en cuenta. Nuestro país es uno de los países más diversos en actividades y lugares turísticos del mundo, hecho que está constatado por la UNESCO a través

de los grandes bienes declarados *Patrimonio de la Humanidad* por esta organización internacional (pág. 2).

En todo el territorio peruano podemos encontrar ciudades y regiones con diversidad en climas, monumentos, costumbres, paisajes culturales de codiciado valor histórico-cultural y áreas naturales protegidas (reservas).

Según MINCETUR (2008), por estas razones y dado el interés en adecuar la oferta turística, infraestructura y comercio de estos destinos emergentes en verdaderas opciones turísticas de calidad, Perú se debe presentar como un destino diverso y renovado (pág. 30).

El turismo ha sido identificado como uno de los sectores económicos de mayor crecimiento, superando incluso las tasas mundiales, según PROMPERU (2000):

“En el Perú el promedio anual de llegadas internacionales fue de 12.9% entre 1990 y 1999; mientras que a nivel mundial, fue un promedio de 4% durante el mismo lapso. Asimismo, esta actividad se ha convertido a partir de 1998 en la segunda fuente de divisas, después de la minería, llegando a generar 1003 millones de dólares durante el año 1999” (pág. 3).

3.1.3 Modalidad de la actividad turística

Este tipo de clasificación corresponde al tipo de actividad que es desarrollado por el visitante, dividiéndose esta clasificación en dos grandes grupos:

- **Turismo convencional**

Esta modalidad se desarrolla mediante la práctica organizada y tradicional de turismo. Generalmente mediante este tipo de prácticas el turista recurre a las agencias de viajes y turismo, servicios de alojamiento, transporte, alimentación, excursiones y cualquier otro programa son convencionales.

- **Turismo no convencional**

Esta modalidad se denomina así debido a la práctica no tradicional de turismo y este posee características especializadas. Como ejemplo a este tipo de prácticas podemos mencionar las caminatas, canotaje, observación de la naturaleza, etc., entre este tipo de actividades se tiene las siguientes clasificaciones:

a) Turismo de aventura: este tipo de turismo se subdivide en las siguientes clases:

- Turismo de aventura suave o soft: caminata o trekking, canotaje en aguas de corrientes suaves, kayak en aguas de corrientes suaves, montañismo, cabalgatas.
- Turismo de aventura fuerte o hard: puenting, ala delta, espeleología, canotaje, kayak, montañismo en lugares de alto riesgo.
- Turismo de naturaleza: desarrollado en zonas naturales con variados tipos de intereses.

b) Ecoturismo: referido al turismo de naturaleza, promueve las diferentes características encontradas en la naturaleza, la conservación de su entorno y la gestión sostenible de los recursos de la misma.

c) Turismo rural: desarrollado en el medio rural de manera sostenible, generalmente es una forma de escape de la rutina hacia la tranquilidad encontrada en el campo, sus habitantes y la naturaleza.

d) Turismo místico o religioso: son turistas con motivaciones estrictamente religiosas.

e) Turismo esotérico: aquellos que buscan la energía cósmica y su influencia en el hombre y la tierra.

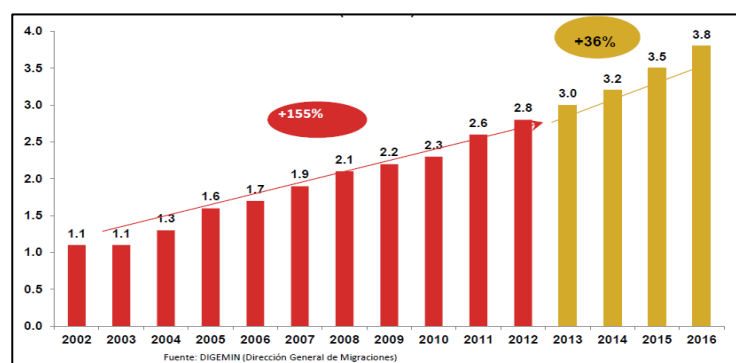
3.1.4 Entorno económico

La evaluación del desarrollo del entorno económico se centra en el análisis de variables claves para el sector turismo, las cuales son revisadas periódicamente por entidades encargadas de la evaluación del sistema económico en dicho rubro, siendo algunas de las variables analizadas las tasas de interés y políticas monetarias, inflación, tipo de cambio, entre otros aspectos adicionales. Analizando el entorno financiero del turismo en el Perú, ya sea receptivo o interno, encontramos la siguiente información observado desde una perspectiva macroeconómica:

“El sector turismo tiene un impacto económico significativo en el mundo, su contribución al PBI mundial asciende a 9%, generando US\$ 6.6 trillones y 260 millones de empleos (1 de cada 11 empleos son generados por la actividad turística). De igual manera, el turismo en el Perú ha venido creciendo de manera sostenida. Entre el 2002 y el 2012, la llegada de turistas internacionales creció de 1.1 a 2.8 millones, la generación de divisas se cuadruplicó pasando de US\$ 837 a US\$3,288 millones. Asimismo, en el 2012 el turismo generó 1’081,035 empleos y su contribución directa al PBI nacional fue de 3.72%.”. (MINCETUR - Plan estratégico nacional de turismo 2012 - 2021 – PENTUR, pág. 11).

Gráfico N° 01

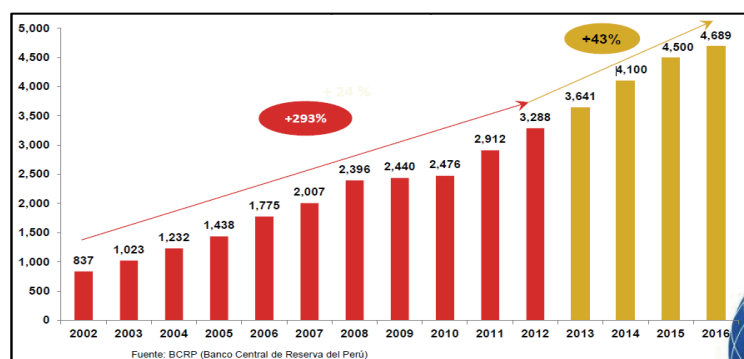
Arribo de turistas internacionales en el Perú
(US\$ millones)



Fuente: DIGEMIN (Dirección General de Migraciones)

Gráfico N° 02

Ingreso de divisas internacionales en el Perú por parte del turismo
(US\$ millones)

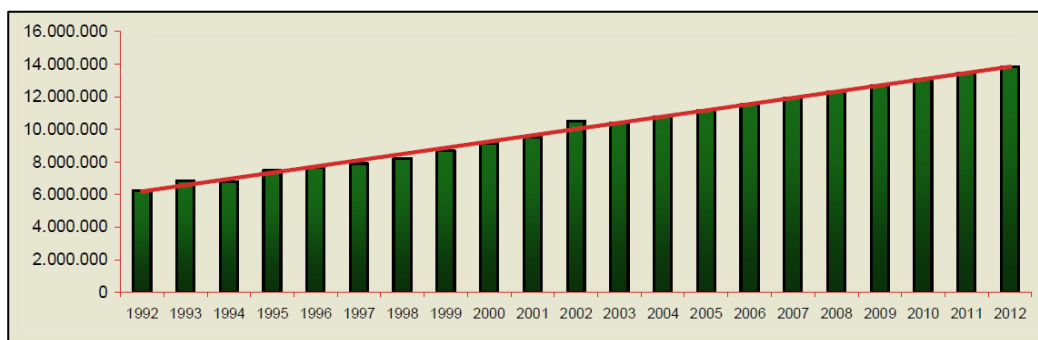


Fuente: BCRP (Banco Central de Reserva del Perú)

De acuerdo a las proyecciones de MINCETUR y otras instituciones interesadas en el desarrollo y evaluación del sector turismo en el Perú, se puede indicar para los próximos años el pronóstico de crecimiento económico a nivel turístico.

Gráfico N° 03

Turismo interno histórico en el Perú años 1992 - 2012



Fuente: BADATUR – OTP. Elaboración: Observatorio turístico del Perú

Según el gráfico N° 03, la tendencia a lo largo del periodo de estos últimos 21 años, anualmente existe un crecimiento turístico interno sostenible, las cuales permiten afirmar que el turismo en el país se está convirtiendo en una de las actividades principales generadoras no solo de empleos y crecimiento comercial, sino también de fomentar mayor interés en el turista nacional a través de las experiencias agradables en los viajes. Se puede intuir un incremento y diversificación de la oferta turística nacional, promoviendo una

cultura turística entre los peruanos, con el interés de instituciones públicas y privadas, entidades del estado entre otras.

3.2 Conglomerado Bietápico o Conglomerados en Dos Fases

El método de Conglomerado Bietápico fué descrito por Chiu, Fang Chen Wang y Jeris el año 2001 en su artículo A Robust and Scalable Clustering Algorithm for MixedType Attributes in Large Database Environment (pág.263-268).

En este artículo, los autores proponen una extensión del modelo basado en medidas de distancia, asumiendo una medida de distancia que permite agrupar datos que posean tanto atributos de tipo cualitativo como cuantitativo. La medida sugerida deriva de un modelo de probabilidad y es equivalente a la disminución del logaritmo de la verosimilitud resultante de la fusión de ambos clusters. Además el algoritmo, determina el número de conglomerados automáticamente.

3.2.1 Pasos del algoritmo

El algoritmo en dos fases puede analizar archivos de datos de gran tamaño mediante la construcción de un árbol de características de conglomerados (CF) que resuma los registros.

El procedimiento consta de los siguientes pasos:

- Paso 1: Pre-clusters de los casos, en esta parte se desarrolla una aproximación secuencial para pre-agrupar los casos. El objetivo consiste en calcular una nueva matriz de datos con una menor cantidad de casos para el siguiente paso. Para conseguir este objetivo, los pre-clusters calculados y sus características son empleados como nuevos casos. Los pre-clusters se definen como regiones densas ubicadas en el espacio de atributos analizado¹.
- Paso 2: Cluster de los casos, se aplica un modelo basado en una técnica jerárquica, en forma similar a las técnicas aglomerativas jerárquicas. El modelo asume las variables continuas X_j ($j=1, \dots, p$) tienen una distribución

¹ Johann Bacher, Knut Wenzing, Melanie Volger - SPSS TwoStep CLUSTER - A First Evaluation

normal con media μ_{ij} y variancia σ_{ij}^2 , mientras las variables categóricas a_j son dentro de cada cluster “ i ” independientes y con distribución multinomial con probabilidades π_{ijl} correspondiente a la l -ésima categoría ($l=1,2,\dots,m$) de la variable a_j ($j=1,2,\dots,q$). Las comprobaciones empíricas internas que se pueden realizar indicarían que este procedimiento es bastante robusto frente a las violaciones tanto del supuesto de independencia como de las distribuciones.

3.2.2 Supuestos

El algoritmo de la segmentación Bietápica se basa en la medida de distancia que da los mejores resultados si todas las variables son independientes, variables continuas que tienen una distribución normal y variables categóricas con distribución multinomial. Sin embargo, el algoritmo está pensado para trabajar razonablemente bien cuando las suposiciones no se cumplan. Esto debido a que las técnicas de segmentación no involucran pruebas de hipótesis y cálculo de niveles de significancia, solo está presente para realizar análisis descriptivos. Es aceptable segmentar datos cuando no se cumplan estos supuestos. Solamente el investigador determinará si los resultados son satisfactorios para sus necesidades (Norusis 2011).

3.2.3 Medidas de distancia

Las medidas de distancia determinan la forma de calcular la similaridad entre dos conglomerados.

Medida de distancia: Función Log Verosimilitud

La función de log-verosimilitud se puede emplear tanto en variables continuas como categóricas y se trata de una distancia basada en probabilidad. Esta distancia entre dos conglomerados se encuentra relacionada a la disminución en log-verosimilitud, ya que se combinan en un solo cluster. En el cálculo de la distancia se trabaja con la distribución normal para las variables cuantitativas y

distribución multinomial para las variables categóricas, así como la independencia entre las variables y casos.

De este modo, la distancia entre los cluster “ i ” y “ s ” se define del modo siguiente:

$$d(i, s) = \xi_i + \xi_s - \xi_{(i,s)}$$

Donde:

$$\xi_i = -n_i \left(\sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{ij}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{ijl} \log(\hat{\pi}_{ijl}) \right)$$

$$\xi_s = -n_s \left(\sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{sj}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{sjl} \log(\hat{\pi}_{sjl}) \right)$$

$$\xi_{(i,s)} = -n_{(i,s)} \left(\sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{(i,s)j}^2 + \hat{\sigma}_j^2) - \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{(i,s)jl} \log(\hat{\pi}_{(i,s)jl}) \right)$$

Además:

$$\hat{\pi}_{vjl} = \frac{n_{vjl}}{n_v}$$

Donde:

- $d(i, s)$: distancia entre el cluster “ i ” y el cluster “ s ”.
- $\langle i, s \rangle$: índice que representa el cluster formado por la combinación de los cluster “ i ” y “ s ”.
- n_v : número de registros en el cluster v
- n_{vjl} : número de registros en el cluster v cuya j -ésima variable categórica toma la l -ésima categoría

- $\hat{\sigma}_j^2$: varianza estimada de la j -ésima variable continua a través de todo el conjunto de datos
- $\hat{\sigma}_{vj}^2$: es la varianza estimada de la j -ésima variable continua en el cluster v .

En general, ξ_v se puede interpretar como la medida de dispersión intra-cluster v (para $v = i,s,<i,s>$) que consta de dos partes. La primera parte $-n_v \sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{vj}^2 + \hat{\sigma}_j^2)$ mide la dispersión de las variables continuas x_j dentro del cluster v . Si $\hat{\sigma}_j^2$ se ignora y solo se usa $\hat{\sigma}_{vj}^2$, $d(i,s)$ sería exactamente la disminución en log-verosimilitud cuando se combinan ambos cluster. Este término se añade para resolver el problema causado cuando $\hat{\sigma}_{vj}^2 = 0$, lo cual llevaría a que el logaritmo natural no esté definido (por ejemplo si el cluster contemplara un solo caso). El segundo término, también conocido como entropía $-n_v \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{vj} \log(\hat{\pi}_{vj})$ se usa para medir la dispersión de variables categóricas.

De modo similar al método aglomerativo jerárquico los cluster con las distancias $d(i,s)$ más pequeñas son fusionados en cada paso. De este modo la función log-verosimilitud para el paso con “ k ” cluster se calcula del modo siguiente:

$$l_k = \sum_{v=1}^k \xi_v.$$

Esta función se puede interpretar como la dispersión dentro de los cluster. Si únicamente se utilizan variables categóricas, l_k es equivalente a la entropía dentro de los clusters.

Medida de distancia: Distancia Euclideana

Esta medida de distancia, que se utiliza en otros procedimientos de agrupación solo puede ser aplicada si todas las variables son continuas. Corresponde a la distancia Euclideana entre los centros de dos clusters, definido por el vector de media de cada variable.

3.2.4 Selección del número de clusters

La metodología puede utilizar el método jerárquico en el segundo paso para evaluar múltiples soluciones de cluster y automáticamente determinar el número de clusters. Una característica del algoritmo jerárquico es que produce una secuencia de particiones en una sola corrida: 1, 2, 3, clusters mientras el algoritmo k-means debería correr varias veces por cada número de clusters especificados.

Para determinar el número de clusters en forma automática, el Conglomerado Bietápico emplea un estimador de dos etapas a través el método jerárquico. (Bacher et. al .2004):

El Criterio de Información de Akaike (AIC):

$$AIC_k = -2l_k + 2r_k$$

Donde r_k es el número de parámetros independientes

El Criterio de Información Bayesiana (BIC)

$$BIC_k = -2l_k + r_k \log n$$

El cual es calculado durante la primera fase. De este modo BIC_k o AIC_k resultan ser buenos estimadores iniciales del máximo número de clusters.

Finalmente el máximo número de clusters es un conjunto igual al número de clusters donde el ratio:

$$\frac{BIC_k}{BIC_l} < c_l$$

- Donde $c_l < 0.04$

En la segunda fase se usa el ratio de cambio R_k en la distancia para k clusters, definido como:

$$R_k = d_{k-1}/d_k$$

Donde d_{k-1} es a distancia su k clusters son fusionados a $k-1$ cluster. La distancia d_k se define de la misma manera. El número de clusters se obtiene para la solución donde ocurre un gran salto en el ratio de cambio.

El ratio se calcula de la siguiente manera:

$$R_{k1}/R_{k2}$$

Para los valores más grandes de $R_{(k)}$. Si el ratio de cambio es más grande que el valor del umbral 1.15, el número de clusters debería ser igual a k_1 de otro modo el numero de clusters es igual a la solución con máximo (k_1, k_2) .

3.2.5 Importancia de las variables

La contribución de cada una de las variables a la creación de cluster puede calcularse para variables continuas y categóricas

Variables continuas

En el caso que la variable sea continua, la medida de la importancia se base en:

$$t = \frac{\hat{u}_{kv} - \hat{u}_{jv}}{\hat{\sigma}_{jv}/\sqrt{n_v}} \sim t(n_v - 1)$$

Tanto la significancia como el estadístico “ t ” pueden tomarse como medida de la importancia.

Variables discretas

En el caso que la variable sea categórica, la media de la importancia se encuentra basada en:

$$\chi^2 = \sum_{l=1}^{L_v} \left(\frac{n_{jvl} - n_{vl}}{n_{vl}} \right)^2 \sim \chi^2_{(L_v)}$$

Tanto la significancia como el estadístico χ^2 pueden tomarse como medida de la importancia.

3.2.6 Tratamiento de valores atípicos

Este grupo permite tratar de manera especial a los valores atípicos durante la conglomeración si se llena el árbol de características de los conglomerados (CF). El árbol CF se considerará lleno si no acepta ningún caso más en un nodo hoja y no se encuentra ningún nodo hoja que se pueda dividir.

3.2.7 Medida de Cohesión y Silueta

De acuerdo con las investigaciones realizadas por Kaufman y Rousseeuw (1990) acerca de la interpretación de estructuras de conglomerados, se puede determinar si los resultados son pobres, correctos o buenos.

- Un resultado “bueno” indicará en los datos una evidencia prudente o sólida de la existencia de una estructura de conglomerados.
- Un resultado “correcto” indica una evidencia débil, y
- Un resultado “pobre” indica, según la valoración, que no existen evidencias obvias.

El coeficiente de calidad o medida de silueta en todos los registros se calcula mediante la fórmula:

$$\text{Coeficiente de silueta} = (B - A) / \text{máx.}(A, B)$$

Donde:

- A: distancia del registro al centro de su conglomerado.
- B: distancia del registro al centro del conglomerado más cercano al que no pertenece.
- Un coeficiente de silueta de 1 implicará que todos los casos están ubicados directamente en los centros de sus conglomerados.
- Un valor de 0 implica, de media, que los casos están equidistantes entre el centro de su propio conglomerado y el siguiente conglomerado más cercano.
- Un valor de -1 significaría que todos los casos se encuentran en los centros de conglomerado de otro conglomerado.

IV. Metodología de la Investigación

4.1 Tipo de Investigación

El tipo de investigación es no experimental. Se considera de carácter exploratorio y descriptivo, ya que es un problema de investigación poco estudiado y no ha sido abordado bajo el mismo enfoque. Por lo que permitirá a futuro realizar una investigación de mayor complejidad e identificar variables que ayuden a diferenciar e identificar el comportamiento del turista nacional de acuerdo al contexto estudiado. Por otro lado es descriptiva ya busca explicar las características más importantes de los segmentos hallados de acuerdo a la diferencias de cada grupo.

4.2 Formulación de la Hipótesis

- Existen grupos diferenciados entre los consumidores de turismo interno en el Perú en los niveles socioeconómicos medio y alto.
- La Segmentación Bietápica permite identificar perfiles de turistas nacionales.

4.3 Identificación y definición operacional de las variables

Las variables independientes que entraran al modelo, son aquellas que se muestran la siguiente Tabla N°01.

Tabla N° 01

	Variable	Descripción	Valores/Unidades
P6	Actividades realizada con mayor frecuencia en el tiempo libre	Actividades frecuentes realizadas en mayor frecuencia por el turista interno	<ul style="list-style-type: none"> • Practicar deporte, acudir al gimnasio • Quedarme a dormir en casa / hacer una siesta • Reunirme con amigos o familiares en casa • Salir de paseo fuera de la ciudad y regresar el mismo día • Pasear por centros comerciales dentro de la ciudad • Salir a comer o almorzar a restaurantes • Ir al cine • Ir al teatro • Jugar juegos de video (Tv, computadora) • Estar en internet • Salir de viaje / hacer turismo • Hacer tareas del hogar (limpiar, lavar, cocinar, etc.) • Ir al sauna / ir al spa • Ir a discotecas / pubs • Ir a casinos • Ver Tv o DVD en casa
P13	Lugar ideal para hacer un viaje familiar	Lugar ideal para hacer un viaje familiar, incluyendo a los hijos	<ul style="list-style-type: none"> • Cusco, Macchu Picchu, Cajamarca, Lima, Trujillo, Máncora, Arequipa, Iquitos, Huaraz, Chiclayo.
P15	Intención de viaje	Intención de viaje del turista interno	<ul style="list-style-type: none"> • Si, No
P46	Posesión de tarjeta de crédito	Posesión de tarjeta de crédito del turista interno	<ul style="list-style-type: none"> • Si, No
P50	Uso o acceso a internet	Uso o acceso a internet	<ul style="list-style-type: none"> • Si , No
P70	Oportunidad de compra	Oportunidad de compra de productos tecnológicos para uso personal o del hogar en el último año	<ul style="list-style-type: none"> • Si, No (Laptop, Netbook, USB 8GB, MP3, Ipod, Iphone, Blackberry, TVLED, Play Station II ó III)
P72	Calificación a la actual situación económica del país	Calificación del turista interno a la situación económica del país	<ul style="list-style-type: none"> • Buena, Regular, Mala
P77	Género	Género del turista interno	<ul style="list-style-type: none"> • Masculino, Femenino
P78	Edad	Edad del turista interno	<ul style="list-style-type: none"> • Años
P79	Nivel socioeconómico	Nivel socioeconómico del turista interno	<ul style="list-style-type: none"> • A,B,C,D,E
P85	Grado de instrucción	Nivel educativo alcanzado del turista interno	<ul style="list-style-type: none"> • Sin estudios/ Primaria completa • Primaria completa/Secundaria incompleta • Secundaria completa/Superior Técnica incompleta • Superior Técnica completa/Superior Universitaria incompleta • Superior Universitaria completa • Postgrado/Maestría • Doctorado/PHD

Las variables que intervienen en el modelo son: P6, P13, P15, P46, P50, P70, P72, P77, P78, P79 y P85 de las cuales P6, P13, P15, P46, P50, P70, y P77 son de naturaleza categórica y la variable P78 es de naturaleza cuantitativa.

4.4 Diseño de la Investigación

El diseño de la investigación es transversal descriptivo.

4.5 Población y muestra

La muestra seleccionada ascendió a 2400 casos (hombres y mujeres de 18 a 64 años de edad) seleccionados en las principales ciudades de afluencia turística Lima, Trujillo, Arequipa, Chiclayo y Huancayo, de los niveles socioeconómicos A, B y C.

La metodología de la siguiente fue la siguiente:

- Muestreo en dos etapas
Hogar: (definida como 1ª unidad de muestreo). Dicho nivel está conformado por los hogares pertenecientes a los niveles socioeconómicos previstos (A, B y C).
Persona: (definida como 2ª unidad de muestreo). Dicho nivel está conformado por el miembro del hogar que participará en la entrevista (turista interno aplica filtro). Esta persona (informante) será seleccionada utilizando el método Kish.
- Método de Manzaneo para el desplazamiento de los encuestadores.
Puntos de partida elegidos al azar, desplazamiento por avenidas, jirones y calles mediante el sistema de "manzaneo". Selección de viviendas de manera sistemática.
- Método de Kish para la selección de las personas.
Selección aleatoria del informante, según el número de habitantes del hogar y el número de encuesta a realizar.
- Definición de niveles socioeconómicos utilizado la fórmula de APEIM.

Tablas N°2

Ciudad	Muestra	NSE A	NSE B	NSE C	Error global
Lima	800	150	300	350	+/- 3.5%
Arequipa	400	50	150	200	+/- 5.0%
Chiclayo	400	50	150	200	+/- 5.0%
Trujillo	400	50	150	200	+/- 5.0%
Huancayo	400	50	150	200	+/- 5.0%
Total	2400	350	900	1150	+/- 2.0%

Fuente: PROMPERU

V. PROCEDIMIENTO Y ANÁLISIS DE DATOS

5.1 Análisis de Conglomerados Bietápico

Se procedió a realizar el análisis con la ayuda del software estadístico SPSS versión 22, el procedimiento del Análisis de Conglomerados en dos fases es una herramienta de exploración diseñada para describir agrupaciones naturales, por lo que en el presente trabajo de investigación se utiliza esta técnica para el análisis de todas las variables.

El procesamiento consta de dos etapas (Chiu et. al. 2001, Bacher et. al. 2004)

- Formación de los pre-cluster: mediante árboles CF.
- Agrupación de los pre-cluster mediante cluster jerárquico aglomerativo.

5.2 Validación mediante la técnica CHAID

Se utilizó el algoritmo CHAID para la validación de los grupos hallados al utilizar el Análisis de Conglomerados Bietápico.

CHAID divide a la población en dos o más grupos distintos basados en categorías del “mejor” predictor de una variable predictora. Luego divide cada uno de estos en grupos más pequeños basados en variables de otros predictores. Este proceso de división continua y termina hasta que no se encuentren mas predictores estadísticamente significativos (o hasta que se cumpla alguna regla de parada).

VI. RESULTADOS

Para el establecimiento de los grupos, recurrimos a la técnica de Conglomerado Bietápico, obteniendo los siguientes resultados:

6.1 Análisis inicial

A continuación se presentan los estadísticos (proporciones) para cada una de las variables dentro de cada conglomerado

Tabla N° 03

Distribución de género por conglomerado

CONGLOMERADO	MASCULINO		FEMENINO	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	749	74.9%	799	57.1%
2	251	25.1%	601	42.9%
Combinados	1000	100.0%	1400	100.0%

Como podemos observar el 74.9% de los individuos del sexo masculino pertenecen al conglomerado 1. Por otro lado el 57.1% de los individuos del sexo femenino pertenece al conglomerado 1 mientras que el 42.9% pertenece al conglomerado 2.

Tabla N° 04

Distribución del nivel socioeconómico por conglomerado

CONGLOMERADO	NSE A		NSE B		NSE C	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	322	92.0%	641	71.2%	585	50.9%
2	28	8.0%	259	28.8%	565	49.1%
Combinados	350	100.0%	900	100.0%	1150	100.0%

Como podemos observar el 92.0% de los individuos del NSE A pertenecen al conglomerado 1 mientras que en el 71.2% de los individuos del NSE B pertenece al conglomerado 1 y en el NSE C el 50.9% pertenece al conglomerado 1 y el 49.1% pertenece al conglomerado 2.

Tabla N° 05

Distribución del grado de instrucción alcanzado por conglomerado

CONGLOMERADO	PRIMARIA INCOMPLETA		SECUNDARIA		COMPLETA/ SUPERIOR		COMPLETA/SUP.	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	0	0.0%	9	10.2%	265	41.4%	631	71.8%
2	9	100.0%	79	89.8%	375	58.6%	248	28.2%
Combinados	9	100.0%	88	100.0%	640	100.0%	879	100.0%

CONGLOMERADO	UNIVERSITARIA		/MAESTRÍA		DOCTORADO / PHD	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	534	79.8%	101	94.4%	8	100.0%
2	135	20.2%	6	5.6%	0	0.0%
Combinados	669	100.0%	107	100.0%	8	100.0%

Como podemos observar el 100.0% de los individuos sin estudios o con primaria incompleta pertenecen al conglomerado 1 mientras que el 89.8% de los individuos con primaria completa o secundaria incompleta pertenecen al conglomerado 2. De los individuos con educación superior técnica o superior universitaria incompleta, el 71.8% pertenecen al conglomerado 1, mientras que el 79.8% de los individuos con educación superior universitaria completa pertenecen al conglomerado 1 y por último de los individuos con post grado o maestría, el 94.4% pertenece al conglomerado 1 y 100% de los que tienen doctorado pertenecen al conglomerado 1.

Tabla N° 06

Distribución de la intención de viaje por motivos de vacaciones o recreación al interior del país, para el periodo 2010-2011 por conglomerado

CONGLOMERADO	VIAJARÉ		VIAJE		VIAJAR	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	372	75.0%	596	69.3%	448	58.8%
2	124	25.0%	264	30.7%	314	41.2%
Combinados	496	100.0%	860	100.0%	762	100.0%

CONGLOMERADO	VIAJARÉ		NS/NC	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	100	44.8%	32	54.2%
2	123	55.2%	27	45.8%
Combinados	223	100.0%	59	100.0%

Respecto a la intención de viaje por motivos de vacaciones o recreación al interior del país vemos que dentro de los que afirman que con toda seguridad viajarán el 75% pertenece al conglomerado 1, de igual modo el mayor porcentaje (69.3%) entre quienes probablemente viajen se ubican en el conglomerado 1, mientras que los que mencionan no estar seguros de viajar (58%) pertenecen al conglomerado 1, y en el conglomerado 2 prevalecen en mayor porcentaje (55%) los que definitivamente no viajarían.

Tabla N° 07

Distribución de la calificación atribuida a la actual situación económica del país por conglomerado

CONGLOMERADO	BUENA		REGULAR		MALA		NS/NC	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	149	82.8%	1072	67.5%	322	51.9%	5	50.0%
2	31	17.2%	517	32.5%	299	48.1%	5	50.0%
Combinados	180	100.0%	1589	100.0%	621	100.0%	10	100.0%

La percepción acerca de la situación económica del país viene siendo de regular(67%) a buena (82%) para el conglomerado 1, mientras que el conglomerado 2 tiene una percepción mas negativa de la situación económica del país, aproximadamente la mitad de ellos atribuyen una calificación negativa.

Tabla N° 08

Distribución al lugar ideal para hacer un viaje familiar, incluyendo a los hijos por conglomerado

CONGLOMERADO	CUSCO		MACCHU PICCHU		CAJAMARCA		LIMA		TRUJILLO	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	304	66.4%	120	58.8%	76	47.8%	64	65.3%	58	61.1%
2	154	33.6%	84	41.2%	83	52.2%	34	34.7%	37	38.9%
Combinados	458	100.0%	204	100.0%	159	100.0%	98	100.0%	95	100.0%

CONGLOMERADO	MANCORA		AREQUIPA		IQUITOS		HUARAZ		CHICLAYO	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	72	75.8%	56	71.8%	59	75.6%	36	60.0%	35	63.6%
2	23	24.2%	22	28.2%	19	24.4%	24	40.0%	20	36.4%
Combinados	95	100.0%	78	100.0%	78	100.0%	60	100.0%	55	100.0%

Respecto al lugar ideal para hacer un viaje familiar, incluyendo a los hijos, se puede observar que, entre todas las opciones que poseen los individuos, entre los 10 lugares más visitados en primer lugar se encuentra Cusco, luego por orden de mérito correspondería a Macchu Picchu, Cajamarca, Lima, Trujillo, Máncora,

Arequipa, Iquitos, Huaraz y por último Chiclayo. Las descripciones de cada uno de los lugares anteriormente mencionados son:

- Cusco (66.4%), Macchu Picchu (58.8%), Lima (65.3%), Trujillo (61.1%), Máncora (75.8%), Arequipa (71.8%), Iquitos (75.6%), (Huaraz) (60%), Chiclayo (63.6%) pertenecen al conglomerado 1.
- Cajamarca: El 52.2% pertenece al conglomerado 2.

Tabla N° 09

Distribución de las actividades que realiza con mayor frecuencia en el tiempo libre por conglomerado

CONGLOMERADO	PRACTICAR DEPORTE, ACUDIR AL GIMNASIO		CASA / HACER UNA SIESTA		REUNIRME CON AMIGOS O FAMILIARES EN CASA	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	716	79.9%	361	52.3%	321	58.2%
2	180	20.1%	329	47.7%	231	41.8%
Combinados	896	100.0%	690	100.0%	552	100.0%

Respecto a las actividades que realiza con mayor frecuencia en el tiempo libre se perfila en el conglomerado 1, siendo los de mayor frecuencia: para practicar deporte, acudir al gimnasio (79.9%), quedarse a dormir en casa/ hacer una siesta (52.3%), reunirse con amigos o familiares en casa (58.2%).

Tabla N° 10

Distribución de la posesión de tarjeta de crédito por conglomerado

CONGLOMERADO	SI		NO		NC	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	685	83.3%	854	54.6%	9	60.0%
2	137	16.7%	709	45.4%	6	40.0%
Combinados	822	100.0%	1563	100.0%	15	100.0%

Respecto a la posesión de tarjeta de crédito se perfila en el conglomerado 1, con respuesta afirmativa (83.3%) y respuesta negativa (54.6%).

Tabla N° 11

Distribución del uso o acceso a internet por conglomerado

CONGLOMERADO	SÍ		NO	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	1487	99.5%	61	6.7%
2	7	.5%	845	93.3%
Combinados	1494	100.0%	906	100.0%

La percepción acerca del uso o acceso a internet se observa que, el 99.5% respondió afirmativamente en el conglomerado 1, mientras que el 93.3% respondió negativamente en el conglomerado 2.

Tabla N° 12

Distribución de la oportunidad de compra de productos para uso personal o de su hogar por conglomerado

CONGLOMERADO	SÍ		NO	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado 1	1006	81.7%	542	46.4%
2	226	18.3%	626	53.6%
Combinados	1232	100.0%	1168	100.0%

En la oportunidad de compra de productos para uso personal o de su hogar se observa que, el 81.7% respondió afirmativamente en el conglomerado 1, mientras que el 53.6% respondió negativamente en el conglomerado 2.

6.2 Análisis definitivo

El objetivo del presente trabajo es determinar grupos de individuos en función a sus características socio-demográficas, hábitos de vida y preferencias con respecto a viajar al interior del país

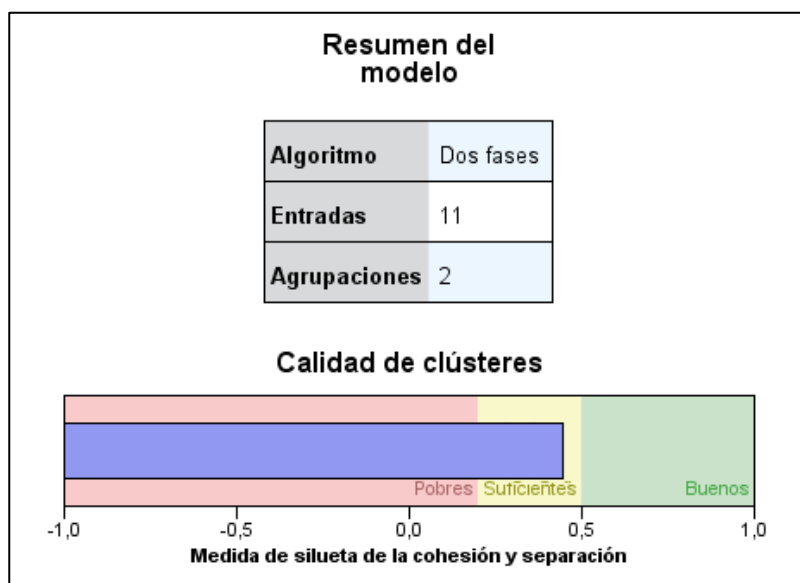
Modelo segmentación Bietápica

Para determinar que el número de conglomerados en el modelo se aplicó la metodología de segmentación de dos etapas a las 10 variables estudiadas, mediante esta técnica se busco encontrar automáticamente (mediante los criterios *BIC* o *AIC*) el número de conglomerados adecuados.

Medida de distancia

Al tratarse de variables categóricas y cuantitativas, la distancia que se utilizó en esta metodología es la log-verosimilitud.

Gráfico N° 04

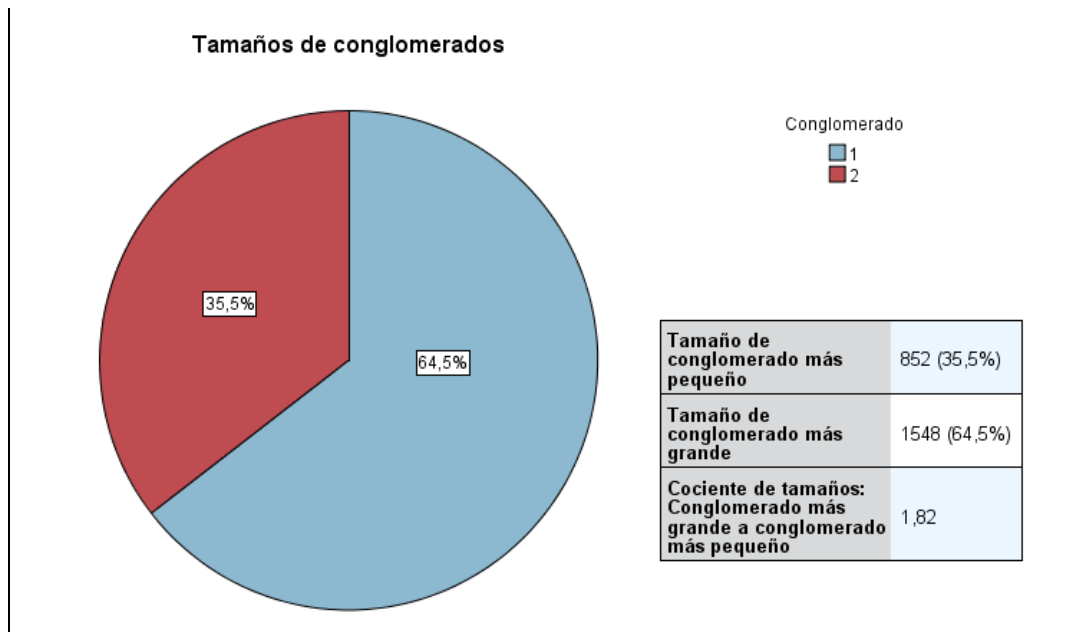


En el gráfico N°4 muestra el resumen del algoritmo, donde indica el tipo de algoritmo empleado en el análisis Conglomerado Bietápico, se utilizó 11 variables empleadas obteniendo 2 conglomerados.

También se observa la medida de calidad de los conglomerados obtenidos, en base a la relación y separación entre los conglomerados hallados, alcanzando un nivel aceptable.

Número de conglomerado

Gráfico N ° 05



Respecto al gráfico N°5 de tamaño de conglomerados, observamos que el tamaño del conglomerado más pequeño (conglomerado 2) corresponde al 35.5% (852 observaciones), y el tamaño del conglomerado más grande (conglomerado 1) corresponde al 64.5% (1548 observaciones). El ratio o cociente de tamaños entre el conglomerado más grande y el pequeño es de 1.82 eso quiere decir que el tamaño del conglomerado 1 es 1.82 veces el tamaño del conglomerado 2.

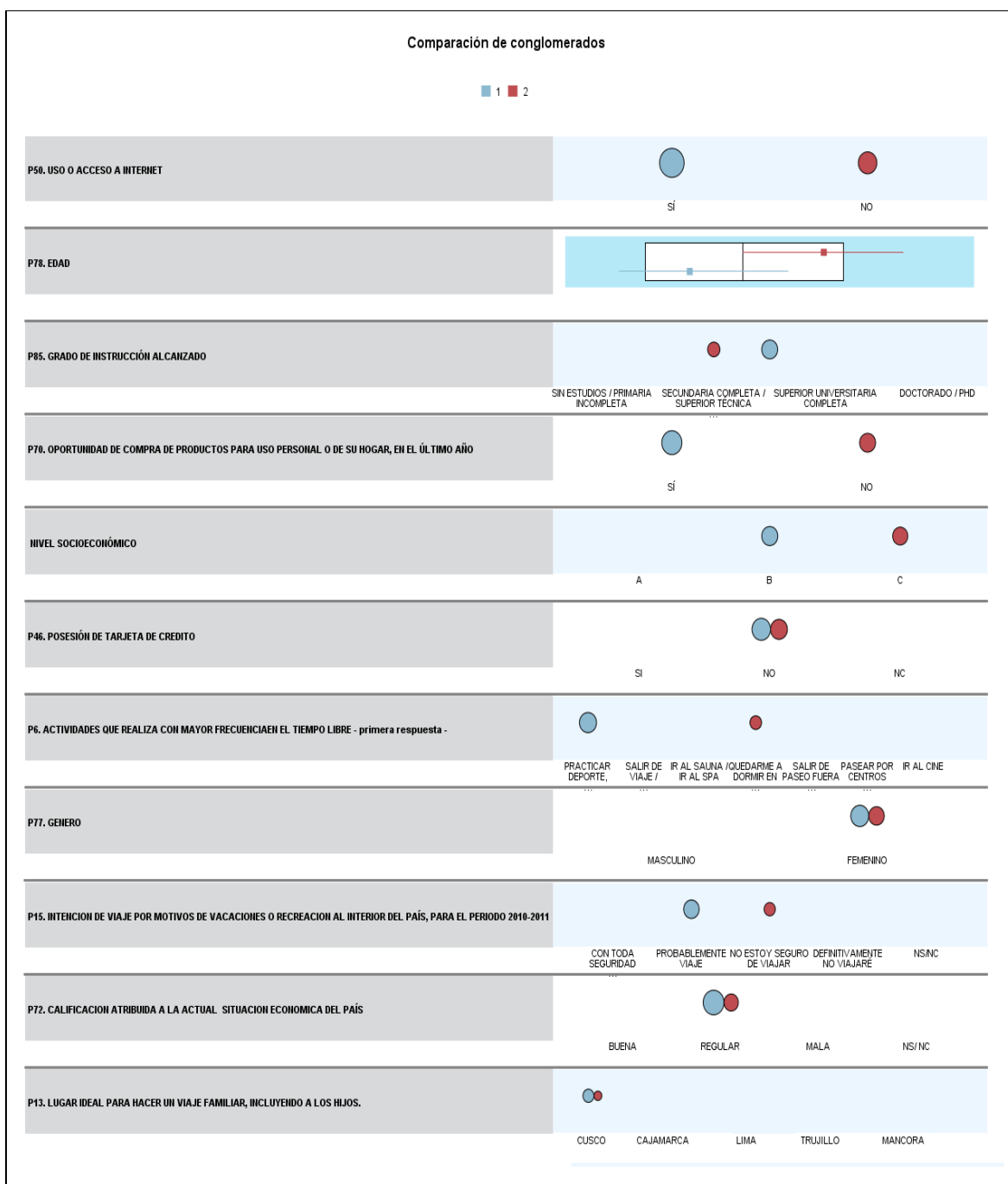
Gráfico N° 06

Conglomerados
Importancia de entrada (predictor)
■ 1,0

Conglomerado	1	2
Etiqueta		
Descripción		
Tamaño	64.5% (1548)	35.5% (852)
Entradas	P50. USO O ACCESO A INTERNET SI (96,1%)	P50. USO O ACCESO A INTERNET NO (99,2%)
	P78. EDAD 34,53	P78. EDAD 46,33
	P85. GRADO DE INSTRUCCION SUP. TECNICA COMPLETA / SUP. UNIVERSITARIA INCOMPLETA (40,8%)	P85. GRADO DE INSTRUCCION SECUNDARIA COMPLETA / SUP. TECNICA INCOMPLETA (44,0%)
	P70. OPORTUNIDAD DE COMPRA DE PRODUCTOS PRODUCTOS DE TECNOLOGIA SI (65,0%)	P70. OPORTUNIDAD DE COMPRA DE PRODUCTOS PRODUCTOS DE TECNOLOGIA NO(73,5%)
	P79. NIVEL SOCIOECONOMICO B(41,4%)	P79. NIVEL SOCIOECONOMICO C(66,3%)
	P46. POSESION DE TARJETA DE CREDITO NO (55,2%)	P46. POSESION DE TARJETA DE CREDITO NO(83,2%)
	P6. ACTIVIDADES QUE REALIZA CON MAYOR FRECUENCIA EN EL TIEMPO LIBRE PRACTICAR DEPORTE ACUDIR AL GIMNASIO(46,3%)	P6. ACTIVIDADES QUE REALIZA CON MAYOR FRECUENCIA EN EL TIEMPO LIBRE DORMIR EN CASA/ HACER UNA SIESTA (38,6%)
	P77. GENERO FEMENINO (51,6%)	P77. GENERO FEMENINO (70,5%)
	P15. INTENCION DE VIAJE POR MOTIVOS VACACIONES AL INTERIOR DEL PAIS PROBABLEMENTE (38,5%)	P15. INTENCION DE VIAJE POR MOTIVOS VACACIONES AL INTERIOR DEL PAIS NO ESTOY SEGURO DE VIAJAR (60,7%)
	P72. CALIFICACION ATRIBUIDA A LA SITUACION ECONOMICA DEL PAIS REGULAR (69,3%)	P72. CALIFICACION ATRIBUIDA A LA SITUACION ECONOMICA DEL PAIS REGULAR (60,7%)
	P13. LUGAR IDEAL PARA HACER UN VIAJE FAMILIAR INCLUYENDO A LOS HIJOS CUSCO(19,6%)	P13. LUGAR IDEAL PARA HACER UN VIAJE FAMILIAR INCLUYENDO A LOS HIJOS CUSCO(18,1%)

En el gráfico N°6 muestra el nivel de significancia de las hipótesis de importancia de variables, con la que se demuestra la evidencia estadística para afirmar que todas las variables resultaron ser importantes para el análisis.

Gráfico N° 07



El gráfico anterior compara el tamaño de ambos conglomerados en relación a su participación que poseen las variables dentro de cada uno:

- En el caso de las variables cualitativas, se denota con un círculo de diferente color (azul para el conglomerado 1 y rojo para el conglomerado 2) la categoría más frecuente. El tamaño del círculo permite comparar el porcentaje de la categoría respectiva dentro de cada conglomerado.

- En el caso de la única variable cuantitativa presente, edad, se presenta un diagrama de caja para ambos conglomerados, representando la distribución de la variable en cada uno mediante una línea del color definido anteriormente. En la caja se muestra la mediana global representada por una línea vertical en el interior de la misma, así como la mediana para cada conglomerado, representada por un cuadrado en la línea respectiva. El tamaño de cada línea se encuentra dado por el primer y tercer cuartil y los extremos verticales de la caja corresponden al primer y tercer cuartil global.

Del mismo modo que en el gráfico de importancia, los valores de las medianas y cuartiles globales y de cada conglomerado no se visualizan directamente, por lo que es necesario el construir una tabla N°13 en base a los valores que brinda el programa acercando el cursor a la línea respectiva.

Estadísticos por conglomerado para la variable Edad

Tabla N° 13

Estadístico	Conglomerado 1	Conglomerado 2	Global
Mediana	32.0	47.09	37.96
Primer Cuartil	24.03	37.94	27.03
Tercer Cuartil	43.06	55.96	49.28

Conglomerado 1:

Como podemos observar, en el cuadro anterior se muestra las medidas resumen de cada variable, dentro de ambos conglomerados, caracterizando cada uno de ellos.

De acuerdo a lo anterior el conglomerado 1 posee 1548 observaciones (equivalente al 64.58%), se encontrará caracterizado por las siguientes descripciones:

- Posee uso o acceso a internet (96.1%).
- La edad media es de 35 años.
- El grado de instrucción alcanzado es Superior Técnica Completa o Superior Universitaria incompleta (40.8%).
- Ha tenido oportunidad de comprar productos para uso personal o de su hogar en el último año (65.0%).
- Pertenece al nivel socioeconómico B (41.4%).
- No posee tarjeta de crédito (55.2%).

- Las actividades que realiza con mayor frecuencia en su tiempo libre son practicar algún deporte o acudir al gimnasio (46.3%).
- Pertenecen principalmente al sexo femenino (51.6%).
- Su intención de viaje por motivos de vacaciones o recreación al interior del país para el periodo 2010-2011 es que probablemente viaje (38.5%).
- La calificación que atribuye a la actual situación económica del país es regular (69.3%).
- El lugar ideal que considera para hacer un viaje familiar incluyendo a los hijos es la ciudad del Cusco (19.6%).

Conglomerado 2:

Del mismo modo el conglomerado 2 conformado de 852 observaciones (equivalente al 35.5%), se encontrará caracterizado porque la mayoría:

- Posee uso o acceso a internet (99.2%).
- La edad media es de 47 años.
- El grado de instrucción alcanzado es Secundaria Completa o Superior Técnica Incompleta (44.0%).
- No ha tenido oportunidad de comprar productos para uso personal o de su hogar en el último año (73.5%).
- Pertenecen al nivel socioeconómico C (66.3%).
- No posee tarjeta de crédito (83.2%).
- Las actividades que realiza con mayor frecuencia en su tiempo libre es quedarse en casa o hacer una siesta (38.6%).
- Pertenecen principalmente al sexo femenino (70.5%).
- Su intención de viaje por motivos de vacaciones o recreación al interior del país para el periodo 2010-2011 es que no está seguro de viajar (36.9%).
- La calificación que atribuye a la actual situación económica del país es regular (60.7%).
- El lugar ideal que considera para hacer un viaje familiar incluyendo a los hijos es la ciudad del Cusco (18.1%).

6.3 Validación de los conglomerados asignados

Con el objetivo de efectuar la validación de los grupos, en función de las variables seleccionadas, se efectuó un análisis de árbol de decisión, ya que la variable dependiente (número de conglomerado) es una variable cualitativa. Para lo cual se procede a seleccionar muestras de entrenamiento y prueba.

La regla utilizada en el presente trabajo, se basa en el principio de Pareto es también conocida como la regla del 80-20. El principio nos dice que “el 80% de los efectos que se observan en cualquier gran sistema son producidos por un 20% de las variables del sistema.

Muestra de entrenamiento y prueba:

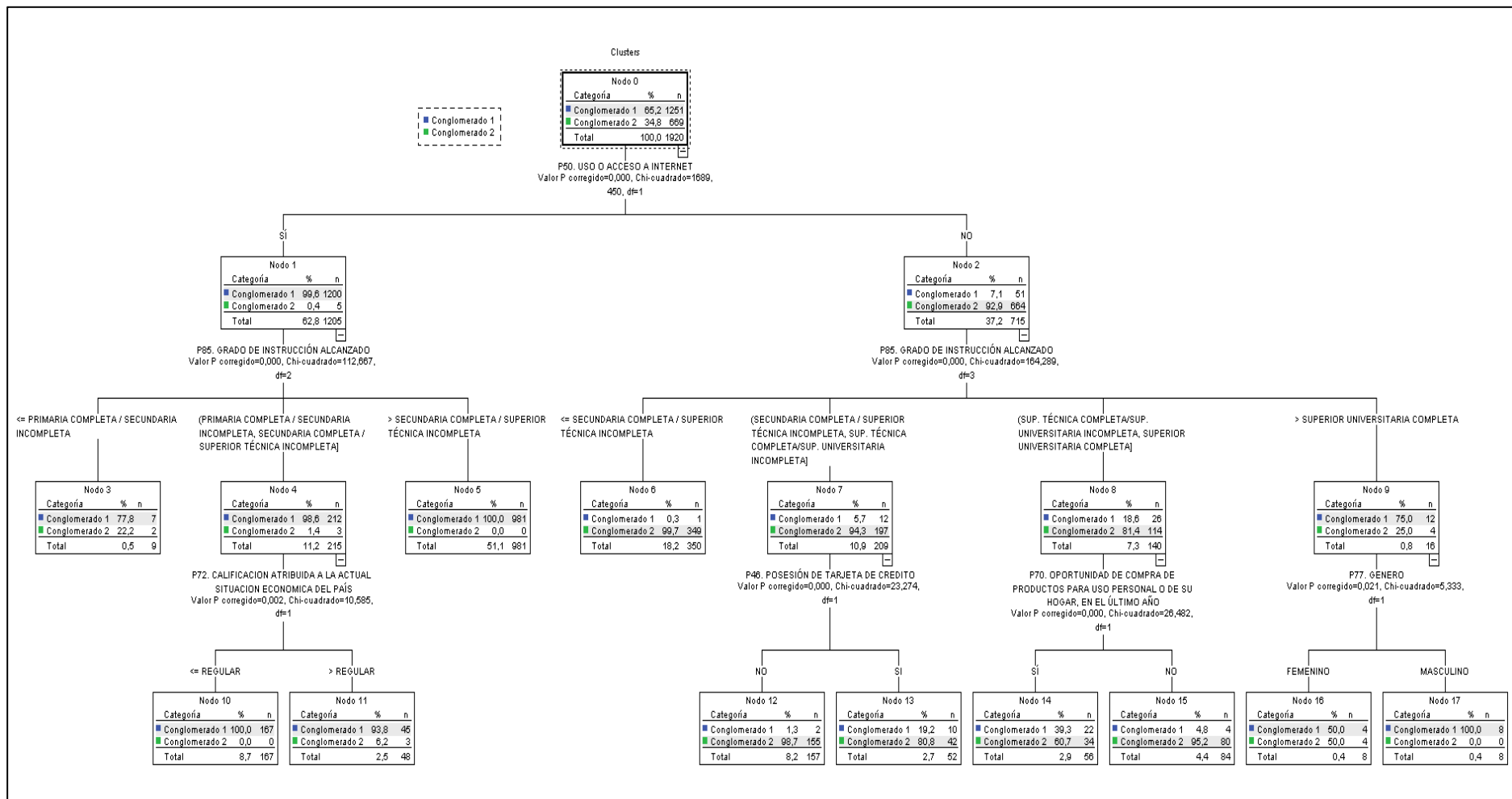
Se divide el conjunto total de los datos en dos subconjuntos complementarios e independientes entre sí: muestra de entrenamiento y muestra de prueba. Se extrae en forma aleatoria una muestra correspondiente al 80% del total de los datos y como muestra de prueba el 20% restante.

Se entrenó el modelo con 1920 casos y se realizó el test con 480 casos, de esta forma se halló el error cometido del modelo utilizado.

Se determinó un número mínimo de casos tanto en el nodo parental como en los nodos terminales igual a 5 y se ejecutó el algoritmo en el software SPSS (versión 22) obteniendo el siguiente resultado que se muestra en el gráfico N° 08.

Gráfico N° 08

Árbol de la muestra de entrenamiento



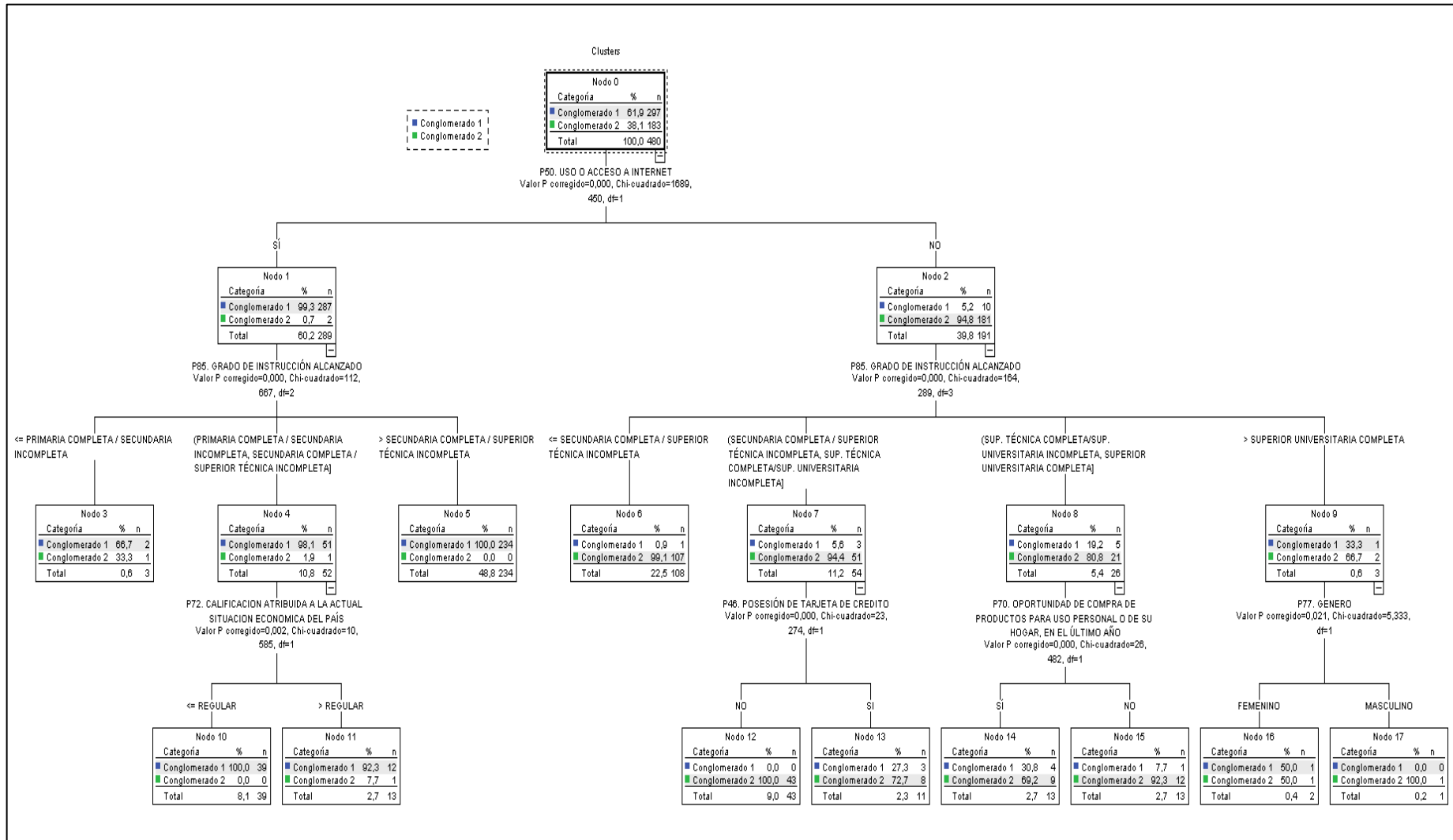
La variable uso o acceso a internet resulta ser el mejor predictor para los grupos hallados. Para la variable uso o acceso de internet el mejor predictor resulta la variable grado de instrucción. De modo que para el grado de instrucción primaria completa/secundaria incompleta y secundaria completa/superior técnica incompleta, el único predictor significativo es la variable calificación atribuida a la actual situación económica del país. Mientras que para el grado de instrucción secundaria completa/superior técnica incompleta y superior técnica completa/superior universitaria incompleta, el único predictor significativo en el modelo es la variable posesión de tarjeta de crédito.

Por otro lado, para la variable grado de instrucción superior técnica completa/superior universitaria incompleta y superior el único predictor significativo hallado es la variable oportunidad de compra de productos tecnológicos para uso personal o de su hogar en el último año. No obstante para la variable grado de instrucción superior universitaria completa, el único predictor significativo es la variable género.

Como paso siguiente se procedió a realizar la validación con los individuos seleccionados en la muestra prueba. Los resultados obtenidos se muestran en el siguiente gráfico.

Gráfico N° 09

Árbol de la muestra de prueba



Como se observa en los gráficos anteriores (Gráfico N°08 y N°09) se obtiene en la muestra de prueba la estructura de árbol similar a la muestra de entrenamiento, por lo que se concluye que la segmentación realizada es adecuada. A continuación se presenta las tablas que ayudarán a entender los resultados obtenidos de la validación.

Tabla N° 14

Riesgo

Muestra	Estimación	Típ. Error
Entrenamiento	,025	,004
Contraste	,027	,007

Métodos de crecimiento: CHAID

Variable dependiente: Clusters

Como se observa, el riesgo o error de clasificación resulta ser de apenas 2.5% para la data de entrenamiento, mientras que para la data de prueba es de 2.7%, siendo el error estándar del riesgo de 0.004 para la data de entrenamiento y de 0.007 para la de prueba.

Tabla N° 15

Clasificación

Muestra	Observado	Pronosticado		
		Conglomerado 1	Conglomerado 2	Porcentaje correcto
Entrenamiento	Conglomerado 1	1212	39	96,9%
	Conglomerado 2	9	660	98,7%
	Porcentaje global	63,6%	36,4%	97,5%
Contraste	Conglomerado 1	288	9	97,0%
	Conglomerado 2	4	179	97,8%
	Porcentaje global	60,8%	39,2%	97,3%

Métodos de crecimiento: CHAID

Variable dependiente: Clusters

La tabla de clasificación muestra los resultados obtenidos en la data de prueba o entrenamiento, se ha clasificado correctamente el 97.0% del conglomerado 1 y el 97.8% del conglomerado 2, siendo el porcentaje de clasificación global igual a 97.3%, por lo que se concluye como correcta la asignación de los conglomerados.

VII. CONCLUSIONES

- Los resultados del presente estudio indican la existencia de distintas de dos grupos diferenciados.
- Las variables que caracterizaron los grupos están conformados por características como la edad, género, nivel socioeconómico, grado de instrucción, intención de viaje, calificación de la situación económica actual del país, actividades más frecuentes en el tiempo libre, posesión de tarjeta de crédito, uso de internet, oportunidad de compra de productos tecnológicos en el último año.
- El algoritmo bietápico que ha dado una salida de dos conglomerados, calcula una medida de cohesión y separación que indica una calidad suficiente de conglomerados.
- La validación del modelo mediante árbol de clasificación CHAID mostró un mínimo error, tomando un 80% de la base de datos como entrenamiento y el 20% como prueba.
- Finalmente, cabe señalar que la presente investigación sobre los perfiles de turistas internos de los niveles socioeconómicos A, B y C tiene implicaciones ciertamente relevantes desde el punto de vista económico. Los datos obtenidos contribuyen a entender mejor los perfiles de turistas nacionales que existen dentro del Perú.

VIII. RECOMENDACIONES

- Limpieza de datos previo al análisis.
- Se recomienda intensificar publicidad de interés para el turista nacional mediante internet.
- Extender la presente investigación a más ciudades del Perú significativas para la actividad “turismo”, para así conocer mejor los tipos de turistas nacionales.

IX. BIBLIOGRAFÍA

- Bacher, J; Wenzing; K; Vogler, M. 2004. SPSS Two-Step Cluster: A first Evaluation. Universitat Erlangen-Nurnberg. Alemania, Lehrstuhl für Soziologie. 23p.
- García, J. (2007). *Turismo: Variantes e impactos* (2 ed.) La Habana: Editorial Universitaria.
- Martínez, V. (2006). *Ocio y turismo en la sociedad actual: los viajes, el tiempo libre y el entretenimiento en el mundo globalizado*. Madrid: McGraw Hill.
- Marzal, M. (2007). Maras: Pueblo en camino hacia el desarrollo turístico. *Espacio y Desarrollo*, 1-92. Extraído el 10 de agosto del 2011 de la base de datos EBSCO Host Academic Search Elite.
- MINCETUR (2010, 15 de julio). *El Turismo Rural*.
http://www.mincetur.gob.pe/turismo/LEGAL/LINEAMIENTOS_TURISMO_RURAL.pdf
- Norusis, M. 2011. Cluster Analysis. A Concise Guide to Market Research, Springer-Verlag. Cap. 9.
- RAPTRAVEL (2005, 02 de octubre). *Actividades y Turismo Vivencial*.
<http://www.raptravel.org/actividades/vivencial.php>
- SPSS - The SPSS TwoStep Cluster Component - SPSS Schweiz
Johann Bacher , Knut Wenzig, Melanie Vogler - SPSS Twostep Cluster - A first evaluation.
- Tom Chiu, Dong Ping Fang, John Chen, Yao Wang - Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment, 2001 SPSS - IBM SPSS Statistics Base 19

X. ANEXOS

ANEXO N°1

P15. DÍGAME, PARA EL PERIODO 2010 - 2011, ¿CUÁL ES SU INTENCIÓN DE VIAJE POR MOTIVOS DE VACACIONES O RECREACIÓN AL INTERIOR DEL PAÍS? (E: **MOSTRAR TARJETA**) (252)

CON TODA SEGURIDAD VIAJARÉ
PROBABLEMENTE VIAJE
NO ESTOY SEGURO DE VIAJAR
DEFINITIVAMENTE NO VIAJARÉ
NS/NC

P6. AHORA, EN SU TIEMPO LIBRE, ¿QUÉ ACTIVIDADES HACE CON MAYOR FRECUENCIA? (E: MÚLTIPLES RESPUESTAS ESPONTÁNEAS).

ACTIVIDADES	ACTIVIDADES (MÚLTIPLE)
Practicar deporte, acudir al gimnasio	1
Quedarme a dormir en casa / hacer una siesta	2
Reunirme con amigos o familiares en casa	3
Salir de paseo fuera de la ciudad y regresar el mismo día	4
Pasear por centros comerciales dentro de la ciudad	5
Salir a comer o almorzar a restaurantes	6
Ir al cine	7
Ir al teatro	8
Jugar juegos de video (tv, computadora)	9
Estar en internet	10
Salir de viaje / hacer turismo	11
Hacer tareas del hogar (limpiar, lavar, cocinar, etc.)	12
Ir al sauna / ir al spa	13
Ir a discotecas / pubs	14
Ir a casinos	15
Ver tv o dvd en casa	16
OTROS	
NS/NC	98

P46. ¿TIENE UD. TARJETA(S) DE CRÉDITO? (400)

1. SÍ
2. NO
3. NC

P50. DÍGAME, ¿ACOSTUMBRA USAR O ACCEDER A INTERNET? (420)

1. SÍ
2. NO

P70. EN EL ÚLTIMO AÑO, ¿HA COMPRADO ALGUNO DE ESTOS PRODUCTOS PARA SU USO PERSONAL O DE SU HOGAR? (E: **MOSTRAR TARJETA. UNA RESPUESTA**)

• Laptop	• MP3	• TV LCD, PLASMA
• NETBOOK	• IPOD	• NINTENDO WII
• USB 8	• IPHONE	• PLAY STATION II Ó III
• GIGABYTES	• BLACKBERRY	

1. SÍ
2. NO

P72. DÍGAME, ¿CÓMO CALIFICARÍA A LA ACTUAL SITUACIÓN ECONÓMICA DEL PAÍS (E: COMPLETAR LA PREGUNTA CON CADA FRASE), COMO BUENA, REGULAR O MALA? (E: UNA RESPUESTA POR CADA FRASE, ANOTAR EN EL RECUADRO)

1. BUENA
2. REGULAR
3. MALA
4. NS/NC

P13. LUGAR IDEAL PARA HACER UN VIAJE FAMILIAR INCLUYENDO A LOS HIJOS

P77. GÉNERO (963)

1. MASCULINO
2. FEMENINO

P78. POR FAVOR, ¿PODRÍA DECIRME CUÁNTOS AÑOS TIENE USTED? (E: ANOTAR EDAD EXACTA) (964)

_____ AÑOS

P85. ¿CUÁL ES EL GRADO DE INSTRUCCIÓN MÁS ALTO QUE HA ALCANZADO? (971)

1. SIN ESTUDIOS / PRIMARIA INCOMPLETA
2. PRIMARIA COMPLETA / SECUNDARIA INCOMPLETA
3. SECUNDARIA COMPLETA / SUPERIOR TÉCNICA INCOMPLETA
4. SUP. TÉCNICA COMPLETA/SUP. UNIVERSITARIA INCOMPLETA
5. SUPERIOR UNIVERSITARIA COMPLETA
6. POST GRADO /MAESTRÍA
7. DOCTORADO / PHD

P79. NSE (963)

1. A
2. B
3. C
4. D
5. E

ANEXO N° 2

ESTADISTICAS DESCRIPTIVAS DE LOS TURISTAS INTERNOS

Gráfico N° A-01

Distribución del uso o acceso a internet para el conglomerado 1

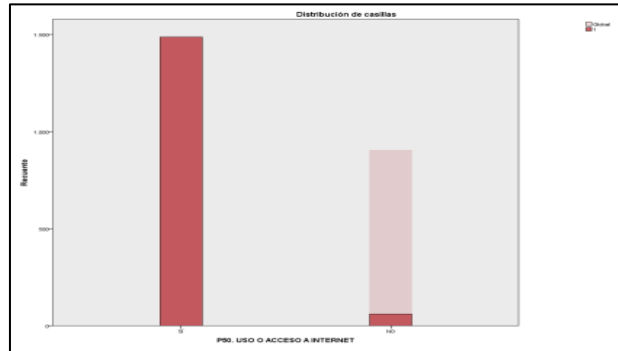


Gráfico N° A-02

Distribución del uso o acceso a internet para el conglomerado 2

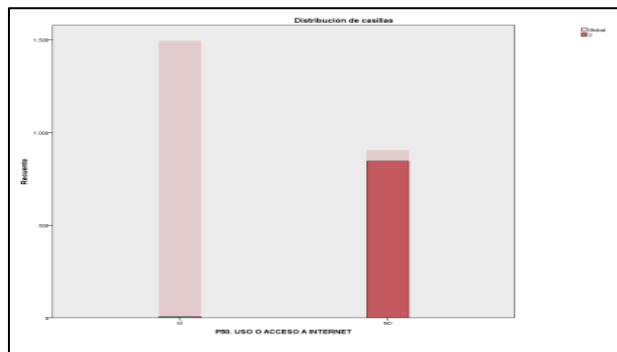


Gráfico N° A-03

Distribución de la edad para el conglomerado 1

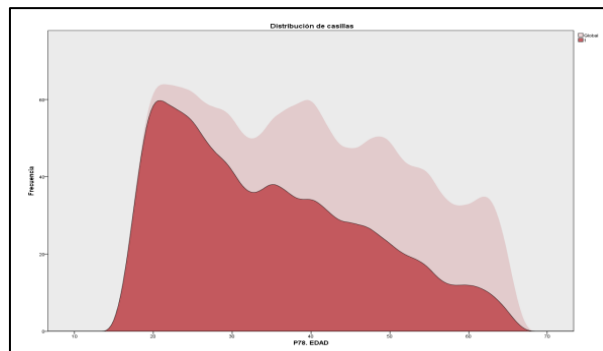


Gráfico N° A-04

Distribución de la edad para el conglomerado 2

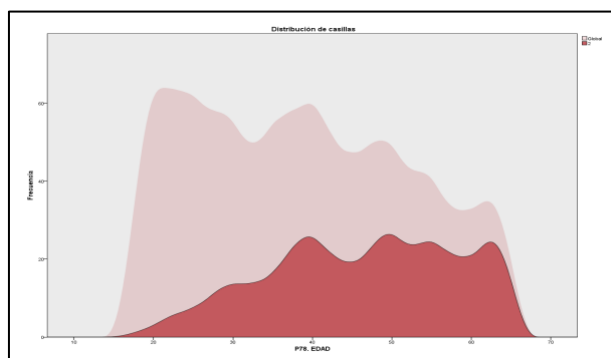


Gráfico N° A-05

Distribución del grado de instrucción alcanzado para el conglomerado 1

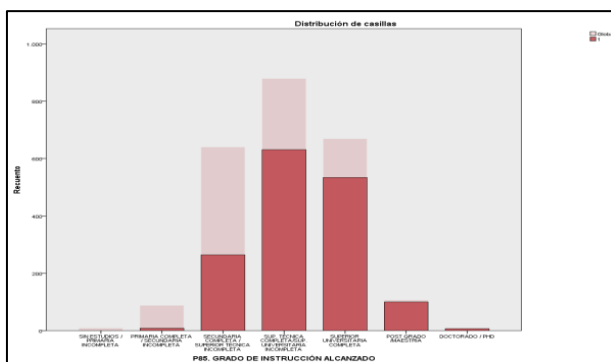


Gráfico N° A-06

Distribución del grado de instrucción alcanzado para el conglomerado 2

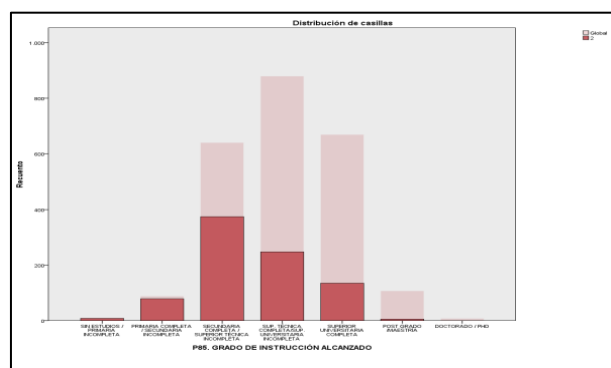


Gráfico N° A-07

Distribución de la oportunidad de compra de productos tecnológicos para uso personal o de su hogar para el conglomerado 1

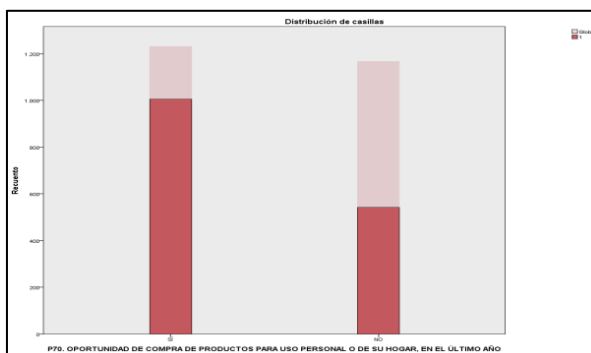


Gráfico N° A-08

Distribución de la oportunidad de compra de productos tecnológicos para uso personal o de su hogar para el conglomerado 2

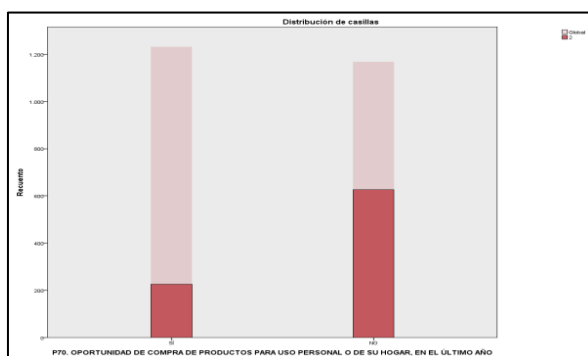


Gráfico N° A-09

Distribución del nivel socioeconómico para el conglomerado 1

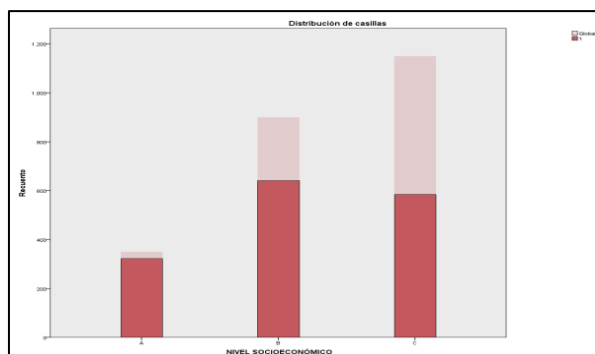


Gráfico N° A10

Distribución del nivel socioeconómico para el conglomerado 2

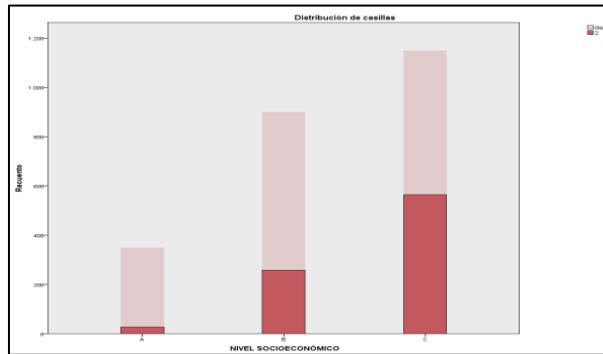


Gráfico N° A11

Distribución de la posesión de tarjeta de crédito para el conglomerado 1

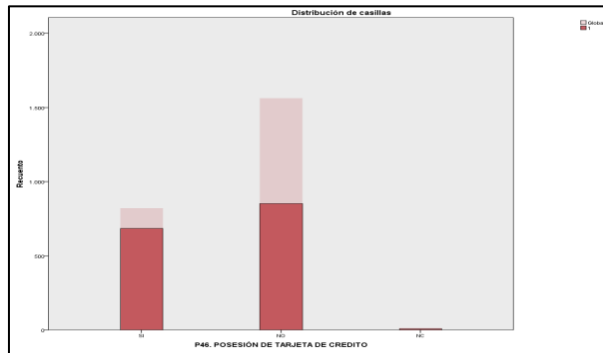


Gráfico N° A12

Distribución de la posesión de tarjeta de crédito para el conglomerado 2

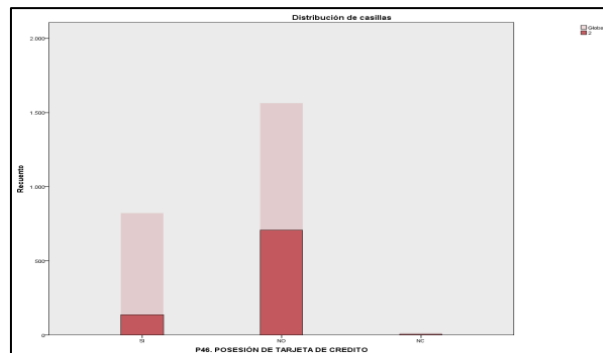


Gráfico N° A13

Distribución de las actividades que realiza con mayor frecuencia en el tiempo libre para el conglomerado 1

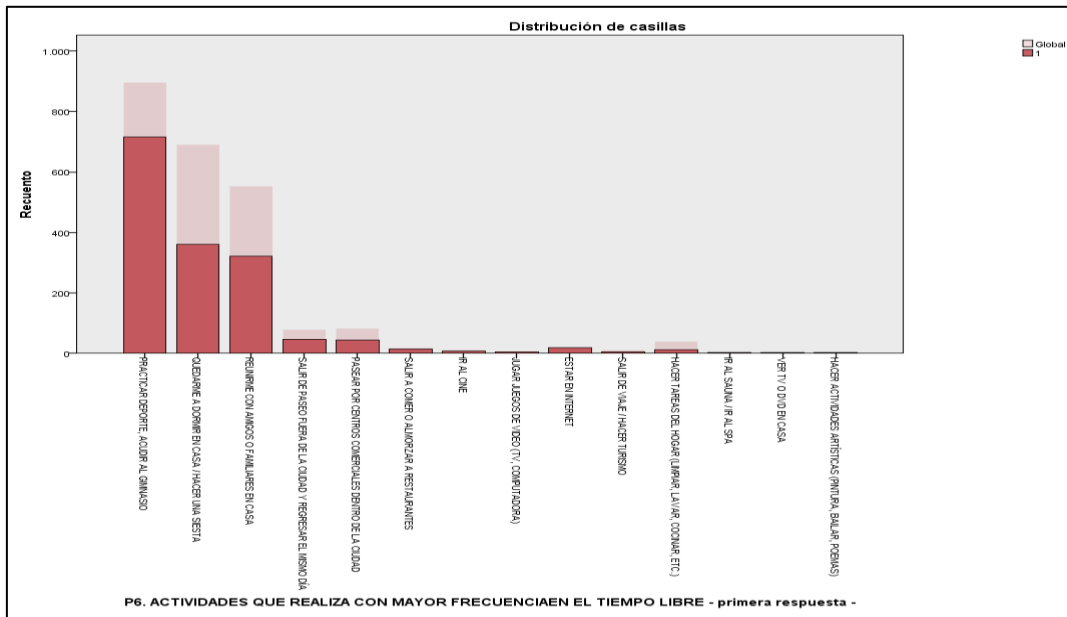


Gráfico N° A14

Distribución de las actividades que realiza con mayor frecuencia en el tiempo libre para el conglomerado 2

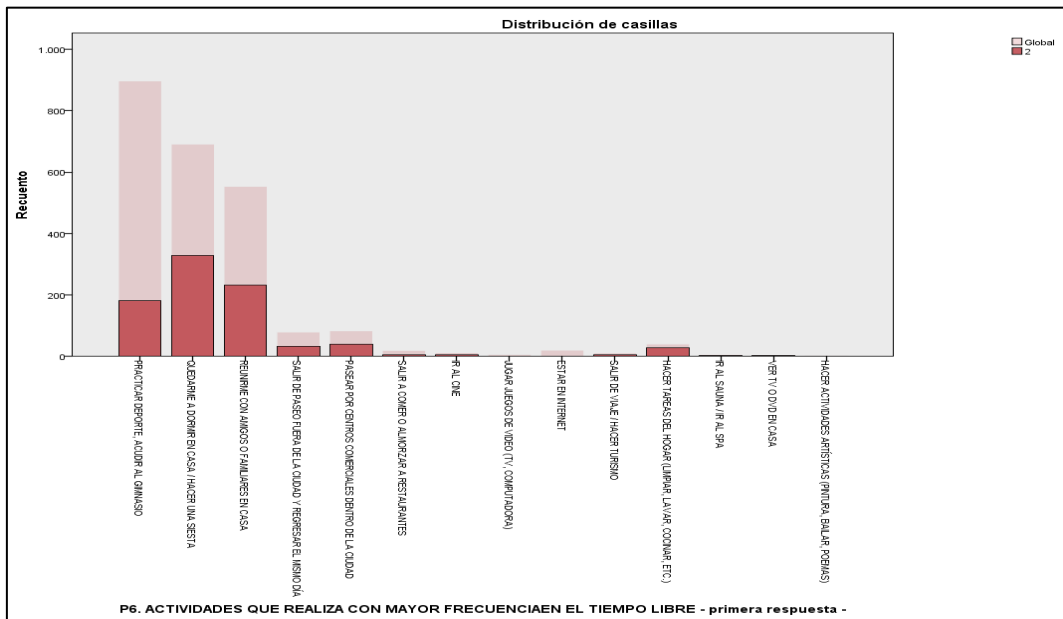


Gráfico N° A15

Distribución de género para el conglomerado 1

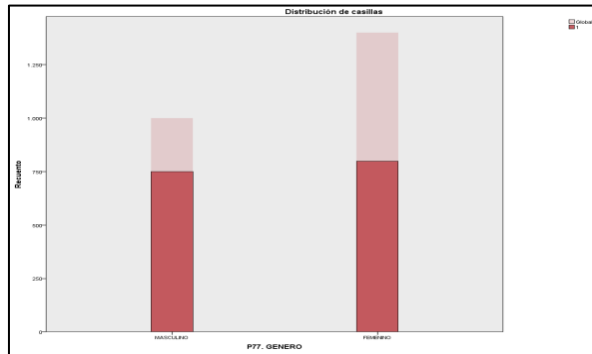


Gráfico N° A16

Distribución de género para el conglomerado 2

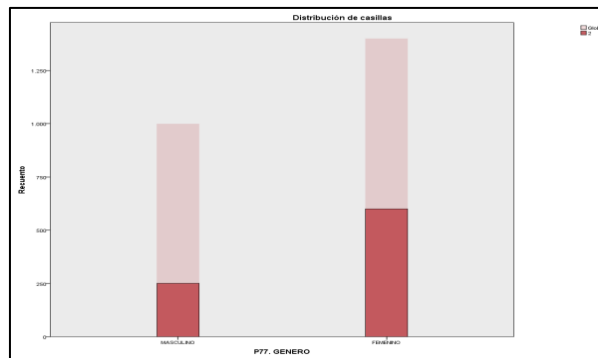


Gráfico N° A17

Distribución de la intención de viaje por motivos de vacaciones o recreacional para el conglomerado 1

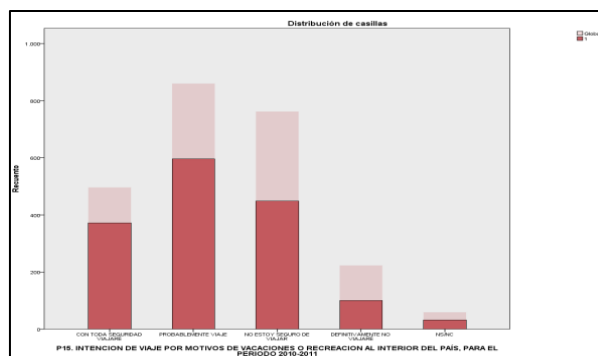


Gráfico N° A18

Distribución de la intención de viaje por motivos de vacaciones o recreacional para el conglomerado 2

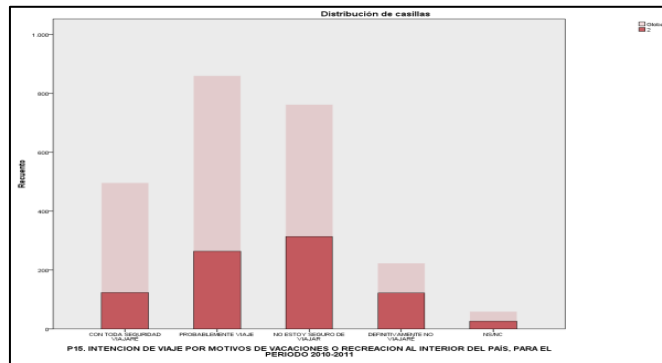


Gráfico N° A19

Distribución de la calificación atribuida a la actual situación económica del país para el conglomerado 1

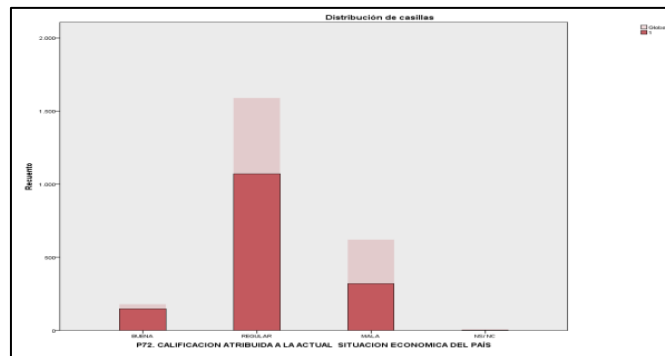


Gráfico N° A20

Distribución de la calificación atribuida a la actual situación económica del país para el conglomerado 2

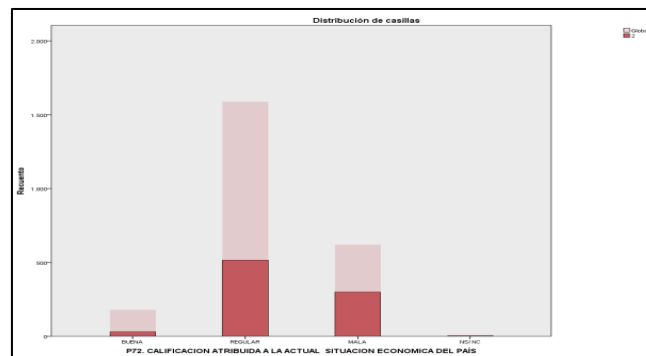


Gráfico N° A21

Distribución del lugar ideal para hacer un viaje familiar para el conglomerado 1

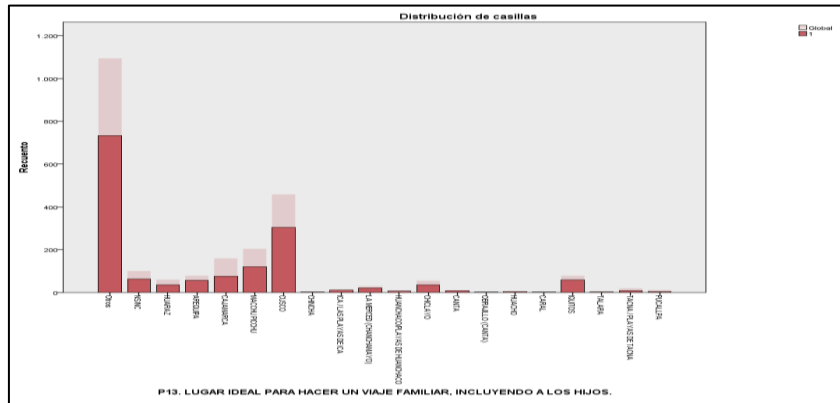


Gráfico N° A22

Distribución del lugar ideal para hacer un viaje familiar para el conglomerado 2

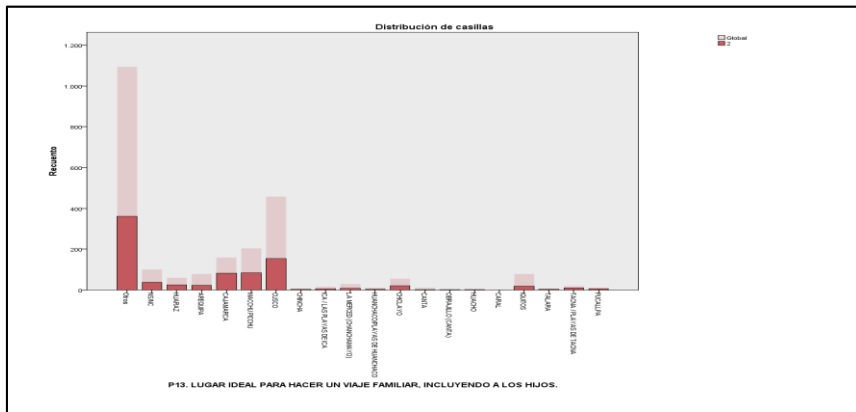


Gráfico N° A23

Distribución de la edad del turista interno

