



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Académico Profesional de Estadística

**Caracterización de los clientes de la Clínica
Odontológica de la UNMSM. Una aplicación del
análisis de correspondencias múltiples**

TESINA

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Flavio Francisco ROMANÍ RAMOS

ASESOR

Wilfredo Eugenio DOMINGUEZ CIRILO

Lima, Perú

2007



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Romaní F. Caracterización de los clientes de la Clínica Odontológica de la UNMSM. Una aplicación del análisis de correspondencias múltiples [Tesina]. Lima: Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Académico Profesional de Estadística; 2007.

RESUMEN

"CARACTERIZACIÓN DE LOS CLIENTES DE LA CLÍNICA ODONTOLÓGICA DE LA UNMSM. UNA APLICACIÓN DEL ANALISIS DE CORRESPONDENCIAS MÚLTIPLES"

FLAVIO FRANCISCO ROMANI RAMOS

2007

Asesor: Mg. Wilfredo Domínguez

Título Obtenido: Licenciado en Estadística

En este trabajo de investigación se realiza la caracterización de los pacientes externos de La Clínica de la facultad de Odontológica de la Universidad Nacional de San Marcos, los cuales fueron analizados por la técnica multivariante de análisis de correspondencia Múltiple. Según los resultados los pacientes que asisten a la clínica odontológica tienen como mayor fuente de publicidad las recomendaciones, son personas adultas y adultas mayores entre 40 a 50 años, consideran el precio cómodo y buen servicio en la atención en la clínica 1,2 y 3.

Palabra claves: Análisis de correspondencia múltiple
Mapas perceptuales
Servicio

ABSTRACT

“UNMSM DENTAL CLINIC’S CLIENTS CHARACTERIZATION. AN APPLICATION OF THE MULTIPLE CORRESPONDENCES ANALYSIS”

**ROMANI RAMOS FLAVIO FRANCISCO
2007**

**Adviser: Mg. Wilfredo Dominguez
Obtained Title: Statistic Licenciature**

In this researching work, the characterization of the external patients of the UNMSM Dental Faculty Clinic is done, which were analyzed using the Analysis Correspondence Multivariate Technics.

According to the results, the attending patients to the dental clinic take as their greater publicity source recommendations, they are adult and major adult persons aged between 40 to 50 years. They consider 1, 2 and 3 clinics to have a comfortable price and good service in the attention.

KEY WORDS:

Multiple Correspondences Analysis
Perceptual Maps
Service

INTRODUCCIÓN

En 1915 se establece la primera Clínica dental en el local de la Facultad de Medicina en la antigua Calle Lechugal. Desde entonces se viene brindando atención odontológica a los pacientes, siendo la primera y única en esas épocas en donde se brindaba atención dental. En la actualidad, la Facultad de Odontología de la UNMSM no es la única que brinda servicios odontológicos ya que existen en Lima seis facultades de odontología, como: San Martín de Porras, Cayetano Heredia, Villarreal, Garcilazo de la Vega y Alas Peruanas, los cuales cuentan con sus respectivas clínicas. Por lo tanto se hace imprescindible por la competencia existente no solo estar frecuentemente a la vanguardia de los avances científicos y tecnológicos sino para tener una comprensión de las necesidades de los usuarios. Hoy en día la Clínica odontología de San Marcos tiene una moderna infraestructura y un prestigio que se fue ganando a través del tiempo. El propósito de esta monografía es principalmente aplicar técnicas estadísticas multivariadas particularmente el Análisis de Correspondencia Múltiple, ACM que permitirá sintetizar la información, esto es, pasar de unas variables categóricas originales a un pequeño número de nuevas variables o factores tales que sintetizen la información de las variables originales”, esto es, en la búsqueda de los factores más próximos con todos los grupos de modalidades o preguntas conjuntamente y de esta manera caracterizar al paciente y buscar perfiles de los mismos para descubrir interrelaciones entre las características medidas, poniéndolas de manifiesto en gráficos bidimensionales que pueden ser interpretados con relativa facilidad, tanto por los expertos como por los usuarios.

CAPITULO I

1.1 PLANTEAMIENTO DEL PROBLEMA

Actualmente vivimos en un mundo globalizado en el que se presentan innumerables cambios que han traído como consecuencia no solo modificaciones en el comportamiento de las sociedades, necesidades de las personas, si no también en los servicios que se brinda a los usuarios. Es así que los profesionales de la salud, quienes tienen como misión proporcionar cuidados de salud a la población, los cuales deberán satisfacer la expectativa que tienen los usuarios de ellos, tanto en el aspecto técnico como en el aspecto humanístico y el aspecto de su entorno.

La calidad de atención en los servicios de salud es responsabilidad del equipo de salud que labora en un hospital, en tal sentido la institución cobra particular importancia para brindar atención directa y estar en contacto el mayor tiempo con el paciente; de ahí que una atención interpersonal de buena calidad es necesario para lograr la satisfacción de las necesidades biopsico-espirituales de los pacientes, debido a la frecuencia de interacción, lo cual da la calidad de atención contribuye en nuestros tiempos al desarrollo de las diferentes actividades . En este sentido es importante conocer la demanda es decir tener una visión clara de las características y conocer la opinión del usuario externo acerca los servicios que se prestan en la clínica odontológica,

Actualmente no se sabe como este perfil del usuario para satisfacer sus necesidades de acuerdo a su realidad , necesidad económica y social , en tal sentido apelaremos a las técnicas estadísticas multivariadas de interdependencia, aplicación de Análisis de Correspondencia múltiple a través de este modelo caracterizaremos al pacientes y realizaremos un diagnóstico acerca en sus requerimiento y necesidades para promover estrategias de mercadeo en su efecto la Clínica esté mejor organizada, brindando un mejor servicio.

1.2 FORMULACIÓN DEL PROBLEMA

¿Cuál es la caracterización de los pacientes externos de La Clínica de la facultad de Odontológica de la Universidad Nacional de San Marcos

1.3 JUSTIFICACIÓN DEL PROBLEMA

En la Clínica Odontológica de la Universidad Nacional de San Marcos no se sabe cual es la característica de los pacientes, cuya importancia es muy valiosa para conseguir más pacientes y sobre todo fidelizar a los pacientes

1.4 LIMITACIONES

Una de las principales limitaciones enfrentadas en el desarrollo de la presente Investigación fue que se encontraron pocos estudios de trabajos similares en el Perú. Otra dificultad fue la falta de base de datos de los pacientes en un sistema de cómputo.

1.5 FORMULACIÓN DE OBJETIVOS

1.5.1 Objetivo General

Caracterizar a los pacientes externos de La Clínica de la facultad de Odontológica de la Universidad Nacional de San Marcos.

1.5.2 Objetivos Específicos

- Determinar las características de los pacientes de la clínica odontológica según publicidad.
- Caracterizar a los pacientes de la clínica odontológica según calidad de atención.
- Caracterizar a los pacientes de la clínica odontológica según comodidad de la sala de espera.
- Caracterizar a los pacientes de la clínica universitaria según el precio

- Caracterizar a los pacientes de la clínica universitaria según calidad del personal administrativo y el operador.
- Encontrar las características de los pacientes según ubicación de la clínica odontológica.

CAPITULO II

2.1 ASPECTOS TEÓRICOS

ANÁLISIS DE CORRESPONDENCIAS SIMPLES

El Análisis de Correspondencias Simple permite estudiar las relaciones de interdependencia entre variables categóricas, por tanto guarda cierta analogía con la prueba Chi cuadrado y el coeficiente de concordancia de Kendall. Es una técnica de composición, que se ha ido haciendo más popular para la reducción dimensional y la elaboración de mapas perceptuales; debido a que el mapa perceptual se basa en la asociación entre objetos y un conjunto de características descriptivas o atributos especificados por el investigador.

El análisis de correspondencia simple es una técnica para analizar la homogeneidad entre las categorías de cada uno de las dos variables respecto a las categorías de la otra.

Esta técnica tiene la ventaja que para su uso no se hace ninguna conjetura respecto a la distribución de probabilidad de la población de la cual se extrajo la muestra multivariante. Esta técnica va más allá de analizar la relación existente entre las variables, porque permite determinar cómo está estructurada esta asociación, describiendo “proximidades” que permiten identificar “categorías causa de asociación”.

Entre las técnicas de composición, el Análisis Factorial es el más parecido; pero el Análisis de Correspondencias va más allá del Análisis Factorial. Su aplicación mas directa es la representación de la “correspondencia” de categorías de variables, particularmente aquellas medidas con escalas nominales. Esta Correspondencia es la base de desarrollo de mapas perceptuales. Los beneficios del Análisis se basan en sus capacidades únicas para presentar filas y columnas, por ejemplo etiquetas y atributos en un mismo espacio.

Por tanto, el objetivo básico del Análisis de Correspondencias Simples (ACS), es el de obtener una representación gráfica (nube de puntos) de una tabla de contingencia, para descubrir asociaciones entre las categorías de dos variables cualitativas o cuantitativas categorizadas.

Este es un método exploratorio y requiere un número grande de observaciones.

COMO SE REALIZA EL ANALISIS DE CORRESPONDENCIAS SIMPLES

En principio, las debemos hacer una tabla de contingencia (ó tabla cruzada) con las dos variables cualitativas.

Luego debemos calcular una medida de homogenización, como ya dijimos anteriormente en estadística se utiliza la prueba de chi-cuadrado, que no es otra cosa que la diferencia elevada al cuadrado de la frecuencia observada menos la frecuencia esperada dividida sobre el total de datos observados.

Ahora toca la parte más importante, definir el número de dimensiones a utilizar. Como estamos estudiando el análisis de correspondencias simples, el número de dimensiones es dos ($k = 2$).

Puesto que estamos trabajando con una tabla cruzada, cada fila puede ser considerada como un punto dotado de masa, en un espacio de f dimensiones. De igual manera, cada columna puede ser considerada como un punto dotado de masa, en un espacio de c dimensiones. Con esto obtendremos un nuevo espacio conformado por C dimensiones:

$$C = \min (f,c) - 1$$

Teniendo en cuenta que cada punto tiene un peso o ponderación igual a su masa, un estadístico adecuado para medir la dispersión de la nube de puntos será la inercia. La inercia es el promedio de las distancias de los distintos puntos a su centro de gravedad, estando cada distancia ponderada por la masa del punto correspondiente. La inercia total será la misma tanto si la

nube de puntos corresponde a la representación de la filas como si corresponde a las columnas.

Luego obtenemos los cuadros de pesos, distancia al origen e inercia de los puntos-filas y de los puntos-columnas. En estos cuadros identificamos que factores son los más determinantes.

TABLAS DE CONTINGENCIA Y DE FRECUENCIAS

EL Análisis de Correspondencias de una tabla de contingencia persigue cuatro objetivos fundamentales los que son importantes detallar para justificar este método:

- Analizar toda la información contenida en una tabla de contingencia.
- Representar gráficamente la estructura de una tabla de contingencia.
- Producir estadísticas de control complementarias.
- Analizar la estructura de una tabla de contingencia respetando el hecho que la misma resume una relación simétrica entre los caracteres observados.

Una tabla de contingencia es una tabla estadística que resume la observación simultánea de dos características dadas de dos conjuntos de unidades de observación (población observada). Si el objetivo de lo observado es describir la manera en que esas características se asocian entre ellas a través de las co-ocurrencias observadas, esa tabla estadística constituye un resumen adecuado de la observación.

Consideremos una población o muestra con **K..** unidades (respuestas) sobre los que se pretende analizar simultáneamente dos variables categóricas, digamos la variable **X** que tiene **I** categorías y la variable **Y** que tiene **J** categorías; la tabla de frecuencias bidimensional que describe a estas **K..** unidades se llama tabla de contingencia o tabla de doble entrada, también se le conoce como tabla T:

TABLA DE DOBLE ENTRADA

TABLA T (I, J)

CATEGORÍAS		VARIABLE Y							
		1	2	3	...	j	...	J	TOTAL
VARIABLE X	1	k ₁₁	k ₁₂	k ₁₃	...	k _{1j}	...	k _{1J}	K_{1.}
	2	k ₂₁	k ₂₂	k ₂₃	...	k _{2j}	...	k _{2J}	K_{2.}
	3	k ₃₁	k ₃₂	k ₃₃	...	k _{3j}	...	k _{3J}	K_{3.}
	⋮	⋮	⋮	⋮	⋮	⋮
	i	k _{i1}	k _{i2}	k _{i3}	...	k _{ij}	...	k _{iJ}	K_{i.}
	⋮	⋮	⋮	⋮	⋮	⋮
	I	k _{I1}	k _{I2}	k _{I3}	...	k _{Ij}	...	k _{IJ}	K_{I.}
	TOTAL	K_{.1}	K_{.2}	K_{.3}	...	K_{.j}	...	K_{.J}	K_{..}

NOTACIONES:

- K_{ij} : Frecuencia bidimensional o Término general.
Corresponde a la cantidad de individuos observados que presentan simultáneamente la modalidad i de la variable X y la modalidad j de la variable Y .
- $K_{i.}$: Es el i -ésimo elemento del margen derecha de la tabla T (no depende de J), corresponde a la cantidad de individuos observados que presenta la modalidad i de la variable X (**fila**), y está definido por la siguiente expresión matemática:

$$K_{i.} = \sum_{j=1}^J k_{ij}; \forall i \in I = (1, 2, \dots, I)$$

- $K_{.j}$: Es el j -ésimo elemento del margen inferior de la tabla T (no depende de I), corresponde a la cantidad de individuos observados que presenta la modalidad j de la variable Y (**columna**), y está definido por la siguiente expresión matemática:

$$K_{.j} = \sum_{i=1}^I k_{ij}; \forall j \in J = (1, 2, \dots, J)$$

- $K_{..}$: Es la suma de cada una de las celdillas de la tabla, y corresponde al total de individuos observados. Esta expresado de la siguiente manera:

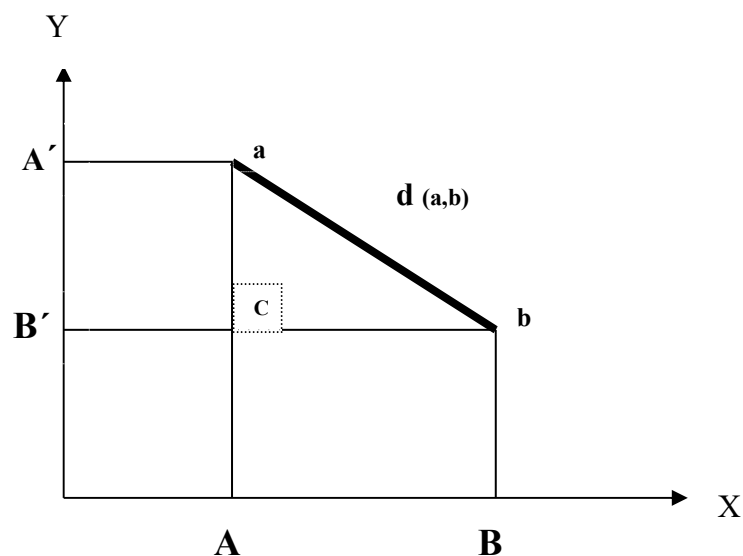
$$K_{..} = \sum_{i=1}^I \sum_{j=1}^J k_{ij} = \sum_{i=1}^I k_{i.} = \sum_{j=1}^J k_{.j}$$

REPRESENTACIÓN DE UNA TABLA DE CONTINGENCIA EN UN ESPACIO EUCLIDIANO

La representación analógica de una tabla de contingencia nos conduce a una apreciación visual de la posición relativa de puntos en un espacio **deformado por los efectos de perspectivas**. Esto puede producir proximidades o alejamientos ficticios entre los puntos, es decir relaciones inexistentes. Podemos evitar esos errores de lectura completando esos gráficos con una relación digital de la posición relativa de esos puntos en esos espacios.

Utilizando las propiedades formales de los espacios de representación euclidianos, podemos calcular la distancia entre dos puntos. También se puede evaluar la posición relativa de una serie de puntos de esos espacios, calculando la suma de todas las distancias, dos a dos, de todos esos puntos.

Recordando el teorema de Pitágoras, podemos calcular fácilmente las distancias entre dos puntos en un plano. Veamos la siguiente la figura:



Los ejes X , Y constituyen el referencial de ese espacio. Los puntos a y b están ubicados por sus coordenadas y sobre esos ejes. A es la coordenada de a sobre el eje X y A' es la coordenada de a sobre el eje Y . La distancia entre el punto a y el punto b está dada por la longitud del segmento de recta entre los puntos a y b .

Calculemos la distancia entre dos puntos. Por el teorema de Pitágoras sabemos que, el cuadrado de la distancia entre los puntos a y b , es igual a la suma de los cuadrados de las diferencias de coordenadas de los puntos a y b sobre los ejes X , Y . Es decir:

$$d^2_{(a,b)} = (B - A)^2 + (B' - A')^2$$

La distancia entre a y b está dada por:

$$d_{(a,b)} = \sqrt{(B - A)^2 + (B' - A')^2}$$

Generalizando al caso de espacios de tres dimensiones. La distancia entre dos puntos a y b , está dada por siguiente relación:

$$d_{(a,b)} = \sqrt{(B - A)^2 + (B' - A')^2 + (B'' - A'')^2}$$

DEFINICION DE LA DISTANCIA ENTRE ELEMENTOS DE UNA TABLA DE CONTINGENCIA

- La expresión general de la distancia entre dos elementos fila de una tabla $T(I,J)$ es la siguiente:

$$d_{(i,i')} = \sqrt{\sum_{j=1}^J (k_{ij} - k_{i'j})^2}; \forall i, i'$$

- La expresión general de la distancia entre dos elementos columna de una tabla $T(I,J)$ es la siguiente:

$$d_{(j,j')} = \sqrt{\sum_{i=1}^I (k_{ij} - k_{ij'})^2}; \forall j, j'$$

ÍNDICES DE SIMILITUD ENTRE LOS ELEMENTOS DE LA TABLA T

Hemos construido ese índice, que llamamos distancia entre dos elementos (fila y columna), a partir de propiedades de los espacios euclidianos que utilizamos como base de representación de las tablas de contingencia.

Demos representar que existe correspondencia de contenidos entre esos dos modos de representación de la información de la tabla de contingencia.

➤ Cuando comparamos dos elementos fila o columna de una tabla establecemos una relación de similitud o de diferencia entre ellos.

En términos de distancias en un espacio euclidiano, esas similitudes o diferencias se expresan por el hecho que los puntos-elementos del espacio se superponen (distancia nula) o no muestran una cierta distancia entre ellos. La distancia entre dos puntos de un espacio euclidiano es siempre nulo o valor positivo. Esto quiere decir que, para todo par de elementos de la tabla T:

$$d_{(i,i')} \geq 0, \forall i, i' \qquad d_{(j,j')} \geq 0, \forall j, j'$$

➤ La comparación de un elemento fila o columna consigo mismo es, en la práctica los resultados de observaciones reales. En términos de distancias en un espacio euclidiano, la similitud de los elementos de la tabla se traduce con la superposición (distancia nula) de los puntos en el espacio (coordenadas idénticas).

El índice de similitud entre elementos que hemos definido es nulo si y solamente si los comparados son idénticos. Para todo par de elementos de la tabla:

$$d_{(i,i')} = 0, \Leftrightarrow i = i' \qquad d_{(j,j')} = 0, \Leftrightarrow j = j'$$

➤ Cuando comparamos dos elementos de la tabla, establecemos la similitud o la diferencia de los mismos cualquiera sea el orden de la comparación. Así mismo en términos de distancias en un espacio euclidiano se traduce que la distancia entre dos puntos es la misma cualquiera sea el sentido (la distancia del punto **a** al punto **b** es la misma que la distancia **b** al punto **a**).

El índice de similitud entre elementos que hemos definido, no cambia el valor si permutamos el orden de los elementos. Es decir que, para todo par de elementos de la tabla:

$$d_{(i,i')} = d_{(i',i)}, \forall i, i' \quad \text{y} \quad d_{(j,j')} = d_{(j',j)}, \forall j, j'$$

TABLA DE FRECUENCIAS RELATIVAS

Esta tabla indica el porcentaje de elementos que simultáneamente tiene las dos categorías, conocida también como la **tabla F**:

CATEGORÍAS		VARIABLE Y							
		1	2	3	...	j	...	J	TOTAL
VARIABLE X	1	f_{11}	f_{12}	f_{13}	...	f_{1j}	...	f_{1J}	$f_{1.}$
	2	f_{21}	f_{22}	f_{23}	...	f_{2j}	...	f_{2J}	$f_{2.}$
	3	f_{31}	f_{32}	f_{33}	...	f_{3j}	...	f_{3J}	$f_{3.}$
	⋮	⋮	⋮	⋮	⋮	⋮
	i	f_{i1}	f_{i2}	f_{i3}	...	f_{ij}	...	f_{iJ}	$f_{i.}$
	⋮	⋮	⋮	⋮	⋮	⋮
	I	f_{I1}	f_{I2}	f_{I3}	...	f_{Ij}	...	f_{IJ}	$f_{I.}$
	TOTAL	$f_{.1}$	$f_{.2}$	$f_{.3}$...	$f_{.j}$...	$f_{.J}$	$f_{..}$

NOTACIONES:

➤ Frecuencia Relativa Conjunta:

$$f_{ij} = \frac{k_{ij}}{K_{..}}$$

En una tabla de frecuencias, cada celdilla es una proporción. El valor de cada celdilla corresponde a la proporción de individuos observados en la población que presentan simultáneamente la variable **X** y la variable **Y**.

➤ Frecuencias Marginales:

$$f_{i.} = \frac{k_{i.}}{K_{..}} = \sum_{j=1}^J \frac{k_{ij}}{K_{..}} = \sum_{j=1}^J f_{ij}$$

$$\forall i \in i = (1, 2, 3, \dots, I)$$

$$f_{.j} = \frac{k_{.j}}{K_{..}} = \sum_{i=1}^I \frac{k_{ij}}{K_{..}} = \sum_{i=1}^I f_{ij} \quad \forall j \in j = (1, 2, 3, \dots, J)$$

- La suma total o suma de los márgenes es evidentemente igual a 1, puesto que la tabla de frecuencias relativas se obtiene dividiendo la tabla de contingencia por $K_{..}$.

$$f_{..} = \sum_{i=1}^I \sum_{j=1}^J \frac{k_{ij}}{K_{..}} = \frac{1}{K_{..}} \sum_{i=1}^I \sum_{j=1}^J k_{ij} = \frac{K_{..}}{K_{..}} = 1$$

Las variables son independientes si se cumplen una de las siguientes condiciones:

- 1) $f_{ij} = f_{i.} \cdot f_{.j}; \forall i, j$
- 2) $f_{i/j} = f_{i.}, f_{j/i} = f_{.j}; \forall i, j$

ÍNDICES DE SIMILITUD ENTRE ELEMENTOS DE LA TABLA F:

- Definición del “índice de similitud” entre filas de una tabla F(IxJ)

$$S_{(i,i')} = \sum_{j=1}^J (f_{ij} - f_{i'j}) = \sum_{j=1}^J \left(\frac{k_{ij}}{k_{..}} - \frac{k_{i'j}}{k_{..}} \right); \forall i, i'$$

- Definición del “índice de similitud” entre columnas de una tabla F(IxJ)

$$S_{(j,j')} = \sum_{i=1}^I (f_{ij} - f_{ij'}) = \sum_{i=1}^I \left(\frac{k_{ij}}{k_{..}} - \frac{k_{ij'}}{k_{..}} \right); \forall j, j'$$

DEFINICION DEL MODO DIGITAL DE LAS DISTANCIAS DE LA TABLA F(I xJ)

- La expresión general de la distancia en el caso de la comparación de dos elementos filas de una tabla F(I, J) es la siguiente:

$$d_{(i,i')} = \sqrt{\sum_{j=1}^J \left(\frac{k_{ij}}{k_{..}} - \frac{k_{i'j}}{k_{..}} \right)^2} = \sqrt{\sum_{j=1}^J (f_{ij} - f_{i'j})^2} =; \forall i, i'$$

- De la misma manera, la distancia entre dos elementos columnas de una tabla F(I, J) es la siguiente:

$$d_{(j,j')} = \sqrt{\sum_{i=1}^I \left(\frac{k_{ij}}{k_{..}} - \frac{k_{ij'}}{k_{..}} \right)^2} = \sqrt{\sum_{i=1}^I (f_{ij} - f_{ij'})^2} =; \forall j, j'$$

DEFINICION DE LOS PERFILES DE LOS ELEMENTOS DE LA TABLA F

Sirve para corregir las distorsiones que pueda existir en la información de una tabla **F**, por ello debemos comparar los perfiles fila y columna.

Los perfiles de fila y de columna de la tabla **F** se obtienen dividiendo los valores de cada celdilla por el valor marginal correspondiente. Esto permite que las filas sean comparables y que las columnas también lo sean. Se puede analizar perfiles filas o perfiles columnas, ambos planteamientos producen los mismos resultados, filas y columnas realizan un papel simétrico en el análisis.

PERFIL FILA: Expresa la distribución de frecuencias condicionales de la subpoblación en la fila **i** para cada una de las características de la columna **j**, y lo podemos expresar de la siguiente manera:

$$F_{ij} = \left[F_{i1} = \frac{f_{i1}}{f_i}, F_{i2} = \frac{f_{i2}}{f_i}, \dots, F_{iJ} = \frac{f_{iJ}}{f_i} \right], i \in I; j \in J$$

$$\text{y} \quad \sum_{j=1}^J F_{ij} = \sum_{j=1}^J \frac{f_{ij}}{f_i} = \frac{1}{f_i} \sum_{j=1}^J f_{ij} = \frac{f_i}{f_i} = 1$$

CATEGORÍAS		VARIABLE Y							
		1	2	3	...	j	...	J	TOTAL
VARIABLE X	1	F ₁₁	F ₁₂	F ₁₃	...	F _{1j}	...	F _{1J}	1
	2	F ₂₁	F ₂₂	F ₂₃	...	F _{2j}	...	F _{2J}	1
	3	F ₃₁	F ₃₂	F ₃₃	...	F _{3j}	...	F _{3J}	1
	⋮	⋮	⋮	⋮	⋮	⋮
	i	F _{i1}	F _{i2}	F _{i3}	...	F _{ij}	...	F _{iJ}	1
	⋮	⋮	⋮	⋮	⋮	⋮
	I	F _{I1}	F _{I2}	F _{I3}	...	F _{Ij}	...	F _{IJ}	1

PERFIL COLUMNA: Expresa la distribución de frecuencias condicionales de la subpoblación en la columna **j** para cada una de las características de la fila **i**, y lo podemos expresar de la siguiente manera:

$$C_{ij} = \left[C_{1j} = \frac{f_{1j}}{f_{.j}}, C_{2j} = \frac{f_{2j}}{f_{.j}}, \dots, C_{Ij} = \frac{f_{Ij}}{f_{.j}} \right], \forall i \in I; \forall j \in J$$

$$y \quad \sum_{i=1}^I C_{ij} = \sum_{i=1}^I \frac{f_{ij}}{f_{.j}} = \frac{1}{f_{.j}} \sum_{i=1}^I f_{ij} = \frac{f_{.j}}{f_{.j}} = 1$$

CATEGORÍAS		VARIABLE Y						
		1	2	3	...	j	...	J
VARIABLE X	1	C ₁₁	C ₁₂	C ₁₃	...	C _{1j}	...	C _{1J}
	2	C ₂₁	C ₂₂	C ₂₃	...	C _{2j}	...	C _{2J}
	3	C ₃₁	C ₃₂	C ₃₃	...	C _{3j}	...	C _{3J}
	⋮	⋮	⋮	⋮	⋮
	i	C _{i1}	C _{i2}	C _{i3}	...	C _{ij}	...	C _{iJ}
	⋮	⋮	⋮	⋮	⋮
	I	C _{I1}	C _{I2}	C _{I3}	...	C _{Ij}	...	C _{IJ}
	TOTAL	1	1	1	...	1	...	1

REPRESENTACION DIGITAL DE LA INFORMACION DE LAS TABLAS PERFILES

- Evaluación de la distancia entre dos perfiles-fila, asociados a una tabla F (I x J):

$$d_{(i,i')} = \sqrt{\sum_{j=1}^J \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2}; \forall i, i'$$

- Evaluación de la distancia entre dos perfiles-columna, asociados a una tabla F (I x J):

$$d_{(j,j')} = \sqrt{\sum_{i=1}^I \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2}; \forall j, j'$$

El índice $d_{(i,i')}$ o $d_{(j,j')}$ evalúa la distancia que caracteriza la semejanza entre dos perfiles en el espacio euclidiano de representación de los mismos.

Para eliminar toda distorsión se debe ponderar cada término de la comparación de dos perfiles por la importancia relativa.

COMPARACIÓN PONDERADA DE PERFILES

Cuando aplicamos la distancia entre dos perfiles-filas, establecemos la suma de los cuadrados de las diferencias de frecuencias condicionales de cada una de las modalidades de la variable en columna. Llamamos “contribución del atributo j a la comparación de dos perfiles-filas” al valor de cada término de esa suma. De la misma manera, llamamos “contribución del atributo i a la comparación de dos perfiles columnas” al valor de cada término de la suma de los cuadrados de las diferencias de frecuencias condicionales de cada una de las modalidades de la variable en filas.

La importancia de cada elemento de la tabla F, expresada por su frecuencia marginal, la llamaremos en adelante el peso de cada elemento.

Para neutralizar la distorsión de la información de la tabla F que persiste

cuando comparamos los perfiles de la misma, debemos considerar la “contribución del atributo j (atributo i) a la distancia entre dos perfiles-filas (columnas)” ya no en valor absoluto, sino en términos relativos al peso del atributo j (atributo i). Procediendo de esta manera hacemos que una diferencias entre perfiles sea tanto mas “contributiva” a la relación que queremos establecer entre dos perfiles filas (columna).

Para ello es suficiente ponderar las contribuciones de cada atributo j (atributo i) a la comparación de dos filas (columnas) con el valor de la inversa del peso de la columna j (fila i).

EVALUACIÓN DE LA COMPARACIÓN DE PERFILES: DISTANCIA Chi²

Modificando la distancia definida anteriormente entre perfiles, definimos otra expresión de un índice de evaluación de similitudes entre perfiles de una tabla F que respeta la exigencia de comparación que hemos explicado.

Ante la desventaja de la prueba Chi-cuadrado se define la distancia Chi-cuadrado entre dos filas (o entre dos columnas) como una distancia euclidiana, al cuadrado, entre sus perfiles, ponderado con las inversas de las frecuencias marginales de las columnas. Esta ponderación evita que las columnas más frecuentes tengan más peso en la distancia.

➤ Distancia del chi² entre dos **perfiles ponderados fila**:

$$d_{(i,i')} = \sqrt{\sum_{j=1}^J \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2} = \sqrt{\sum_{j=1}^J \frac{1}{p_j} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2}, \forall i, i' \in I$$

➤ Distancia del chi² entre dos **perfiles ponderados columna**:

$$d_{(j,j')} = \sqrt{\sum_{i=1}^I \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{i'j'}}{f_{.j'}} \right)^2} = \sqrt{\sum_{i=1}^I \frac{1}{p_i} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{i'j'}}{f_{.j'}} \right)^2}, \forall j, j' \in J$$

Esta distancia es llamada “distancia del Chi²” o “distancia de Benzecri”. Es una distancia euclidiana.

Decimos entonces que la distancia del χ^2 es un índice de evaluación de la información que neutraliza todas las distorsiones en la representación de la información de la tabla de contingencia. La menor o mayor distancia entre los puntos indica el grado de similitud o disimilitud entre los perfiles.

El objetivo del análisis de correspondencia es reducir la dimensión de estos espacios para facilitar la interpretación del grado de similitud de las distintas categorías.

Importante: Se puede analizar los perfiles fila o los perfiles columna, ambos planteamientos producen los mismos resultados, filas y columnas juegan un papel simétrico en el análisis.

REPRESENTACIÓN DE LA INFORMACIÓN EN TERMINOS DE INERCIA DE UNA NUBE DE PUNTOS

Una tabla de contingencia puede ser analizada sin distorsiones si expresamos esa información por medio de:

- Dos espacios euclidianos de representación de los perfiles asociados a la tabla, dotado de sus respectivos pesos.
- Un índice de evaluación de la similitud entre perfiles ponderados asociados a la tabla F (o la tabla T), que llamamos distancia del χ^2 .

Entonces, puesto que la representación analógica de la información se presenta ahora bajo la forma de una cierta distribución de masas en dos espacios euclidianos, no podemos expresar en ello la forma de la información de la tabla F, o la tabla T, en términos de la posición relativa de punto-perfiles, puesto que el peso asociado a esos puntos modifica sus posiciones relativas.

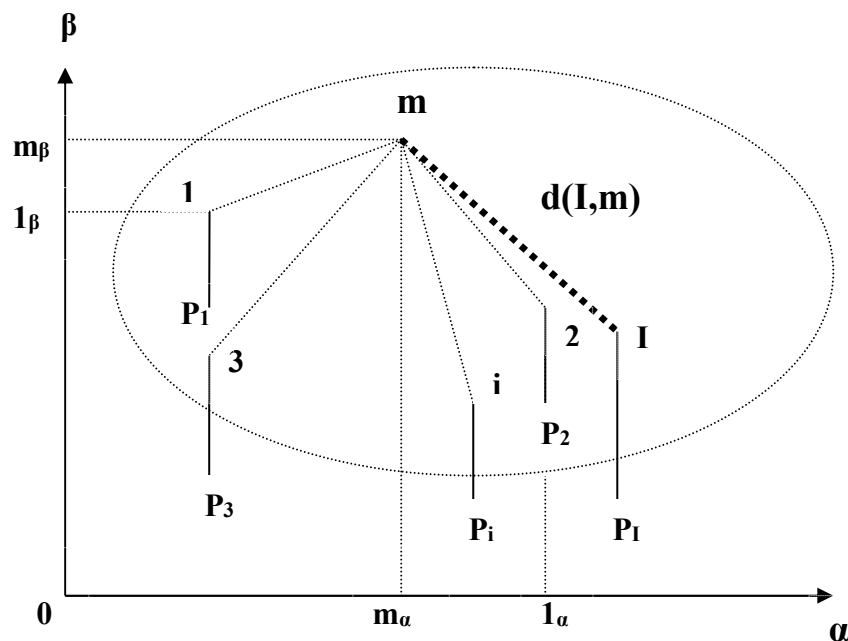
Evocando el concepto físico de inercia veremos que la forma de la información de la tabla F debe ser expresada ahora en términos de dispersión de esos puntos-masa con respecto a un punto de referencia.

a) INERCIA DE UNA NUBE DE PUNTOS CON REPECTO A UN PUNTO DE REFERENCIA

Definamos inercia de una nube puntos perfiles $N(I)$ dotados de peso con respecto a un punto cualquiera m , a la dispersión de esa nube de puntos con respecto al punto m tomado como referencia.

Sea un conjunto de puntos definido de la siguiente manera: $I = (1, 2, 3, \dots, I)$, dotados de pesos, $p_i = (p_1, p_2, p_3, \dots, p_i)$, que llamamos nube de puntos $N(I)$.

En esta figura, esos puntos están ubicados en el espacio euclidiano mediante sus coordenadas en los ejes α y β .



La dispersión de esa nube de puntos con respecto a un punto m cualquiera de ese espacio esta dada por la expresión siguiente:

$$I_m = \sum_{i=1}^I p_i d^2_{(I,m)}, \forall i \in I$$

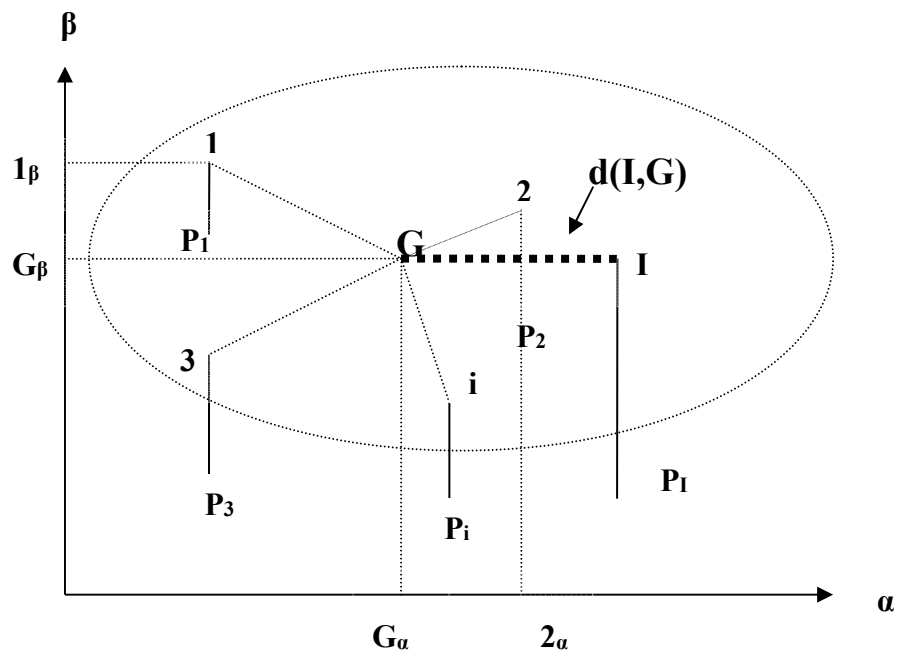
Expresamos entonces que la inercia (dispersión), con respecto a m , de una nube de puntos $N(I)$ es igual a la suma de los productos del cuadrado de la distancia a m de cada punto de la nube multiplicada por el peso asociado a cada punto.

La **contribución del punto i** a la inercia con respecto a m de la nube de puntos $N(I)$ es igual al término correspondiente del punto i en la suma que mide la dispersión de la nube $N(I)$ con respecto a m . Por lo tanto la contribución del punto i a la inercia con respecto a m de la nube de puntos es definida así:

$$\text{Contr}I_m(i) = p_i d^2_{(I,m)}, \forall i \in I$$

Consideremos ahora el mismo sistema de masas en un plano y veamos la dispersión de esos puntos con respecto a un punto particular del plano: El punto en torno al cual esas masas están en equilibrio. Decimos que esos puntos están en equilibrio en torno al punto G por que consideramos que la rotación del sistema de masas en torno a G es nula.

Como vemos en el siguiente grafico:



Los tres puntos ubicados a la derecha del punto G imponen una cierta cantidad de movimiento (en el sentido señalado por la flecha que parte del punto i) que se compensa, teniendo en cuenta los pesos relativos de esos puntos con el movimiento (en el sentido señalado por la flecha que parte del punto 3) impuesto por los dos puntos ubicados a la izquierda de G, tomando en cuenta sus pesos relativos.

Definiendo así la condición de equilibrio de ese sistema de masas, diremos que el punto G es el centro de gravedad del sistema, o bien el punto baricentro. Si G es el centro de gravedad del sistema, una fuerza impuesta en G (paralela a las fuerzas que representan los pesos de cada punto del espacio pero de sentido contrario) y de magnitud igual a la suma de los pesos de cada punto del espacio, mantiene el sistema de punto “suspendido” en el espacio, sin alterar la forma del sistema de masas.

Si el punto G es el baricentro del sistema de masas, las coordenadas de G sobre cada eje del referencial de ese espacio es igual a la media ponderada de las coordenadas de los puntos masa en el mismo punto referencial.

Llamando “Inercia” esa dispersión, así evaluada, de los puntos de $N(I)$ con respecto a un punto m cualquiera “evocamos” el concepto físico de “cantidad de movimiento” de un conjunto de masas en torno a un punto de referencia. Sin embargo no podemos considerar que la dispersión de puntos dotados de masas en los espacios de representación que hemos creado, cumplen con las mismas leyes, formuladas en física, que rigen los equilibrios de masas reales.

INERCIA DE LOS PUNTOS PERFILES PONDERADOS CON RESPECTO A LOS CENTROS DE GRAVEDAD

**INERCIA DE LOS PUNTOS PERFILES FILAS PONDERADOS CON
RESPECTO A LOS PUNTOS G_L Y G_C :**

$$I_{G_L}^{N(I)} = \sum_{i=1}^I \text{Contr}G_L(i) \quad \text{y} \quad I_{G_C}^{N(J)} = \sum_{j=1}^J \text{Contr}G_C(j)$$

En los dos espacios de representación de la información de las tablas de Perfiles Ponderados, la inercia de la nube de puntos $N(I)$ con respecto a G_L es la misma que la inercia de la nube de puntos $N(J)$ con respecto a G_C .

DESvíO A LA SITUACION DE INDEPENDENCIA: ASOCIACION DE DOS VARIABLES CUALITATIVAS

Para medir la asociación entre dos variables cualitativas hay que evaluar la importancia con la cual lo observado se aleja de la situación de independencia.

Tabla de independencia o tabla teórica:

Sea $f_{ij}^* = f_{i.} \cdot f_{.j}$ el término general

Entonces:

$$\sum_{i=1}^I f_{i.}^* = \sum_{i=1}^I f_{i.} = 1 \quad \text{y} \quad \sum_{j=1}^J f_{.j}^* = \sum_{j=1}^J f_{.j} = 1$$

Las distribuciones de la tabla teórica son proporcionales a sus distribuciones marginales:

$$\sum_{j=1}^J f_{ij}^* = f_{i.} \sum_{j=1}^J f_{.j} = f_{i.}; \forall i \in I$$

$$\sum_{i=1}^I f_{ij}^* = f_{.j} \sum_{i=1}^I f_{i.} = f_{.j}; \forall j \in J$$

➤ Tabla de desvíos a la situación de independencia:

$$\text{Término general: } e_{ij} = f_{ij} - f_{ij}^*; \forall i \in I; \forall j \in J$$

➤ Tabla de desvíos ponderados:

$$\text{Término general: } e_{ij}^* = \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}; \forall i \in I; \forall j \in J$$

➤ Interés de esta tabla:

La suma de los valores de esta tabla es igual al coeficiente del φ^2

$$\varphi^2 = \sum_{i=1}^I \sum_{j=1}^J e_{ij}^* = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*}; \forall i \in I; \forall j \in J$$

Los coeficientes del φ^2 puede ser expresado con los elementos de la Tabla F(I,J).

$$\varphi^2 = \sum_{i=1}^I \sum_{j=1}^J e_{ij}^* = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}}; \forall i \in I; \forall j \in J$$

La inercia de las nubes N(I) y N(J), con respecto al punto G correspondiente, mide el grado de asociación de las variables cualitativas de la Tabla F(I,J).

Demostración:

$$\begin{aligned} I_{G_j}^{N(J)} &= \sum_{j=1}^J p_j d^2(j, G_j) \\ &= \sum_{j=1}^J \frac{1}{f_{.j}} \left(\sum_{i=1}^I \frac{1}{f_{i.}} (f_{ij} - f_{i.}f_{.j})^2 \right) \\ &= \sum_{j=1}^J \sum_{i=1}^I \frac{(f_{ij} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}} \\ &= \varphi^2 \end{aligned}$$

LOS PUNTOS PERFILES MARGINALES G DE ESOS ESPACIOS SON LOS CENTROS DE GRAVEDAD DE LOS SISTEMAS DE MASA

Hemos visto que la masa asociada al punto perfil marginal en cada uno de esos espacios de representación es igual a la suma de las masas asociadas a cada punto perfil de esos espacios.

Cálculo de la media ponderada de las coordenadas de los puntos perfiles sobre cada eje del espacio

- Considerando los valores absolutos de las masas:

$$p_{G_L}^* = k_1 + k_2 + k_3 + \dots + k_i = K \quad \text{y} \quad p_{G_C}^* = k_1 + k_2 + k_3 + \dots + k_j = K$$

- Las coordenadas de los puntos perfiles columnas sobre el eje i:

$$\left[\frac{f_{i1}}{f_{i.}\sqrt{f_{.1}}}, \frac{f_{i2}}{f_{i.}\sqrt{f_{.2}}}, \frac{f_{i3}}{f_{i.}\sqrt{f_{.3}}}, \dots, \frac{f_{iJ}}{f_{i.}\sqrt{f_{.J}}} \right]$$

- De modo que la media ponderada por el peso de esos perfiles de esas coordenadas es la siguiente:

$$\begin{aligned} \bar{X} &= \frac{1}{K_{..}} \sum_{j=1}^J k_{.j} \left(\frac{f_{ij}}{f_{.j}\sqrt{f_{i.}}} \right) = \frac{1}{K_{..}} \sum_{j=1}^J k_{.j} \sum_{i=1}^I \frac{f_{ij}}{f_{.j}\sqrt{f_{i.}}} \\ &= \frac{1}{\sqrt{f_{i.}}} \sum_{j=1}^J \frac{f_{ij}}{f_{.j}} = \frac{1}{\sqrt{f_{i.}}} \frac{\sum_{j=1}^J f_{ij}}{\sum_{j=1}^J f_{.j}} = \frac{1}{\sqrt{f_{i.}}} \frac{f_{i.}}{1} = \sqrt{f_{i.}} \end{aligned}$$

Vemos que la media ponderada de las coordenadas de los **puntos-perfiles columnas** sobre el eje i es igual a la coordenada del perfil marginal en columna sobre ese mismo eje.

- Las coordenadas de los puntos perfiles fila sobre el eje j:

$$\left[\frac{f_{1j}}{f_{.j}\sqrt{f_{.1}}}, \frac{f_{2j}}{f_{.j}\sqrt{f_{.2}}}, \frac{f_{3j}}{f_{.j}\sqrt{f_{.3}}}, \dots, \frac{f_{Ij}}{f_{.j}\sqrt{f_{.I}}} \right]$$

- De modo que la media ponderada por el peso de esos perfiles de esas coordenadas es la siguiente:

$$\begin{aligned}\bar{X} &= \frac{1}{K_{..}} \sum_{i=1}^I k_i \left(\frac{f_{ij}}{f_i \sqrt{f_{.j}}} \right) = \frac{1}{K_{..}} \sum_{i=1}^I k_i \sum_{j=1}^J \frac{f_{ij}}{f_i \sqrt{f_{.j}}} \\ &= \frac{1}{\sqrt{f_{.j}}} \sum_{i=1}^I \frac{f_{ij}}{f_i} = \frac{1}{\sqrt{f_{.j}}} \frac{\sum_{i=1}^I f_{ij}}{\sum_{i=1}^I f_i} = \frac{f_{.j}}{1} \frac{1}{\sqrt{f_{.j}}} = \sqrt{f_{.j}}\end{aligned}$$

REPRESENTACIÓN FACTORIAL DE LA INFORMACIÓN DE UNA TABLA DE CONTINGENCIA

Veamos como se crean los ejes factoriales que permiten la representación factorial de la información de una tabla de contingencia.

Apoyándonos en las principales propiedades de los factores de una nube de puntos-perfiles, podremos presentar luego el conjunto de índices estadísticos que completan la representación factorial de la información de una tabla de contingencia. Índices que constituyen la componente digital complementaria de ese modo analógico de presentar dicha información.

De ese modo de representación de la información aportada por una tabla de contingencia admite la representación conjunta de la inercia global de las nubes de puntos-perfiles $N(I)$ y $N(J)$. Esta propiedad permite ver en los gráficos factoriales las correspondencias entre las características observadas puestas en evidencia por el análisis.

También veremos la importancia de elementos suplementarios en el análisis de la información de una tabla de contingencia y la manera de representar esos elementos en el espacio factorial de representación de información de una tabla de contingencia.

PROPIEDADES DE LOS FACTORES DE UNA NUBE DE PUNTOS- PERFILES (COORDENADAS DE LOS PUNTOS PERFILES)

Las coordenadas de los puntos perfiles (en fila y en columna) en el nuevo referencial se obtienen por simple proyección ortogonal de los puntos-perfiles (centrados en G) sobre las rectas del nuevo referencial de cada espacio. Esas coordenadas están definidas de la siguiente manera:

$$F_{\alpha}(i) = \sum_{j=1}^J Z_{ij} U_{\alpha j} \quad \mathbf{y} \quad G_{\alpha}(j) = \sum_{i=1}^I Q_{ij} W_{\alpha i}$$

Decimos entonces que: $F_1(i)$ es la coordenada de la i -ésima fila de la tabla T sobre el primer eje factorial; $F_2(i)$ es la coordenada de la i -ésima fila de la tabla T sobre el segundo eje factorial; etc... Por otra parte, $G_1(j)$ es la coordenada de la j -ésima columna de la tabla T sobre el primer eje factorial; $G_2(j)$ es la coordenada de la j -ésima columna de la tabla T sobre el segundo eje factorial; y así sucesivamente.

a) COORDENADAS DEL CENTRO DE GRAVEDAD

Los centros de gravedad de las nubes de puntos $N(I)$ y $N(J)$, respectivamente, están ubicados por construcción, en el origen de esos nuevos sistemas de ejes de referencia que son los factores de esas nubes de puntos-perfiles. En consecuencia sus coordenadas de sus espacios factoriales de representación son nulas, es decir:

$$F_{\alpha}(G_L) = 0, \forall \alpha = 1, 2, \dots, p \quad \mathbf{y} \quad G_{\alpha}(G_C) = 0, \forall \alpha = 1, 2, \dots, p$$

Sabemos que las **coordenadas del centro de gravedad** en los espacios factoriales son todas nulas. Lo cual podemos expresar de la siguiente manera:

$$\forall i \in I, \mathbf{y} \forall \alpha, \sum_{i=1}^I f_i F_{\alpha}(G_L) = 0 = \sum_{j=1}^J f_j G_{\alpha}(G_C), \forall j \in J, \mathbf{y} \forall \alpha$$

De modo entonces que los “factores” de las nubes de puntos-perfiles constituyen una serie de p variables (cuantitativas, a valores de R) de media nula, es decir una serie de p variables centradas.

b) VARIANZA DE LOS FACTORES DE UNA NUBE DE PUNTOS-PERFILES

Considerando las variables cuantitativas que llamamos “factores” de una nube de puntos-perfiles, podemos calcular la varianza de las mismas.

Sabemos que: $\bar{X}_\alpha = 0$, de modo que: $\forall \alpha = 1, 2, 3, \dots, p$

$$S_\alpha^2 = \sum_{i=1}^I p_i (x_{\alpha i} - \bar{x}_\alpha)^2 = \sum_{i=1}^I p_i (F_\alpha(i) - 0)^2 = \sum_{i=1}^I p_i (F_\alpha(i))^2$$

En esa expresión: $F_\alpha(i)$ es la coordenada del punto i sobre el eje α , es decir la distancia del origen del eje (que corresponde al centro de gravedad G_L) de la proyección ortogonal del punto i sobre el eje α ; y p_i es peso asociado al punto perfil i . De modo que la varianza del factor α es igual a la inercia de la nube de puntos $N(I)$ a lo largo de ese eje.

Pero se puede verificar que la inercia a lo largo de cada eje factorial de una nube de puntos-perfiles y esta dada por el valor propio asociada a cada eje. Podemos entonces establecer la relación siguiente:

$$S_\alpha^2 = \lambda_\alpha = \sum_{i=1}^I p_i (F_\alpha(i))^2, \forall \alpha = 1, 2, \dots, p$$

c) VALORES PROPIOS PERTINENTES DE UN ANÁLISIS

Si un valor propio es muy pequeño (próximo a cero), concluimos que la parte de la asociación observada entre las variables de la tabla de contingencia que es asumida por el eje correspondiente es ínfima. El complemento de información que aporta ese eje a la información de representación global de la tabla es insignificante o despreciable.

En la práctica del empleo de este método de análisis, sobre tablas de contingencia de grandes dimensiones, el valor propio límite puede ser fijado calculando el valor propio medio.

$$\text{Es decir: } \bar{\lambda} = \frac{I_{G_F}(N(I))}{p} \quad \text{o bien} \quad \bar{\lambda} = \frac{I_{G_C}(N(J))}{p}$$

No debemos utilizar esta regla de manera absoluta cuando procedemos a la interpretación de los resultados de un análisis. Si se analiza, por ejemplo, una tabla de contingencia de grandes dimensiones cuya inercia global no sea muy elevada, el valor propio medio será ya muy próximo a cero. Los valores propios próximos precedentes pueden también ser casi nulo, es decir que aportan una información complementaria que es ya despreciable.

La traza de la matriz inercia de una nube puntos-perfiles es la medida de la inercia global de esa nube de puntos con respecto al centro de gravedad de la misma. La inercia global de esa nube de puntos con respecto al centro de gravedad, es igual al índice φ^2 de asociación de las variables que componen la tabla de contingencia.

De modo que:
$$I_G = \sum_{\alpha=1}^p \lambda_{\alpha} = \varphi^2$$

d) TAZA DE INERCIA DE UN FACTOR

Sabiendo que la inercia I_G de la nube de puntos-perfiles $N(I)$ (o $N(J)$) es igual a la suma de los p valores propios λ_{α} no nulos, podemos expresar el aporte de cada eje a la inercia total de la nube por medio del índice τ_{α} , que expresa la tasa de inercia del factor y que definimos a continuación:

$$\tau_{\alpha} = \frac{\lambda_{\alpha}}{\sum_{\alpha=1}^p \lambda_{\alpha}} * 100 = \frac{\lambda_{\alpha}}{I_G} * 100$$

De modo que:

$$\sum_{\alpha=1}^p \tau_{\alpha} = 100 \%$$

e) TAZA DE INERCIA DE UN PLANO FACTORIAL

Si nos interesamos en la inercia de una nube de puntos perfiles proyectada sobre un plano definido por dos ejes factoriales α_i y α_{i+k} , sabemos que la inercia de la nube sobre ese plano esta medida por: λ_{α_i} y $\lambda_{\alpha_{i+k}}$ (puesto que el eje α_i es ortogonal al eje α_{i+k}), de modo que la taza de la inercia asociada al plano $(i, i+k)$ es defina así:

$$\tau_{plano (i,i+k)} = \frac{(\lambda_{\alpha_i}, \lambda_{\alpha_{i+k}})}{\sum_{\alpha=1}^p \lambda_{\alpha}} * 100$$

La taza de inercia expresa la parte de inercia global que es representado a lo largo de un eje factorial o sobre un plano factorial. Pero el índice ϕ^2 de asociación entre las variables que sirve de referencia al índice τ_{α} , no constituye una medida significativa de la información que aporta la tabla de contingencia.

COMPONENTES DIGITALES DEL MENSAJE

a. CONTRIBUCION DE LOS PUNTOS PERFILES A LA INERCIA A LO LARGO DE UN EJE FACTORIAL

La contribución de un punto-perfil a la inercia de un factor es un índice que juega un rol determinante cuando se trata de interpretar los resultados del análisis de la información aportada por una tabla de contingencia.

La contribución de un punto-perfil conjuga el peso y la coordenada ligados a cada punto. Esto permite evitar los errores de interpretación que produciría una lectura "inocente" de los gráficos de los ejes factoriales que tome en cuenta solamente las coordenadas de los puntos-perfiles, olvidando que cada punto tiene asociado un peso (no representado) y que el objetivo del grafico del

eje factorial es rendir cuenta de una parte de la inercia global de la nube de puntos-perfiles.

La contribución del punto i a la inercia del eje α es la siguiente:

$$CTR_{\alpha}(i) = \frac{p_i (F_{\alpha}(i))^2}{\lambda_{\alpha}} * 100, \forall i \in N(I)$$

Contribución del punto j al eje α :

$$CTR_{\alpha}(j) = \frac{p_j (F_{\alpha}(j))^2}{\lambda_{\alpha}} * 100, \forall j \in N(J)$$

b. CALIDAD DE REPRESENTACIÓN DE LOS PUNTOS PERFILES

Podemos expresar por medio de un **índice de calidad de representación** del punto en cada eje, la parte de la distancia de un punto i (o del punto j) al centro de gravedad, que es representada por la coordenada del punto i (o del punto j) sobre cada eje factorial.

Calidad de representación del punto i sobre el eje α :

$$COR_{\alpha}(i) = \frac{(F_{\alpha}(i))^2}{d_{(i,G)}^2}, \forall \alpha = 1, 2, 3, \dots, p$$

donde:
$$\sum_{\alpha=1}^p COR_{\alpha}(i) = 1, \forall i \in N(I)$$

Calidad de representación del punto j sobre el eje α :

$$COR_{\alpha}(j) = \frac{(G_{\alpha}(j))^2}{d_{(j,G)}^2}, \forall \alpha = 1, 2, 3, \dots, p$$

donde:
$$\sum_{\alpha=1}^p COR_{\alpha}(j) = 1, \forall j \in N(J)$$

Si un punto perfil i presentara un $COR_{\alpha}(i)=100\%$ para un eje α dado, esto significa que el eje α representa perfectamente al punto perfil i puesto que $F_{\alpha}(i) = d_{(i,0)}$. Decimos entonces que el punto perfil i está completamente explicado por el eje α . Considerando solamente ese eje. Considerando solamente ese eje, podemos explicar completamente la posición de ese punto i con respecto a todos los otros puntos perfiles de la nube $N(I)$.

En cambio, si un punto perfil i presentara un $COR_{\alpha}(i) = 0\%$ para un eje α , esto significa que $F_{\alpha}(i) = 0$, es decir que ese punto está situado en un subespacio ortogonal al eje α . Ese eje no está asociado a ese punto.

De modo entonces que el coeficiente $COR_{\alpha}(i)$ o $COR_{\alpha}(j)$ se interpreta como un coeficiente de correlación múltiple entre cada variable de la tabla T y el conjunto de p variables factoriales construidas por el Análisis Factorial de Correspondencias de la tabla T .

OBTENCIÓN DE LOS EJES FACTORIALES:

El análisis se puede hacer bien en la nube $N(I)$ o en la nube $N(J)$, los resultados de ambos análisis están relacionados.

Utilizando la nube $N(I)$, el objetivo es obtener una representación simplificada de los puntos fila cuyas coordenadas son f_{ij} / f_i , $j = 1, \dots, p$.

Estos puntos están afectados de un peso o masa f_i y la distancia entre ellos se mide con la distancia χ^2 .

$$d^2(i, i') = \sum_{j=1}^J \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_{i'}} \right)^2 = \sum_{j=1}^J \left(\frac{f_{ij}}{f_i \sqrt{f_j}} - \frac{f_{i'j}}{f_{i'} \sqrt{f_j}} \right)^2$$

Buscamos el centro de gravedad de la nube de puntos $\frac{f_{ij}}{f_i \sqrt{f_j}}$, teniendo en cuenta que cada punto tiene un peso f_i .

$$g_j = \sum_{i=1}^I \frac{f_{ij}}{f_i \sqrt{f_j}} f_i = \sum_{i=1}^I \frac{f_{ij}}{\sqrt{f_j}} = \frac{f_j}{\sqrt{f_j}} = \sqrt{f_j}$$

$$\Rightarrow G_L = (g_1, \dots, g_J)$$

Luego centramos la nube, trasladando el origen al centro de gravedad. Las nuevas coordenadas de los puntos i son:

$$\frac{f_{ij}}{f_i \sqrt{f_j}} - \sqrt{f_j}$$

La proyección de un punto sobre un nuevo eje de vector unitario u_1 viene dada por

$$F_1(i) = \sum_{j=1}^J \left(\frac{f_{ij}}{f_i \sqrt{f_j}} - \sqrt{f_j} \right) u_{1j}$$

Necesitamos hallar el vector u_1 que maximice la inercia de la nube proyectada, es decir

$$\text{máx} \sum_{i=1}^I f_i F_1^2(i)$$

Lo cual equivale a diagonalizar (buscar vectores propios) de la matriz Z , de término general

$$z_{ij'} = \sum_{i=1}^I \left(\frac{f_{ij} - f_i f_j}{\sqrt{f_i} \sqrt{f_j}} \right) \left(\frac{f_{ij'} - f_i f_{j'}}{\sqrt{f_i} \sqrt{f_{j'}}} \right)$$

Esta matriz se puede expresar como $Z = X'X$ siendo X la matriz de término general

$$x_{ij} = \frac{f_{ij} - f_i f_j}{\sqrt{f_i} \sqrt{f_j}}$$

Se puede demostrar que todos los vectores propios de $Z = X'X$, para todo $\alpha \neq J$, son también vectores propios de $S = W'W$, siendo

$$w_{ij} = \frac{f_{ij}}{\sqrt{f_i} \sqrt{f_j}}$$

El vector u_J es también vector propio de S pero asociado al valor propio 1. En consecuencia el análisis puede realizarse sobre la tabla W no centrada con la condición de despreciar este vector propio.

Las coordenadas de los puntos i sobre el eje α están dadas por

$$F_\alpha(i) = \sum_{j=1}^J \frac{f_{ij}}{f_i \sqrt{f_j}} u_{\alpha j}$$

De igual manera se trabaja en la nube $N(J)$, donde la matriz a diagonalizar es XX' . Cuyos vectores propios coinciden con los de WW' . Los vectores propios obtenidos serán v_i . Y la proyección de los puntos j sobre el eje α será:

$$G_{\alpha}(j) = \sum_{i=1}^I \frac{f_{ij}}{f_{j\cdot} \sqrt{\lambda_i}} v_{\alpha i}$$

INTERPRETACIÓN DE LOS RESULTADOS:

Si se consideran los dos primeros ejes tenemos:

1. Si la filas (columnas) tienen una estructura semejante, su situación será próxima sobre el plano, lo inverso no siempre es cierto depende de la calidad de representación.
2. La situación cercana de un punto fila i y un punto columna j sólo se puede interpretar si están alejados del origen, en la periferia de la nube
3. Cuando una línea tiene un perfil próximo al perfil medio, tienen un comportamiento medio, se encontrará próxima al origen.

AYUDAS A LA INTERPRETACIÓN:

Para interpretar la imagen hay que conocer la significación de los ejes sobre los cuales hemos proyectado la nube y estudiar la forma de esta. Para ello es necesario calcular unos coeficientes o ayudas a la interpretación.

Para cada fila y columna se calculan:

- Contribución Absoluta $CTA_{\alpha}(i)$: expresa la participación que tiene el elemento i en la inercia explicada por el factor α .

$$CTA_{\alpha}(i) = \frac{f_i F_{\alpha}^2(i)}{\lambda_{\alpha}}$$

- Contribución Relativa $CTR_{\alpha}(i)$: Recoge la participación del factor α en la explicación del elemento i . Mide la calidad de representación de i sobre el eje α .

$$CTR_{\alpha}(i) = \frac{F_{\alpha}^2(i)}{d^2(i, G)} ; \sum_{\alpha} CTR_{\alpha}(i) = 1$$

INTERPRETACIÓN DE LOS EJES:

- Primero se buscan aquellos puntos i (j) de mayor CTA.
- Dentro de estos se separan los puntos que se proyectan del lado positivo de los que intervienen en el lado negativo, que estarán en oposición.
- Se estudia la calidad de representación CTR de estos puntos. Si un punto tiene un CTR pequeño es de suponer que tenga un papel importante sobre otro eje, y para el estudio sería necesario considerar el conjunto de los ejes.
- Se buscan aquellos puntos i (j) que si bien no contribuyen a la formación del factor, sí se encuentran bien representados (CTR alto). Estos puntos son ilustrativos en la significación del eje.

REPRESENTACIÓN SIMULTÁNEA.

Las dos nubes de puntos se pueden representar en el mismo espacio, porque existe una relación entre los dos espacios, ya que

$$v_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} X u_{\alpha} \quad y \quad u_{\alpha} = \frac{1}{\sqrt{\lambda_{\alpha}}} X' v_{\alpha}$$

De donde se obtiene

$$v_{\alpha i} = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{j=1}^J \frac{f_{ij}}{\sqrt{f_i} \sqrt{f_j}} u_{\alpha j} = \frac{1}{\sqrt{\lambda_{\alpha}}} F_{\alpha}(i) \sqrt{f_i}$$
$$\Rightarrow F_{\alpha}(i) = \frac{\sqrt{\lambda_{\alpha}} v_{\alpha i}}{\sqrt{f_i}}$$

De la misma manera se obtiene

$$G_{\alpha}(j) = \frac{\sqrt{\lambda_{\alpha}} u_{\alpha j}}{\sqrt{f_j}}$$

De donde se obtienen las siguientes relaciones:

$$G_{\alpha}(j) = \sum_{i=1}^I \frac{f_{ij}}{f_j \sqrt{f_i}} v_{\alpha i} = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^I \frac{f_{ij}}{f_j \sqrt{f_i}} F_{\alpha}(i) \sqrt{f_i}$$

$$G_{\alpha}(j) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^I \frac{f_{ij}}{f_j} F_{\alpha}(i)$$

De igual manera:

$$F_{\alpha}(i) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{j=1}^J \frac{f_{ij}}{f_{i.}} G_{\alpha}(j)$$

La proyección del punto i (j) sobre el eje α , $F_{\alpha}(i)$, $G_{\alpha}(j)$, es el baricentro de las proyecciones de los puntos j (i) sobre el mismo eje., excepto por el factor de dilatación $1/\sqrt{\lambda_{\alpha}}$. Luego hay que multiplicar los coeficientes por este factor para ubicar ambas nubes en el mismo espacio.

El SPSS para esto multiplica por el factor $1/\sqrt{\lambda_{\alpha}}$.

ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

Este análisis es una generalización del análisis de correspondencias simples. Analiza las asociaciones entre más de dos variables categóricas.

El análisis Factorial de Correspondencias Múltiples es un instrumento de investigación interesante cuando se trata de analizar la información aportada por tablas de contingencia de grandes dimensiones. Pero pocas son las áreas de investigación en ciencias humanas en las cuales se hace uso de ese tipo de tablas.

En efecto muy a menudo los investigadores proponen estrategias de observación de una población dada (conjuntos de unidades de observación), a fin de relevar sobre ese conjunto numerosas características, cada una de ellas compuesta de un pequeño número de modalidades. Por medio "Encuestas", se elaboran tablas que resumen las p características observadas sobre n unidades de observación.

El análisis Factorial de Correspondencias Múltiples es un instrumento particularmente adaptado al tratamiento estadístico de los datos producidos por esos tipos de observaciones. Este método de análisis estadístico responde a la doble exigencia de objetividad en ese proceso de reducción y de exploración de lo observado y de tratamiento de la información con el nivel de síntesis de lo adecuado al cuadro conceptual utilizado.

El Análisis Factorial de Correspondencias Múltiples responde a esa exigencia metodológica de analizar la información aportada por la observación de diferentes dimensiones cualitativas u ordinales de una población, respetando el nivel de síntesis que impone el cuadro conceptual con el cual el investigador concibe su problemática de estudio y diagrama el dispositivo de observación.

Esta adecuación metodológica del Análisis Factorial de Correspondencias Múltiples explica su rápida difusión en diferentes campos de investigación en ciencias humanas, como también el hecho que el método del Análisis Factorial de Correspondencias Binarias es empleado solamente en ciertos sectores de investigación que elaboran regularmente tablas de contingencia de grandes dimensiones.

TABLAS DE CONTINGENCIA MÚLTIPLE DE ESTADÍSTICAS DE RESUMEN

a) TABLA DE CODIGOS CONDENSADOS:

Presentamos aquí el modelo general de una tabla de códigos condensados y las principales características de las mismas:

TABLA DE CODIGOS CONDENSADOS TCC (n x p)

INDIVIDUOS	CODIGO DE LA CARACTERISTICA					
	1°	2°	...	j-ésima	...	p-ésima
1	1	2	...	2	...	4
2	2	1	...	1	...	3
⋮	⋮	⋮	⋮
i	k_{i1}	k_{i2}	...	k_{ij}	...	k_{ip}
⋮	⋮	⋮	⋮
n	k_{n1}	k_{n2}	...	k_{nj}	...	k_{np}

- Cada fila contiene todos los códigos correspondientes a las modalidades atribuidas a un individuo, para cada una de las características observadas.
- En la intersección de la i-ésima fila y de la j-ésima columna figura el valor k_{ij} = código numérico que fue dado a la modalidad “atribuida” al i-ésimo individuo a propósito de la j-ésima característica observada.
- Tratándose de una tabla de códigos, la tabla TCC (n x p) no posee propiedades numéricas

b) TABLA LOGICAS:

A partir de una Tabla de Códigos Condensados se puede generar una tabla numérica capaz de resumir la misma información.

Considerando la primera columna de la tabla TCC (nxp) tenemos:

Tabla TCC (nx1)		Tabla TL (nx3)				
INDIVIDUO	CÓDIGO DE LA PRIMERA CARACTERISTICA OBSERVADA	INDIVIDUO	CODIGO DE LA CARACTERISTICA			Suma Total
			Código: 1	Código: 2	Código: 3	
1	1	1	1	0	0	1
2	2	2	0	1	0	1
⋮	⋮	⋮	⋮	⋮	...	⋮
i	k_{i1}	i	X_{i1}	X_{i2}	X_{i3}	1
⋮	⋮	⋮	⋮	⋮	...	⋮
n	K_{n1}	n	0	0	X_{n3}	K_{np}
		Suma Total	$n.1$	$n.2$	$n.3$	n

- A la primera columna de la tabla TCC(nx p) asociamos una tabla TL(nx3), porque esa columna de TCC (nx p) comporta tres modalidades mutuamente exclusivas.
- La tabla TL (nx3) contiene las tres “variables indicadoras” de las modalidades de la 1° característica observada (variables de presencia-ausencia).

Repitiendo esa operación para cada columna de TCC (nx p), podemos asociar a ella una **tabla TL (nx k)**.

TABLA LOGICA O TABLA DISYUNTIVA COMPLETA CORRESPONDIENTE A LA TABLA TCC (nxp)

INDIVIDUO	PRIMERA CARACTERISTICA Modalidades			(...)	j-ÉSIMA CARACTERISTICA Modalidades			(...)	p-ÉSIMA CARACTERISTICA Modalidades			MARGEN
	1	(...)	k ₁		k _{j-1+1}	(...)	k _{j-1+k_j}		k _{p-1+1}	(...)	k	
1	0	1	0	...	1	0	0	...	1	0	0	p
2	1	0	0	...	0	0	1	...	0	0	1	p
⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮
i	0	1	0	...	0	X _{ij}	0	...	1	0	0	p
⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮
n	1	0	0	...	1	X _{nj}	0	...	0	1	0	p
MARGEN	n ₁	...	n _{k₁}	(...)	n _{k_{j-1+1}}	...	n _{k_{j-1+k_j}}	(...)	n _{k_{p-1+1}}	...	n _k	np

$$K = \sum_{m=1}^P k_m : \text{total de modalidad, de las } p \text{ características observadas.}$$

La suma en fila de la tabla es igual al margen constante, igual a p.

La margen inferior de la tabla TL (nxK) es igual a la distribución de frecuencias brutas de todas las características observadas.

- La tabla TL (nxK) resume lo observado en los mismos términos que la tabla de códigos condensados correspondientes.
- La tabla lógica presenta propiedades numéricas.
- La tabla TL (nxK) es igual a la tabla de correspondencias de tipo particular.

ANÁLISIS FACTORIAL DE CORRESPONDENCIAS DE UNA TABLA LÓGICA TL (nxK)

a) OBJETIVOS DEL ANÁLISIS

- Facilitar la construcción de esas topologías de individuos permitiendo la comparación de todas las unidades de observación a través de todas las modalidades de las características observadas.
- Estudiar la relación existente entre las características observadas.
- Resumir el conjunto de características observadas en un pequeño número de variables cuantitativas relacionadas con el conjunto de variables cualitativas estudiadas.
- Permitir la comparación de modalidades de las características observadas.

b) DISTANCIA ENTRE INDIVIDUOS (FILAS) DE UNA TABLA LÓGICA

- j-ésima coordenada del i-ésimo individuo:

$$f_{ij} = \frac{x_{ij}}{np} \quad ; \quad f_{i.} = \frac{1}{n} \quad ;$$

$$f_{.j} = \frac{\frac{n_j}{np} \frac{x_{ij}}{np}}{\frac{1}{n} \sqrt{\frac{n_j}{np}}} = \frac{x_{ij}}{p \sqrt{\frac{n_j}{np}}}$$

- j-ésima coordenada del i'-ésimo individuo:

$$f_{i'j} = \frac{x_{i'j}}{np} \quad ; \quad f_{i'.} = \frac{1}{n} \quad ; \quad f_{.j} = \frac{\frac{n_j}{np} \frac{x_{i'j}}{np}}{\frac{1}{n} \sqrt{\frac{n_j}{np}}} = \frac{x_{i'j}}{p \sqrt{\frac{n_j}{np}}}$$

➤ Distancia del χ^2 entre los puntos-individuos i e i' :

$$d_{(i,i')}^2 = \sum_{j=1}^K \left(\frac{x_{ij}}{p \sqrt{\frac{n_j}{np}}} - \frac{x_{i'j}}{p \sqrt{\frac{n_j}{np}}} \right)^2 = \sum_{j=1}^K \frac{np}{n_j} \left(\frac{x_{ij}}{p} - \frac{x_{i'j}}{p} \right)^2$$

$$d_{(i,i')}^2 = \frac{1}{p} \sum_{j=1}^K \frac{n}{n_j} (x_{ij} - x_{i'j})^2$$

Cada término de esa sumatoria $(x_{ij} - x_{i'j})^2$ solamente toma el valor 0 o 1.

- Este factor vale 1 solamente si los individuos considerados no presentan simultáneamente la j -ésima modalidad.
- Vale 0 en el caso contrario (presencia simultánea la j -ésima modalidad).

La distancia entre los mismos crece a medida que aumentan las diferencias de modalidades presentadas por los individuos i o i' .

Cada modalidad interviene en el cálculo de la distancia entre dos individuos con el factor n/n_k , es decir la inversa del peso de la modalidad. La distancia entre los individuos será más grande si presentan el mismo número de divergencias, a propósito de modalidades muy frecuentes (de gran peso).

La distancia χ^2 entre individuos de una tabla lógica respeta el criterio de comparación de individuos que hemos adoptado.

c) DISTANCIA ENTRE MODALIDADES (COLUMNAS) DE UNA TABLA LOGICA

- Cada columna de una tabla lógica TL (nxK) es una variable que indica la modalidad de una de las características observadas.
- Coordenadas de modalidades j y k de la tabla lógica TL (nxK):

$$\text{Modalidad j: } \left[\frac{\frac{x_{1j}}{np}}{\frac{n_j}{np} \sqrt{\frac{p}{np}}}; \frac{\frac{x_{2j}}{np}}{\frac{n_j}{np} \sqrt{\frac{p}{np}}}; \dots; \frac{\frac{x_{ij}}{np}}{\frac{n_j}{np} \sqrt{\frac{p}{np}}}; \dots; \frac{\frac{x_{nj}}{np}}{\frac{n_j}{np} \sqrt{\frac{p}{np}}} \right]$$

$$\text{Modalidad k: } \left[\frac{\frac{x_{1k}}{np}}{\frac{n_k}{np} \sqrt{\frac{p}{np}}}; \frac{\frac{x_{2k}}{np}}{\frac{n_k}{np} \sqrt{\frac{p}{np}}}; \dots; \frac{\frac{x_{ik}}{np}}{\frac{n_k}{np} \sqrt{\frac{p}{np}}}; \dots; \frac{\frac{x_{nk}}{np}}{\frac{n_k}{np} \sqrt{\frac{p}{np}}} \right]$$

- La distancia entre esas dos modalidades :

$$d_{(j,k)}^2 = \sum_{i=1}^n \left(\frac{\frac{x_{ij}}{np}}{\frac{n_j}{np} \sqrt{\frac{1}{n}}} - \frac{\frac{x_{ik}}{np}}{\frac{n_k}{np} \sqrt{\frac{1}{n}}} \right)^2 = \sum_{i=1}^n \left(\frac{x_{ij}}{n_j \sqrt{\frac{1}{n}}} - \frac{x_{ik}}{n_k \sqrt{\frac{1}{n}}} \right)^2$$

- La distancia del chi² entre dos modalidades :

$$d_{(j,k)} = \sqrt{\sum_{i=1}^n n \left(\frac{x_{ij}}{n_j} - \frac{x_{ik}}{n_k} \right)^2}$$

Desarrollando esta expresión tenemos:

$$d_{(j,k)}^2 = \sum_{i=1}^n n \left(\frac{x_{ij}^2}{n_j^2} + \frac{x_{ik}^2}{n_k^2} - 2 \frac{x_{ij}}{n_j} \frac{x_{ik}}{n_k} \right)$$

Si llamamos:

n_{jk} : La frecuencia bruta de individuos que presentaron simultáneamente la modalidad j y la modalidad K.

n_j : La frecuencia bruta de individuos que presentaron solamente la modalidad j.

n_k : La frecuencia bruta de individuos que presentaron solamente la modalidad k.

Luego el factor entre corchetes expresa:

- $\left(\frac{1}{n_j^2}\right)$ para $(n_j - n_{jk})$ individuos que presentaron solamente la modalidad j.

Es decir los individuos para los cuales $x_{ij} = 1$ mientras que $x_{ik} = 0$.

- $\left(\frac{1}{n_k^2}\right)$ para $(n_k - n_{jk})$ individuos que presentaron solamente la modalidad k.

Es decir los individuos para los cuales $x_{ik} = 1$ mientras que $x_{ij} = 0$.

- $\left(\frac{1}{n_j^2} + \frac{1}{n_k^2} - \frac{2}{n_j n_k}\right)$ para n_{jk} individuos que presentaron simultáneamente la

modalidad j y la modalidad k. Es decir los individuos para los cuales $x_{ik} = 1$ mientras que $x_{ij} = 1$.

- Ese factor vale 0 para todos los individuos que no presentaron la modalidad j ni la modalidad k.

Remplazando esos valores obtenemos:

$$d_{(j,k)}^2 = n \left[\left(\frac{n_j - n_{jk}}{n_j n_k} \right) + \left(\frac{n_k - n_{jk}}{n_j n_k} \right) \right]$$

El cuadrado de la distancia entre las modalidades j y k es igual a la proporción de individuos que presentaron la modalidad j, pero no la modalidad k., más la proporción de individuos que presentaron la modalidad k, pero no la modalidad j.

La distancia entre dos modalidades de una tabla lógica crece en función de la proporción de individuos que presentaron una y solo una de las dos modalidades.

Por otra parte, dicha distancia es inversamente proporcional a la importancia relativa de cada una de las modalidades.

De modo entonces que:

- Dos modalidades de una misma característica (mutuamente exclusivas) estarán obligatoriamente muy alejadas en el espacio de representación.
- Dos modalidades comunes a la mayoría de los individuos observados estarán representados en un mismo punto de espacio.
- Dos modalidades muy raras estarán representadas muy alejadas de todas las otras.

La distancia χ^2 asume correctamente el criterio de comparación de dos modalidades que hemos definido.

d) INERCIA TOTAL CON RESPECTO AL CENTRO E GRAVEDAD DE LA NUBE DE PUNTOS N (J)

Contribución de la j-ésima modalidad a la inercia global con respecto al centro de gravedad G_C es definida así:

$$Contr_{(j)} aI_{G_C} = p_j d_{(j,G_C)}^2 = \frac{n_j}{np} d_{(j,G_C)}^2$$

El cuadrado de la distancia de la j-ésima modalidad al centro de gravedad G_C :

$$d_{(j,G_C)}^2 = \sum_{i=1}^n \left(\frac{nx_{ij}^2}{n_j^2} + \frac{1}{n} - 2 \frac{x_{ij}}{n_j} \right)^2$$

Sabiendo que :

$x_{ij} = 1$ para los n_j individuos que presentaron la j-ésima modalidad.

$x_{ij} = 0$ en el caso de los $(n-n_j)$ individuos que no presentaron la j-ésima modalidad.

$$d_{(j,G_C)}^2 = n_j \left(\frac{n}{n_j^2} + \frac{1}{n} - \frac{2}{n_j} \right) + (n-n_j) \frac{1}{n} = \frac{n}{n_j} + \frac{n_j}{n} - 2 + 1 - \frac{n_j}{n} = \frac{n}{n_j} - 1$$

Por lo tanto:

$$Contr_{(j)} aI_{G_c} = \frac{n_j}{np} d_{(j,G_c)}^2 = \frac{n_j}{np} \left(\frac{n}{n_j} - 1 \right) = \frac{1}{p} \left(1 - \frac{n_j}{n} \right)$$

Esta expresión indica una modalidad rara, el cual es una fuerte contribución a la inercia global de la nube de puntos N(J).

La inercia global con respecto a G_c de la nube de puntos N(J).

$$\sum_{j=1}^K Contr_{(j)} aI_{G_c} = \sum_{j=1}^K \frac{1}{p} \left(1 - \frac{n_j}{n} \right) = \sum_{j=1}^K \frac{1}{p} - \sum_{j=1}^K \frac{n_j}{np}$$

Es decir que:

$$I_{G_c}^{N(J)} = \frac{K}{p} - 1$$

De modo entonces que la inercia global de la nube de puntos N(J) de una tabla lógica no depende mas que del numero de características observadas (p) y del numero total de modalidades que presentan esas características (K). decimos entonces que la inercia global de una tabla lógica no depende mas que de su "forma".

RELACIONES DE TRANSICIÓN:

Las relaciones de transición y tratándose del análisis de una tabla lógica, la definición es de la siguiente manera:

$$F_{\alpha}(i) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{j=1}^K \frac{x_{ij}}{p} G_{\alpha}(j); \forall i = 1, 2, \dots, n$$

$$G_{\alpha}(j) = \frac{1}{\sqrt{\lambda_{\alpha}}} \sum_{i=1}^n \frac{x_{ij}}{p} F_{\alpha}(i); \forall j = 1, 2, \dots, K$$

Como x_{ij} vale 0 o 1, la interpretación de esas relaciones de transición es la siguiente:

- La primera relación muestra que la proyección de un individuo i (i-ésima fila de la tabla lógica) sobre un eje factorial α corresponde (con un coeficiente de dilatación) al baricentro de las modalidades que fueron observadas sobre el i-ésimo individuo.

- La segunda relación muestra que la j -ésima modalidad de la tabla lógica es representada sobre un eje factorial α (con un coeficiente de dilatación) por la coordenada media de los individuos observados que presentaron la modalidad j .

La proximidad sobre un eje factorial α , de dos modalidades k y m traduce gráficamente una fuerte asociación entre esas variables indicadoras. La proximidad sobre un eje factorial α , entre dos individuos observados traduce gráficamente la similitud de las características presentadas por esos individuos.

TABLA DE BURT

La tabla de Burt $TB(K, K)$ es una tabla simétrica que presenta el conjunto de las tablas de contingencia que pueden ser construida cruzando dos a dos las p características observadas.

Llamamos, K la suma de las modalidades de la p características observadas sobre los n individuos, la tabla de Burt $TB(K \times K)$ presenta:

- A la intersección de la j -ésima línea y la j -ésima columna, el valor n_{jj} es decir el numero de individuos que presentaron la j -ésima modalidad de una característica dada.
- A la intersección de la j -ésima línea y de la k -ésima columna, el valor $n_{jk} = 0$ si la j -ésima modalidad y la k -ésima modalidad pertenecen a una misma característica observada.
- A la intersección de la p -ésima línea y de la k -ésima columna, el valor n_{pk} , es decir el numero de individuos que presentaron, simultáneamente, la p -ésima modalidad de una característica dada y la k -ésima modalidad de otra característica observada.

Tratándose de una tabla simétrica, construida con la yuxtaposición de todas las tablas de contingencia definidas por las p características observadas (tomadas dos a dos) de todas las variables cualitativas contenidas en ella.

La diagonal de la tabla TB (KxK) esta constituida por la distribución de frecuencias brutas de todas las modalidades, de todas las variables cualitativas.

a) ANALISIS FACTORIAL DE CORRESPONDENCIAS DE UNA TABLA DE BURT

Considerando la tabla TB (KxK) como una tabla de correspondencias, la información aportada por esta tabla TB(K , K) resulta de la comparación de todas las modalidades que la constituyen. Es necesario entonces, definir el criterio de comparación con el cual debemos tratar la información de dicha tabla.

Cada variable cualitativa realiza una partición de los n individuos observados en tantas clases como modalidades constituyan la variable. Dos clases dadas (dos modalidades) serán consideradas cuando presenten distribuciones de frecuencia equivalentes de las variables cualitativas consideradas. Dicho de otra manera, dos clases de individuos serán semejantes cuando presenten perfiles equivalentes de distribución del conjunto de variables cualitativas consideradas.

Podemos definir un espacio euclidiano de representación de la información aportada por una tabla de Burt. Representamos los puntos-perfiles ponderados de las k modalidades de la tabla en un referencial de k dimensiones. Dotando ese espacio de la distancia Chi² entre modalidades, se puede verificar que respeta el criterio de comparación de las mismas que hemos adoptado.

Tratándose de una tabla simétrica, la representación de una sola de las dos las nubes de puntos perfiles ponderados N(I) o N(J) es suficiente para representar toda la información contenida en una tabla de Burt.

La distancia del Chi² entre dos modalidades i e i' esta definida entonces de la siguiente manera:

$$d_{(i,i')}^2 = \sum_{j=1}^K \left(\sqrt{\frac{n}{p}} \frac{n_{ij}}{n_i \cdot n_j} - \sqrt{\frac{n}{p}} \frac{n_{i'j}}{n_i' \cdot n_j} \right)^2 = \frac{1}{p} \sum_{j=1}^K \frac{n}{n_j} \left(\frac{n_{ij}}{n_i} - \frac{n_{i'j}}{n_i'} \right)^2$$

Expresando esa distancia en términos de perfiles-ponderados, la distancia del χ^2 entre dos filas i e i' de una tabla de Burt queda definida de la siguiente manera:

$$d_{(i,i')} = \sqrt{\frac{1}{p} \sum_{j=1}^K \frac{1}{p^2 f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2}$$

Centrando ese espacio de representación en el centro de gravedad de esa nube de puntos-perfiles ponderados, y calculando la matriz de inercia de la nube centrada de puntos-modalidades, podemos determinar sus direcciones principales de alargamiento y definir un nuevo referencial de representación: los ejes factoriales.

El análisis factorial de Correspondencias de una tabla de Burt produce una buena representación grafica de la información aportada por dicha tabla.

b) COEFICIENTES DE AYUDA A LA SELECCION DE EJES FACTORIALES

➤ VALORES PROPIOS

Para cada eje factorial, la suma de las contribuciones a la inercia a lo largo del eje α de todas las modalidades de las p variables es igual a la inercia de la nube de puntos-modalidades proyectada a lo largo de dicho eje (λ_α).

La inercia a lo largo del eje α de todas las modalidades de las p variables es igual al valor medio del índice de correlación η^2 , de las p variables cualitativas con el eje α .

$$\text{Si } \lambda_\alpha = \frac{1}{p} \sum_{j=1}^p \eta_{(G_\alpha, j)}^2, \text{ cuando } \eta_{(G_\alpha, j)}^2 = 1, \forall j = 1, 2, \dots, p; \Rightarrow \lambda_\alpha = 1$$

Los valores propios de un análisis de correspondencias múltiples se encuentran en el intervalo [0,1].

En la practica de investigación, los p valores propios λ_α resultantes de un análisis factorial de correspondencias múltiples son $0 < \lambda_\alpha < 1$.

➤ **TASA DE INERCIA DE UN FACTOR**

La tasa de inercia de un factor no es mas que una proporción cuyo denominador carece de significación en términos de información total aportada por la tabla.

$$\tau_{\alpha} = \frac{\lambda_{\alpha}}{\sum_{\alpha=1}^p \lambda_{\alpha}} \times 100 = \frac{\lambda_{\alpha}}{I_G} \times 100$$

La tasa de la inercia de un factor no puede ser interpretada en términos de índice del grado de importancia o de interés que presenta un factor.

Los valores propios y las tazas de inercia de los ejes factoriales que resulta del Análisis Factorial de Correspondencias Múltiples de una tabla de “individuos-variables cualitativas” no son índices de naturaleza tal que permitan decidir la importancia relativas de los ejes factoriales.

➤ **CONTRIBUCIÓN DE LOS INDIVIDUOS Y DE LAS MODALIDADES A LA INERCIA A LO LARGO DE UN EJE**

En el caso del Análisis Factorial de Correspondencias Múltiples de una tabla lógica, los programas de aplicación calculan y editan las contribuciones de los puntos-individuos y de los puntos-modalidades a la inercia proyectada a lo largo de los ejes factoriales.

Las contribuciones de esos elementos son definidas de la siguiente manera:

$$CTR_{\alpha}(i) = \frac{\frac{1}{n} [F_{\alpha}^2(i)]}{\lambda_{\alpha}} \times 100; \forall i \in N(I) \qquad CTR_{\alpha}(j) = \frac{\frac{n_{.j}}{n} [G_{\alpha}^2(i)]}{\lambda_{\alpha}} \times 100; \forall j \in N(J)$$

Este índice nos permite definir el “grado de generalidad” de cada eje factorial. Decimos que un eje factorial representa un alto grado de generalidad cuando la inercia proyectada a lo largo del mismo resulta de la contribución de una gran cantidad de elementos (individuos o modalidades).

AYUDAS A LA INTERPRETACIÓN:

Se define la contribución de una variable J_k al factor α como la suma de las contribuciones de las modalidades de la variable.

$$CTA_{\alpha}(J_k) = \sum_{j \in J_k} CTA_{\alpha}(j)$$

LAS INERCIAS

1. La parte de la inercia debida a una modalidad de respuesta j es mayor cuanto menor sea el efectivo de esa modalidad, ya que esta inercia está dada por:

$$I(j) = \frac{1}{Q} \left(1 - \frac{n_j}{n} \right)$$

Por lo tanto se deben eliminar las modalidades elegidas muy pocas veces, uniéndola a la modalidad más próxima.

2. La parte de la inercia debida a una pregunta es función creciente del número de modalidades de respuesta que tiene ya que es la suma de las inercias de sus modalidades.

$$I(J_k) = \sum_{j \in J_k} I(j) = \frac{1}{Q} (m_k - 1)$$

Si una pregunta tiene muchas modalidades conviene reagrupar las modalidades en un número razonable. O podemos proyectar las modalidades sin agrupar como suplementarias.

3. La inercia total es la suma de las inercias de todas las preguntas.

$$I = k \sum I(J_k) = \frac{J}{Q} - 1$$

En general, en el análisis de la tabla Z, las tasas de inercia de los ejes dan una idea pésima de la información extraída por ellos.

INTERPRETACIÓN:

La interpretación se hará siguiendo las mismas normas que en ACS, pero debemos tener en cuenta las características particulares de la inercia en este análisis.

CAPITULO III

3.1 METODOLOGIA

3.1.1 Tipificación de la investigación

Por el tipo de conocimiento: científica

Por la naturaleza del objeto del estudio: formal

Por el tipo de pregunta: Descriptiva

Por el método de estudio de las variables: cuantitativa

Por el método de variables: Multivariada.

Por el ambiente en que se realiza: de campo

Por la fuente de datos que se emplean: primaria

Por le enfoque utilitario predominante: pragmático

Por la profundidad en el tratamiento del tema; estudio propiamente dicho

Por el tiempo de aplicación de la variable: transversal o sincrónica

3.1.2 Diseño de Investigación

Es No experimental, transversal y descriptivo Multivariado. Diremos que es no experimental dado que no se intervendrá a los pacientes para observar un efecto especificado. Es transversal puesto que se levantara la información en una fecha especificada. Y es descriptivo multivariado porque nos va permitir describir las características de los pacientes y su satisfacción acerca el precio, calidad, ubicación y la publicidad de la clínica odontológica de la UNMSM a través del Análisis de correspondencia múltiple.

3.1.3 Variables

- X1: Edad
- X2: Sexo
- X3: Grado de instrucción
- X4: Publicidad
- X5: Comodidad de la sala de espera
- X6: Servicio de atención
- X7: Atención del personal administrativo
- X8: Atención por parte del operador
- X9: Calidad
- X10: Precio
- X11: ubicación

3.1.4 Operacionalización de las variables

Variabes	tipo	Indicador	Escala
X1: Edad	Cuantitativa	De 18 a 28 años De 29 a 39 años De 40 a 50 años De 51 a 61 años	Razón
X2: Sexo	Cualitativa	Femenino Masculino	Nominal
X4: Grado de instrucción	Cualitativa	Primaria Secundaria Técnica Superior	Ordinal
X4: Publicidad	Cualitativa	Por recomendación Paginas amarillas Internet Radio	Nominal
X5: Comodidad de la sala de espera	Cualitativa	Si No	Nominal

X6: Servicio de atención	Cualitativa	Clinica 1,2,3 Odontopediatria Diagnostico Radiología	Nominal
X7: Atención del personal administrativo	Cualitativa	Excelente Buena Regular Mala Pésima	ordinal
X8: Atención por parte del operador	Cualitativa	Excelente Buena Regular Mala Pésima	ordinal
X9: Calidad	Cualitativa	Excelente Buena Regular Mala Pésima	ordinal
X10: Precio	Cualitativa	Precios altos Precios justos Precios económicos	
X11: ubicación	Cualitativa	Excelente Buena Regular Mala Pésima	

3.1.5 Población

El presente estudio tomará como fuente informativa a la población conformada por los pacientes que han recibido atención en la clínica Odontológica de UNMSM en el mes de Mayo del 2008.

3.1.6 Muestra

La muestra de estudio de esta investigación corresponde a las denominadas muestra probabilísticas, Se estableció el tamaño de la muestra y el tipo de muestreo que se utilizará para seleccionar la muestra de los alumnos

3.1.6.1 Tamaño de la Muestra

Se aplicara el muestreo aleatorio simple apara proporciones con un margen de error de 5% y un nivel de confiabilidad de 95%

Fórmula:

$$n = \frac{N(P)(Q)(Z^2)}{(N - 1)e^2 + (P)(Q)Z^2}$$

Donde:

N: Población muestreada del estudio (N=292)

P: Proporción de pacientes que opinan favorablemente P=0.8212

Q: 1-0.08212 = 1-0.1788 complemento de P

Z: Coeficiente de confiabilidad al 95% igual a 1.96 con = 0.95

E: Máximo error permisible en la investigación e = 0.055%)

Aplicando la fórmula:

$$n = \frac{292(0.8212)(0.1788)(1.96^2)}{(292 - 1)0.055^2 + (0.8212)(0.1788)1.96^2} = \frac{164,518748}{1,29156428} = 127,38 \approx 127$$

Se seleccionará 127 pacientes de la clínica odontológica

3.1.7 Instrumentos de Recolección de Datos

Para el efecto de la recolección de datos se seleccionó como técnica la entrevista y como instrumento se utilizó el cuestionario, el cual fue aplicado en forma personal a cada paciente por parte de la investigador; considerándose a estos como medios efectivos para recolectar datos reales sobre el presente estudio siendo su objetivo el de obtener información acerca aspectos sociodemográficas y opinión sobre la publicidad, precio, calidad y ubicación de la clínica odontológica

3.1.8 Confiabilidad

La confiabilidad del instrumento se determinó a través de la aplicación de una prueba piloto que se realizó la clínica odontológica, tomando los 30 pacientes de muestra piloto con similares características a la del estudio pero que no pertenece a la población. Así se aplicaron formulas estadísticas tales como el coeficiente de alfa de Cronbach para la confiabilidad, obteniéndose como resultado:

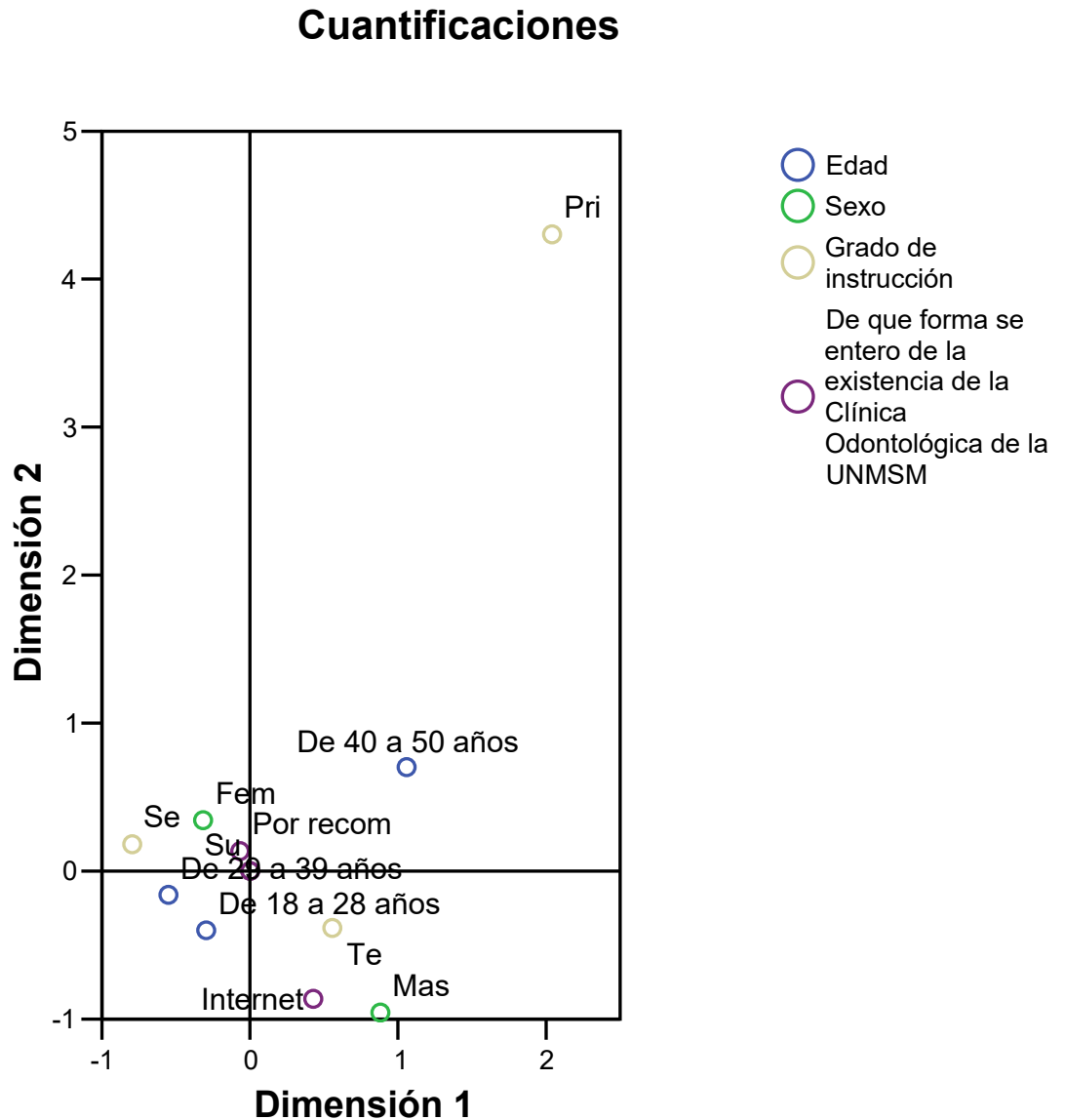
Instrumento para medir la variable aspiraciones estudiantiles Se utilizo el Alfa de Cronbach = 0.99 muy alto .5 (Anexo "A"). Habiéndose encontrado resultado confiable y consistente a través del tiempo para todo el instrumento en general.

Observándose que luego de su aplicación, los resultados son confiables. Estos resultados permitieron reestructurar y/o discriminar items, garantizando su consistencia y coherencia interna así como reducir al máximo el margen de error.

CAPITULO IV

4.1 RESULTADOS

FIGURA N°1: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO, GI,

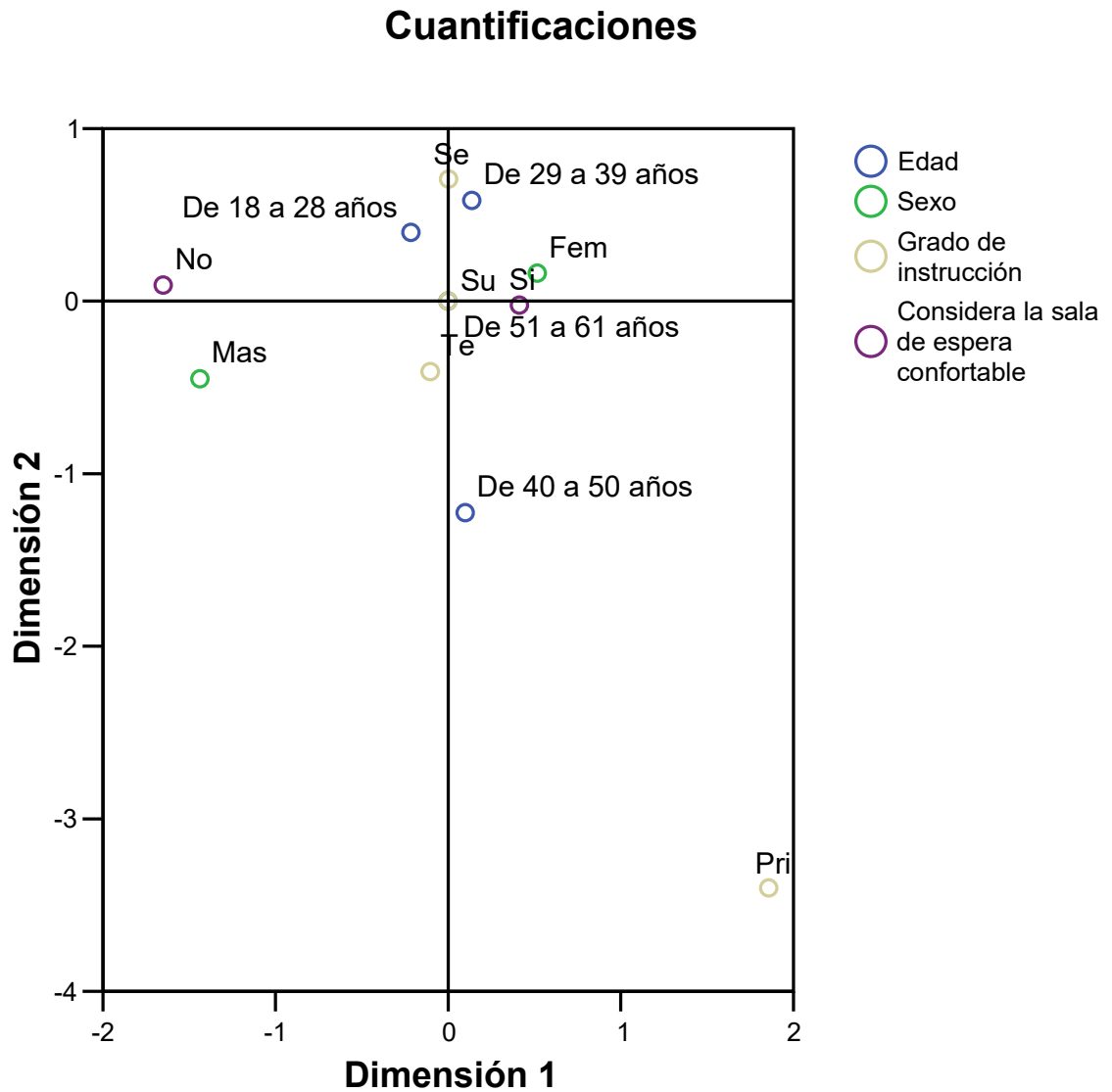


Del análisis de los gráficos de la Figura 1 se aprecia que:

-Las personas adultas y adultos mayores de sexo femenino con grado de instrucción secundaria se enteraron de la clínica universitaria por medio de la recomendación.

-Las clientes de 18 a 28 des sexo masculino con grado de instrucción de técnica, se entero de la clínica universitaria a través del Internet.

FIGURA N°2: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “Considera la sala de espera confortable”

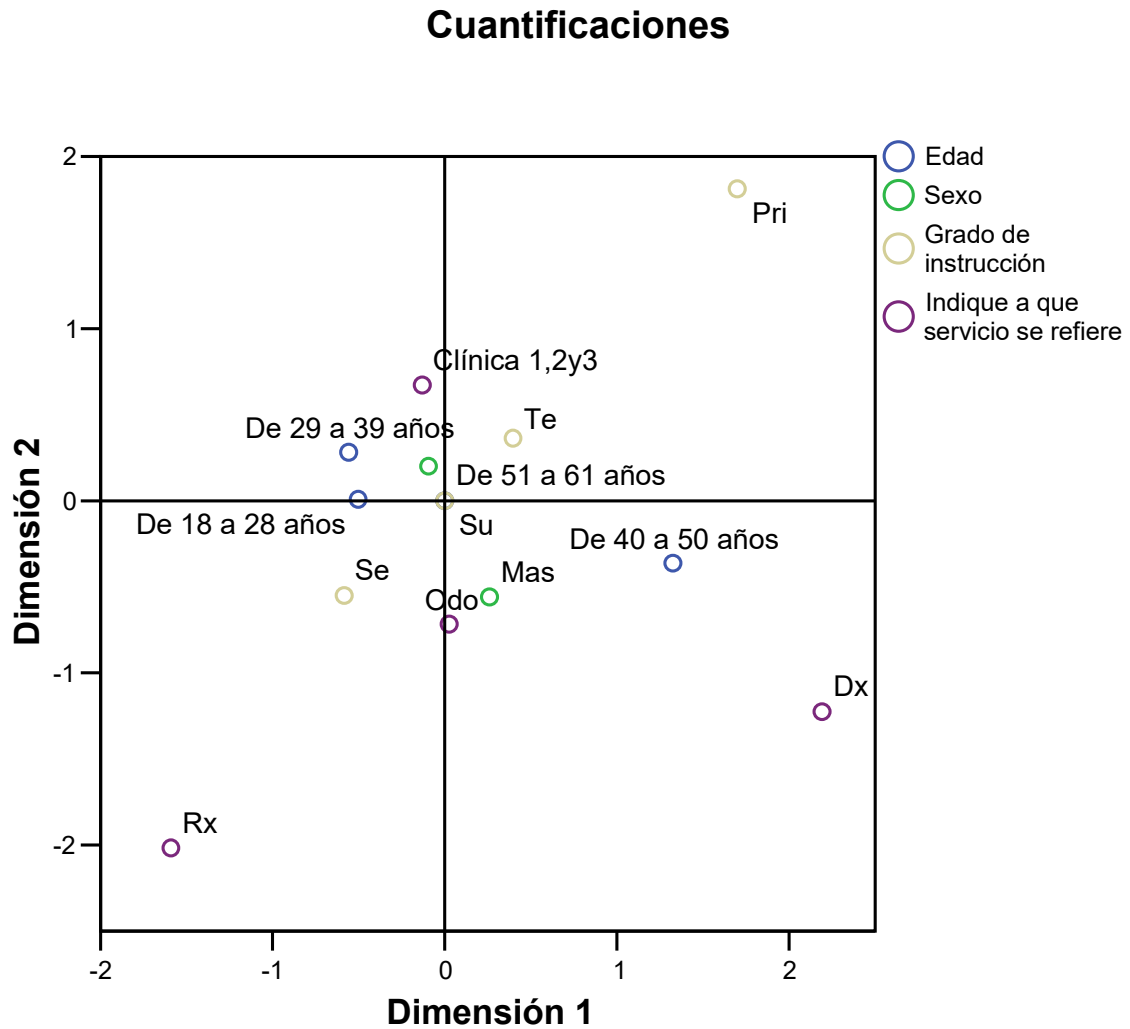


Del análisis de los gráficos de la Figura 2 se aprecia que:

-Las personas adultas y adultas mayor de sexo femenino con grado de instrucción secundaria a superior **considera que Considera que la sala de espera si confortable**

-Las clientes de 18 a 28 des sexo masculino , **Considera que la sala de espera no confortable**

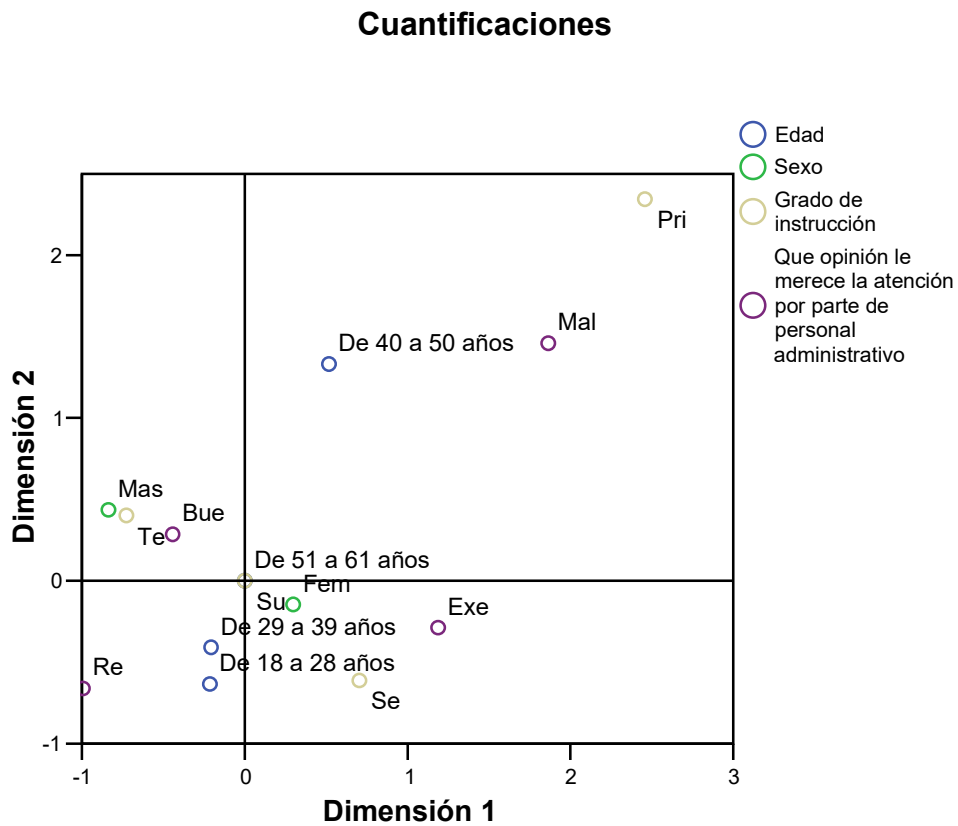
FIGURA N°3: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “Indique a que servicio se refiere”



Del análisis de los gráficos de la Figura 4 se aprecia que:

-Los clientes se dirigen con mayor frecuencia a clínica 1,2 y 3 y son de edad de 29 a 39 años y las personas adultas y adulta mayor (51 a 61 años) se dirige odontopediatría.

FIGURA N°4: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “Que opinión le merece la atención por parte del personal administrativo?”



Del análisis de los gráficos de la Figura 4 se aprecia que:

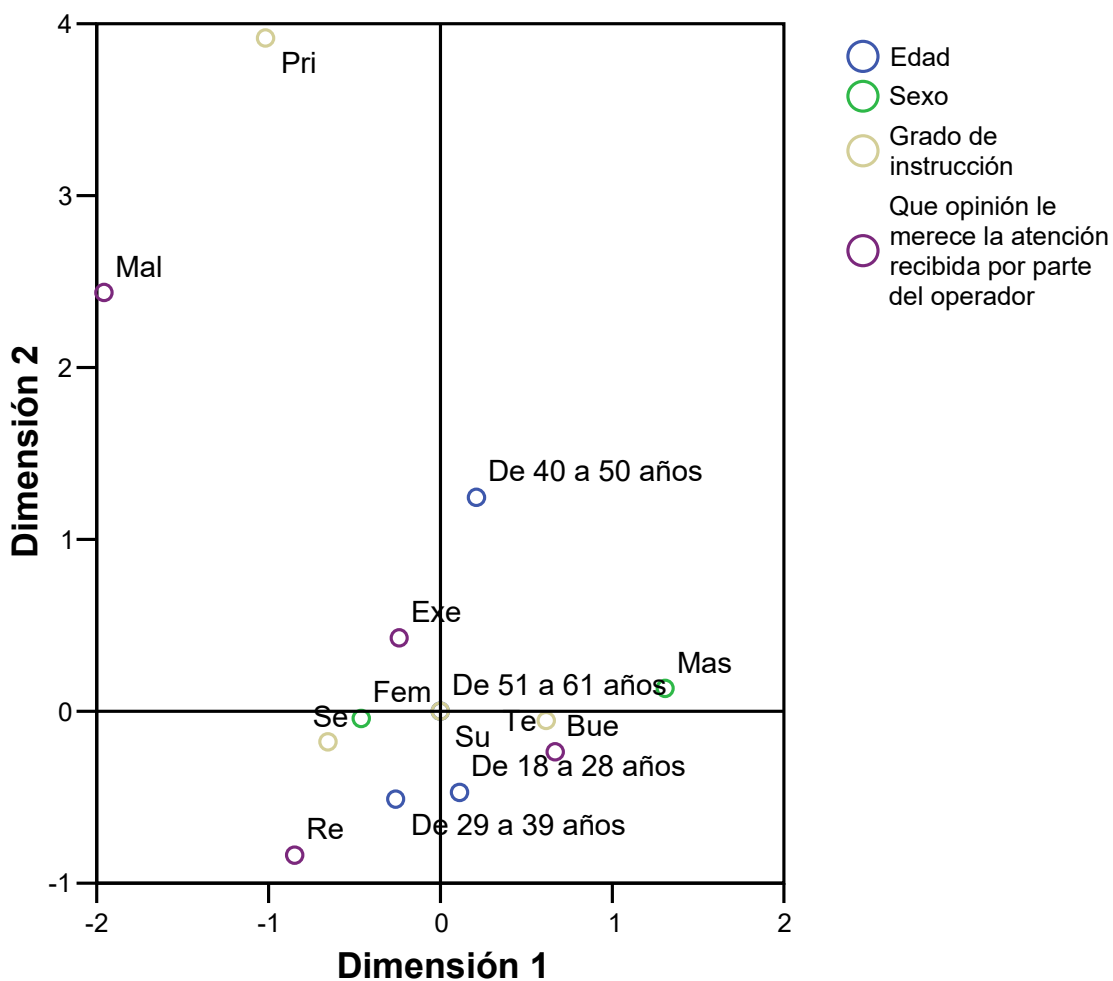
-Los clientes de 40 a 50 años con grado de instrucción primaria consideran que el trato del personal administrativo es mala.

-Los clientes de masculino con grado de instrucción técnica consideran que el trato es bueno

-Los clientes femeninos de 18 a 28 y 29 a 39 años con grado de instrucción superior consideran que el trato del personal administrativo es excelente

FIGURA N°5: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “Que opinión le merece la atención recibida por parte del operador”

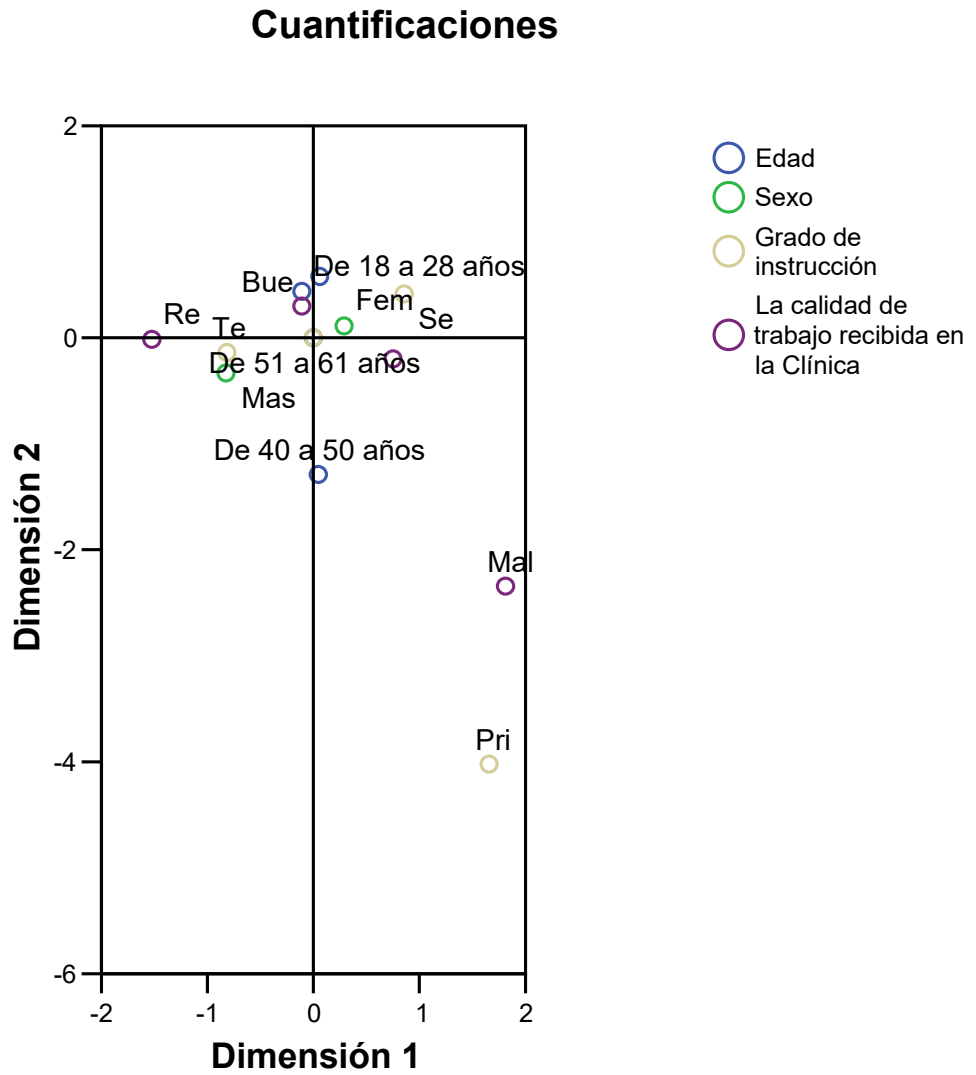
Cuantificaciones



Del análisis de los gráficos de la Figura 5 se aprecia que:

- Los clientes tienen una buena y excelente opinión acerca de la atención recibida por el operador
- Los clientes de 40 a 50 años tienen una mala percepción acerca de la atención del operador.

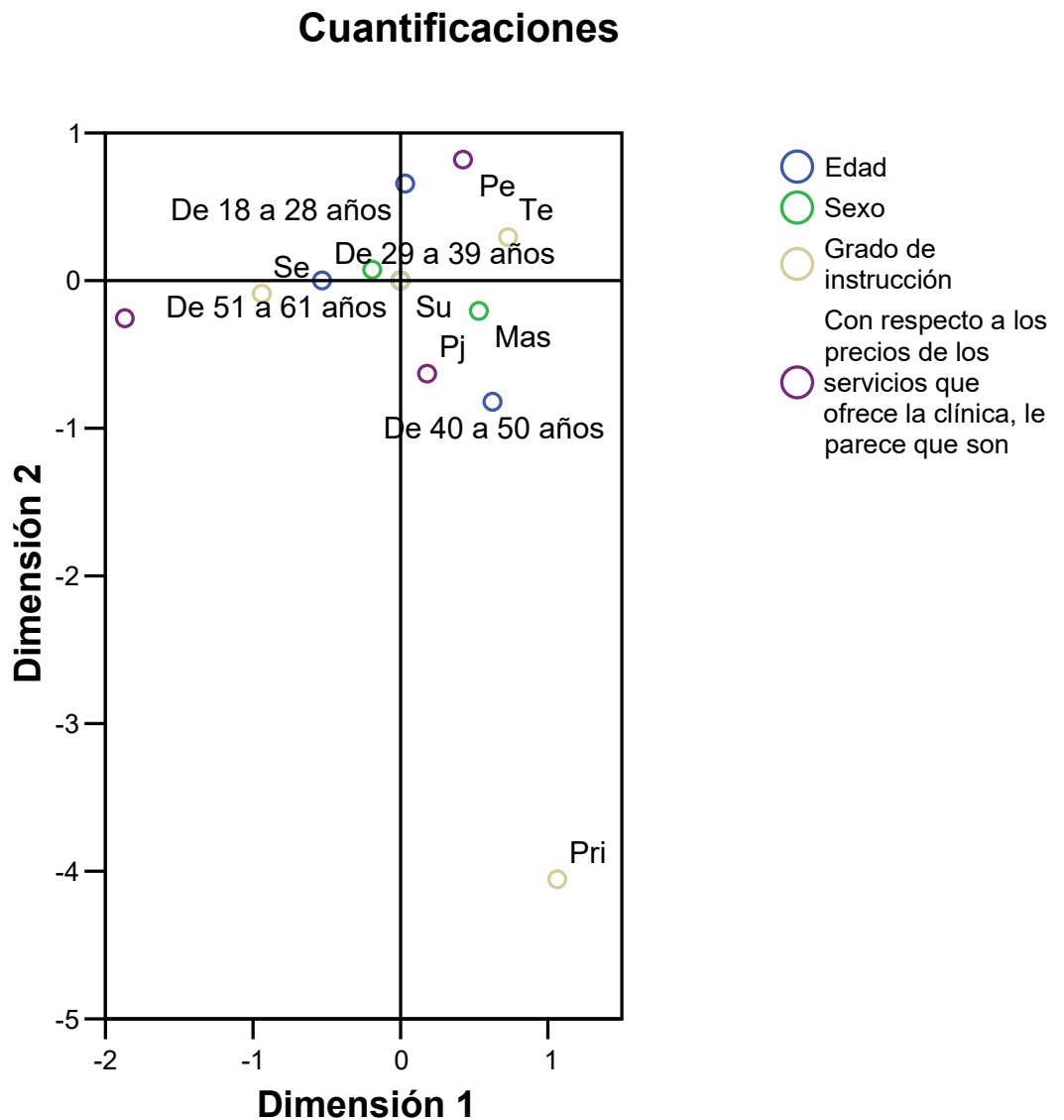
FIGURA N°6: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “La calidad de trabajo recibida en la Clínica



”

Del análisis de los gráficos de la Figura 6 se aprecia que:

FIGURA N°7: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “La calidad de trabajo recibida en la Clínica



Del análisis de los gráficos de la Figura 7 se aprecia que:

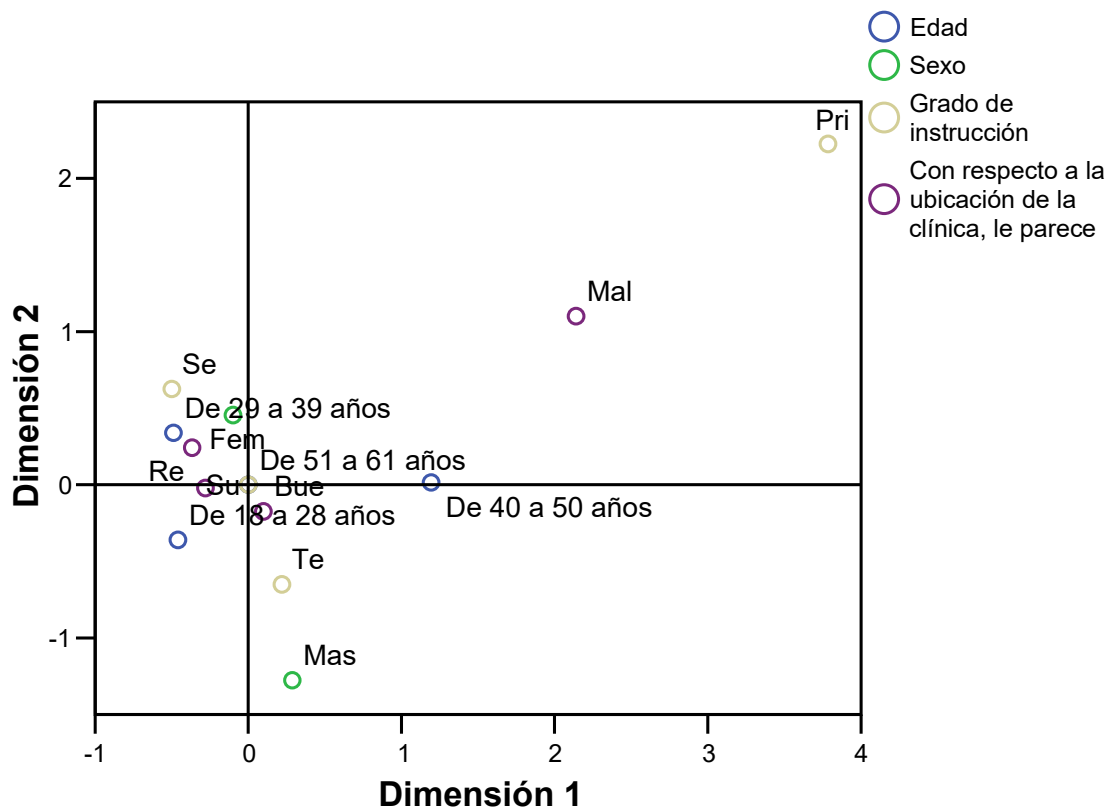
-Los clientes de 40 a 50 años de sexo masculino y grado instrucción superior consideran que son precios justos

-De 51 a 61 años creen que son precios caros

-De 18 a 28 años consideran que so precios económicos

FIGURA N°8: GRÁFICO DE LAS PUNTUACIONES OBTENIDAS PARA EDAD SEXO , GI y la variable “Con respecto a la ubicación de la clínica, le parece”

Cuantificaciones



Del análisis de los gráficos de la Figura 8 se aprecia que:

-Los clientes femeninos de 18 a 28 consideran que la ubicación es buena ; Los clientes femeninos de 29 a 39 consideran que la ubicación es regular ,nadie considera que la ubicación es mala.

CONCLUSIONES

1. La clínica odontológica tiene como mayor fuente de publicidad las recomendaciones de los pacientes, reflejadas en las personas adultas y adultas mayores, este grupo de personas están satisfechas con la comodidad de la clínica y son los que se dirigen a la clínica 1,2 y 3
2. Los clientes de 40 a 50 años consideran que trato es malo por parte del operador y el personal administrativo y consideran que los precios son justos
3. Se encontró que los clientes menos de 40 años están satisfechos con el operador.
4. los pacientes jóvenes (18 a28 años) se informa de la clínica a través del Internet , tienen buena percepción de la calidad de servicio pacientes consideran que la calidad es buena en ubicación , publicidad, trato y comodidad.
5. A lo que respecta el servicio un mayor porcentaje de pacientes consideran que la atención en general es buena.
6. En el precio que da la clínica por sus servicios, debido a que los pacientes lo consideran en un mayor porcentaje económicos.
7. En los elementos: servicio, precio, esta considerado buena por los pacientes, esto significa que estamos por buen camino pero podemos lograr la excelencia, por ella se darán estrategias de marketing en las recomendaciones.

RECOMENDACIONES

- Por convenios con laboratorios y/o entidades privadas crear una señaleptica internas y externas que auspicien la clínica.
- Dar facilidades a las empresas proveedoras de insumos dentales como: cepillos, pastas dentales, etc. Para el servicio del cliente.
- Realizar publicidad mediante volantes, cartillas a las personas que están en nuestra clínica y a los que viven a los alrededores de dicha institución.
- Publicitar notas de prensa a los diferentes medios de comunicación para informar de la existencia de la clínica (son gratuitas).
- Invitar canales de televisión para conocer instalaciones de la clínica para promocionarlas.
- Crear una base de datos de los pacientes para enviar cartas. Informándolos de las promociones que se realizan.
- Publicitar un convenio con el Ministerio de Salud; consistiría en derivar a los pacientes de dicha institución a nuestra clínica, en caso de no poder concluirles algún tratamiento.
- Programar charla de higiene bucal y promocional para los pacientes en la clínica y a los pacientes potenciales.
- Debemos contar con computadoras para registrar todos los datos de los pacientes, historia, citas, etc.
- Por medio de anfitrionas orientar al paciente desde su llegada a la clínica, con el fin de resolver sus dudas e informar sobre todos los servicios que se brindan.
- Realizar promociones en días especiales como: DIA del odontólogo, DIA del niño, DIA de la madre, padre, anciano o por cambios de estación.
- Crear un seguro dental para los alumnos de la universidad y colegios aledaños a la clínica.

BIBLIOGRAFÍA

DESDE UN PUNTO DE VISTA PRÁCTICO:

HAIR, J., ANDERSON, R., TATHAM, R. y BLACK, W. (1999). *Análisis Multivariante*. 5ª Edición. Prentice Hall.

DESDE UN PUNTO DE VISTA TEÓRICO-PRÁCTICO:

GERI (1996) *L'Analyse des données évolutives : methods et applications*. Editions Technip. (Un buen libro sobre Análisis de Correspondencias Dinámico)

GIFI, A. (1990). *NonLinear Multivariate Analysis*. Wiley (Un buen libro para profundizar en el Análisis de Correspondencias Múltiples)

JOBSON, J.D. (1992) *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods*. Springer-Verlag.

LEBART, L; MORINEAU, A. and PIRON, M. (2000). *Statistique Exploratoire Multidimensionnelle*. 3ª Edition. DUNOD.

ENFOCADOS HACIA SPSS:

AGUILERA, A. (2000). *Tablas de Contingencia Bidimensionales*. Cuadernos de Estadística. Editorial La Muralla

JORAISTI, L. y LIZOSAIN, L.(2000) *Análisis de Correspondencias*. . Cuadernos de Estadística. Editorial La Muralla.

PEREZ, César (2001). *Técnicas Estadísticas con SPSS*. Prentice-Hall

VISAUTA, B. (1998) *Análisis Estadístico con SPSS para WINDOWS (Vol II. Análisis Multivariante)*. Mc-Graw Hill.

ANEXOS

CUESTIONARIO

El presente instrumento forma parte del trabajo de investigación titulada: Para caracterizar al paciente externo de la clínica odontológica de UNMSM .Por lo que solicitamos su participación, desarrollando cada pregunta de manera objetiva y veraz. La información es de carácter confidencial y reservado; ya que los resultados serán Manejados solo para la investigación. Agradezco anticipadamente su valiosa colaboración.

1. Edad:.....

2. Sexo

a. Masculino () b. Femenino ()

3. Distrito:

4. Grado de instrucción

a. Primaria () b. Secundaria () c. Superior ()

5. De que forma se entero de la existencia de la Clínica Odontológica de la UNMSM

a. Por recomendación. () b. Paginas amarillas. () c. Internet. () d. Radio ()

6. Considera la sala de espera confortable

a. Si () b. No ()

7. Indique a que servicio se refiere:

a. Clínica 1,2y3 () b. Odontopediatria () c. Diagnostico () d. Radiología ()

8. Que opinión le merece la atención por parte de personal administrativo.

a. Excelente.() b. Buena. () c. Regular. () d. Mala.() e. Pésima.()

9. Que opinión le merece la atención recibida por parte del operador.

a. Excelente.() b. Buena.() c. Regular.() d. Mala. () e. Pésima.

10. La calidad de trabajo recibida en la Clínica:

a. Excelente. () b. Buena.() c. Regular.() d. Mala.() e. Pésima.

11. Con respecto a los precios de los servicios que ofrece la clínica, le parece que son:

- a. Precios altos. () b. Precios justos.() c. Precios económicos. ()

12. Con respecto a la ubicación de la clínica, le parece:

- a. Excelente.() b. Buena.() c. Regular.() d. Mala.() e. Pésima. ()

TABLAS

TABLA N° 1

Distribución de Edad de los clientes entrevistados

	Frecuencia	Porcentaje	Porcentaje Acumulado
De 18 a 28 años	50	35.7	35.7
De 29 a 39 años	50	35.7	71.4
De 40 a 50 años	40	28.6	100.0
Total	140	100.0	

TABLA N° 2

Distribución del genero de los clientes entrevistados

	Frecuencia	Porcentaje
Masculino	37	26.4
Femenino	103	73.6
Total	140	100.0

TABLA N° 3

Distribución del Grado de instrucción de los clientes entrevistados

	Frecuencia	Porcentaje
Primaria	4	2.9
Secundaria	62	44.3
Tecnica	74	52.9
Total	140	100.0

TABLA N° 4

Distribución de la forma que se entero de la existencia de la Clínica Odontológica de la UNMSM

	Frecuencia	Porcentaje
Por recomendacion	121	86.4
Internet	19	13.6
Total	140	100.0

TABLA N° 5

Considera la sala de espera confortable

	Frecuencia	Porcentaje
Si	112	80.0
No	28	20.0
Total	140	100.0

TABLA N° 6

Indique a que servicio se refiere

	Frecuencia	Porcentaje
Clínica 1,2y3	86	61.4
Odo	32	22.9
Dx	12	8.6
Rx	10	7.1
Total	140	100.0

TABLA N° 7

Que opinión le merece la atención por parte de personal administrativo

	frecuencia	Porcentaje
Excelente	41	29.3
Buena	68	48.6
Regular	25	17.9
Mala	4	2.9
Pésima	2	1.4
Total	140	100.0

TABLA N° 8

Que opinión le merece la atención recibida por parte del operador

	Frecuencia	Porcentaje
Excelente	56	40.0
Buena	55	39.3
Regular	24	17.1
Mala	3	2.1
Pesima	2	1.4
Total	140	100.0

TABLA N° 9

La calidad de trabajo recibida en la Clínica

	Frecuencia	Porcentaje
Excelente	57	40.7
Buena	56	40.0
Regular	24	17.1
Mala	1	.7
Pesima	2	1.4
Total	140	100.0

TABLA N° 10

Con respecto a los precios de los servicios que ofrece la clínica, le parece que son

	Frecuencia	Porcentaje
Pa	19	13.6
Pj	65	46.4
Pe	56	40.0
Total	140	100.0

TABLA N° 11

Con respecto a la ubicación de la clínica, le parece

	Frecuencia	Porcentaje
Excelente	50	35.7
Buena	60	42.9
Regular	22	15.7
Mala	7	5.0
Pesima	1	.7
Total	140	100.0