

03043



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS
APLICADAS Y EN SISTEMAS

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS
Y DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

TESINA:

**CLASIFICACIÓN DE INDIVIDUOS POR MEDIO DE CONDUCTAS
ALIMENTARIAS DE RIESGO APLICANDO EL MÉTODO
MULTIVARIADO DE ANÁLISIS DE CONGLOMERADOS DE K-MEDIAS**

PRESENTADO POR LA ALUMNA: M. EN C. ARACELI ESCOBAR ROJAS

ASESORA PRINCIPAL: M. EN C. HORTENSIA MORENO MACÍAS

ASESORA EXTERNA: DRA. CLAUDIA UNIKEL SANTONCINI



MÉXICO, D.F.

2005

m347669



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Atacell Escobar Rojas

FECHA: 9/sep/05

FIRMA: [Firma]
A.V.

*Cada uno está obligado a comprender el mundo
y a comprenderse a sí mismo simultáneamente.*

*El hombre que se va de este mundo sin saber
quién es ni donde estuvo, es como un tonto en
vísperas, aunque sea un tonto loco, desesperado
y genial.*

*Juan José Arreola
La palabra educación, 1977.*

*Profesar una cátedra, son los actos equivalentes que indican un común
afán de conocimiento por parte del maestro y del alumno.*

*Agradezco a los profesores de la Especialidad en Estadística Aplicada
su gran dedicación y amor a su profesión.*

*Agradezco también a un alma sensible y noble que ha sabido encaminar
con paciencia las ideas y sugerencias expresadas en este trabajo.
Hortensia Moreno.*

Contenido

	Páginas
Resumen	4
Introducción	5
Antecedentes	8
Objetivo	10
Metodología	10
Análisis estadístico	15
Resultados	23
Discusión	52
Conclusiones	54
Anexo	55
Bibliografía	67

Resumen

Objetivo: Clasificar individuos, a través de su patrón de respuesta a conductas alimentarias de riesgo, aplicando la técnica multivariada de análisis de conglomerados (*cluster analysis*) no jerárquico, en una muestra de estudiantes adolescentes y adultos jóvenes del estado de Morelos, México.

Material y métodos: Es un estudio transversal, observacional, retrospectivo, y descriptivo. La información proviene de la segunda evaluación del estudio poblacional: “Estilos de vida en adolescentes y adultos jóvenes y su relación con el desarrollo longitudinal de enfermedades crónicas”, cuyo marco muestral son las escuelas públicas de nivel medio, medio superior y superior del estado de Morelos, México; del ciclo escolar 2001-2002. Las conductas alimentarias de riesgo se obtuvieron de un cuestionario autoaplicable de 13 preguntas (α de Cronbach de 0.54); de las cuales, diez de ellas derivan del cuestionario EAT-40 y las otras tres preguntas, fueron elaboradas para el estudio de adolescentes, con una escala de respuesta de tipo Likert. El algoritmo empleado para clasificar individuos con riesgo o con síndrome parcial de trastornos de la conducta alimentaria, fueron dos criterios cuantitativos, basados en el análisis de conglomerados de k-medias, donde se utilizan datos ordinales y datos binarios.

Resultados: La prevalencia del grupo de riesgo, encontrado para el criterio cualitativo, clasificación sugerida por expertos, fue de 3.06% en los hombres y de 3.24% en las mujeres. En cambio, la prevalencia obtenida para los criterios cuantitativos con datos ordinales fue de 4.2% en los hombres y 4.6% en las mujeres. El resultado para los conglomerados elaborados con datos binarios fue de 4.6% para ambos géneros. Cuando se comparan el criterio cualitativo con los conglomerados ordinal y binario, se observa un valor de concordancia elevado en ambas agrupaciones.

Conclusiones: La clasificación obtenida mediante la técnica multivariada de análisis de conglomerados, con datos ordinales y con datos binarios puede llegar a ser una herramienta a considerar para futuros análisis de estos trastornos mentales, ya que permite dividir individuos con riesgo, sin la utilización del análisis cualitativo, metodología más sujeta a la subjetividad del investigador.

Introducción

Epidemiología psiquiátrica

La epidemiología psiquiátrica tradicionalmente se rezaga de otras ramas de la epidemiología por las dificultades contenidas en la conceptualización y en la medición de trastornos mentales. Las encuestas epidemiológicas aplicadas en adultos revelan que los desórdenes mentales se establecen a edades tempranas, por ello, la investigación de factores de riesgo en niños y adolescentes, es un buen avance en este tipo de estudios, sin embargo, existen algunos obstáculos en estas poblaciones, que hacen un gran reto para la medición de estos trastornos, tal es el caso de la presencia de desórdenes mentales en etapas tempranas de la vida, los cuales son menos “cristalizados” que los desórdenes en los adultos; de tal manera que en ocasiones los diagnósticos son errados. Cuando se requiere obtener información de los auto reportes, el reto se traduce, en que los encuestados, tengan la habilidad suficiente para interpretar adecuadamente las preguntas en los cuestionarios;¹ estos problemas tienen gran relevancia en el presente estudio, debido a que la muestra está constituida por una muestra de adolescentes y adultos jóvenes quienes respondieron a un cuestionario autoaplicado.

Trastornos de la conducta alimentaria

En la actualidad, uno de los trastornos psiquiátricos que crece continuamente es los trastornos de la conducta alimentaria (TCA), padecimiento que agobia a muchas mujeres y que cada vez se diagnostica a edades más tempranas.² La Encuesta Nacional de Epidemiología Psiquiátrica (ENEP), efectuada en el año 2000, detectó que el 1.8% de la población femenina de 18 a 65 años de edad, alguna vez ha padecido bulimia nerviosa.³

La presencia de conductas anómalas para controlar el peso corporal, también se han incrementado en niños y en jóvenes del sexo masculino.⁴ Posiblemente, estas conductas sean una manifestación a la distorsión corporal que también sufren los varones; en un estudio efectuado en tres países desarrollados, Austria, Estados Unidos y Francia, se encontró una alta discrepancia entre lo que son los “ideales de belleza masculina”, predominantemente musculosa y la percepción que tienen de su cuerpo; además se encontró que los varones encuestados, creen que las mujeres prefieren hombres musculosos.⁵

Los trastornos de la conducta alimentaria, presentan actitudes y conductas inapropiadas, que les caracterizan, como el sometimiento a dietas, los atracones, el uso de laxantes, diuréticos, el vómito autoinducido, el ejercicio excesivo, el miedo a ganar peso, etc., (DSM-IV Manual Diagnóstico y Estadístico de los Trastornos mentales).⁶ Cuando estas actitudes y conductas anómalas no cumplen en frecuencia y duración con los requerimientos diagnósticos, se les denomina conductas alimentarias de riesgo (CAR).⁷ De tal manera que un individuo con cierta cantidad de estas conductas en una escala de respuesta de alta frecuencia (“con frecuencia” o “siempre”), evidencia un síndrome parcial o subclínico de los TCA.⁸

En la mayoría de las ocasiones, los individuos van a ser clasificados en dos categorías: sanos y enfermos. Esto obliga a disponer de criterios diagnósticos no siempre suficientemente definidos.⁹

La necesidad práctica de estos criterios es obvia, aunque, la distinción entre enfermo y sano en ocasiones es una convención médica útil para decidir si el individuo ha de ser tratado o no, pues la mayoría de las enfermedades constituyen fenómenos progresivos o continuos y no categóricos.¹⁰ Por lo tanto la definición de criterios diagnósticos, estén basados en síntomas, signos o pruebas clínicas, constituye el primer paso en la medición de cualquier problema de salud.

En el caso de las conductas alimentarias de riesgo, la metodología empleada para clasificar a individuos con riesgo o síndrome parcial, se han realizado por medio de la clasificación cualitativa, en donde, el criterio del investigador es fundamental, de tal manera que la realización frecuente de tres o más conductas alimentarias de riesgo, pueden llegar a determinar riesgo en la persona que las realiza.⁸

La clasificación descrita anteriormente, posee varias interrogantes, por ejemplo, ¿qué sucede con las personas que poseen una muy alta frecuencia en dos de las conductas alimentarias, o cuando la persona reporta varias conductas con una frecuencia media, se consideran personas con riesgo?

Por otra parte, en el proceso de validación de escalas que indagan conductas alimentarias de riesgo en cuestionarios, se efectúa un análisis discriminante, para darle validez predictiva dicha escala. Para realizar este método multivariado, generalmente se compara una muestra que tiene trastornos psiquiátricos (valorados clínicamente), contra una muestra sana; y de esta manera predecir la pertenencia de una observación en particular, con base en un conjunto de variables predictoras.¹¹

Posteriormente, se determina un punto de corte de la escala que permita diferenciar a los individuos con riesgo de los sanos y donde la sensibilidad (la habilidad de la prueba para identificar correctamente a los enfermos) y la especificidad (la habilidad de la prueba para identificar a las personas que no son enfermos) deben ser elevadas.¹² Finalmente, se determina la puntuación total del individuo (escala ordinal), por medio de la suma de las calificaciones de las variables consideradas y se compara con el punto de corte definido, si la persona cuenta con una puntuación total mayor a este punto, ésta se considera con riesgo.¹³ En esta tesina no se lleva a cabo este análisis porque no se cuenta con una muestra clínica.

En el presente trabajo se pretende dar dos alternativas de solución cuantitativas al problema de clasificar individuos, -sanos y no sanos (en riesgo)-, que presentan una escala de respuesta categórica (escala de Likert) en algunas conductas alimentarias de riesgo; para lograr este objetivo, se presentan estrategias de clasificación por medio de la técnica multivariada conocida como análisis de conglomerados.

El hecho de que no haya a *priori* una clasificación en la muestra sugiere que es fundamental realizar análisis de conglomerados. Luego entonces, es crucial la elección de variables que agruparán los datos, para la obtención de grupos homogéneos.¹¹ Uno de los usos de este análisis es para definir o redefinir categorías de diagnóstico en psiquiatría. Este análisis es una técnica exploratoria, que trabaja con muchas medidas de similitud.¹⁴

En general, este método multivariado es utilizado para distintos objetivos,¹⁵ entre ellos:

- 1.- Desarrollo de una tipología o clasificación.
- 2.- Investigación de esquemas conceptuales para agrupar entidades.
- 3.- Generación de hipótesis a través de la exploración de datos.
- 4.- Pruebas de hipótesis, o ensayos para determinar si las clasificaciones definidas a través de otros procedimientos están en realidad presentes en los datos que se analizan.

En esta tesina, la utilización de este análisis tiene el objetivo de cubrir el primer punto. No se desarrollarán los métodos jerárquicos debido a que los datos no presentan una estructura jerárquica, sin embargo, se describen en el anexo, por considerarlos de suma importancia.

Antecedentes

El problema de la medición en variables psicológicas

En la vida cotidiana la palabra “medición” posee un significado claro y conciso en variables tangibles. La situación es diferente cuando queremos medir variables psicológicas. Una variable psicológica se define como una propiedad o característica que poseen diferentes individuos en cantidades distintas. Al medir variables como aptitudes de aprendizaje, depresión o conductas alimentarias de riesgo, hay problemas de escalamiento que se toman muy complejos. Medir es dar magnitud de cierta propiedad a uno o más objetos con ayuda del sistema numérico.¹⁶

Una de las escalas más utilizadas para medir actitudes y conductas anómalas en la alimentación es la escala de Likert (1932),¹⁷ dicho procedimiento exige una respuesta graduada a cada manifestación y se expresa frecuentemente en función de cinco categorías (categorías ordenadas), que van de “nunca”, “rara vez”, “a veces”, “frecuentemente” y “siempre”; y se califican, siguiendo el orden anterior: 1, 2, 3, 4 y 5, esto permite diferenciar claramente las manifestaciones favorables de las desfavorables. Además, las respuestas son mutuamente excluyentes y exhaustivas. Un criterio de clasificación está basado en la suma de las calificaciones de los elementos, esto representa la puntuación total del individuo, y una vez que se tiene un punto de corte, se compara esta puntuación total para determinar si la persona encuestada se encuentra en riesgo.¹⁶

El problema de clasificación de individuos con riesgo

Diversos estudios muestran que los puntos de corte utilizados para clasificar individuos -sanos, y enfermos-, en investigaciones epidemiológicas no son fijos. La elección de un punto de corte depende del objetivo del estudio. De manera que un punto de corte relativamente alto minimiza el número de falsos positivos, pero puede incrementar el número de falsos negativos. Además, es importante interpretar las puntuaciones elevadas o bajas de los diversos instrumentos, dentro del contexto del individuo. Esto es, si se encuesta a una persona que tiene una profesión que demanda estar delgado(a), el problema consistirá en la interpretación adecuada de este fenómeno. Otra situación problemática es cuando existen personas con TCA, con puntuaciones bajas, en los instrumentos.¹⁸ La determinación de un punto de corte es un reto para el investigador, ya que deberá sugerir una clasificación que permita medir lo que realmente desea, sin embargo, esto no siempre se puede cumplir ya que depende de varios factores, como la experiencia del investigador en el tema, cambios de nomenclatura del diagnóstico de los TCA, de la metodología empleada, del tipo de muestra con la que se cuenta, así como del cuestionario que se utilizó;¹⁹ todo esto hace difícil el establecimiento de un punto que defina claramente, lo normal y lo patológico.

Los puntos de corte establecidos en algunos cuestionarios, que indagan trastornos de la conducta alimentaria y que han sido validados en poblaciones mexicanas, son:

- Para detectar bulimia nerviosa en poblaciones estudiantiles, se ha empleado el auto reporte llamado BULIT (Bulimia Test), de Smith y Thelen, creado en 1984,²⁰ cuestionario de 36 preguntas, con una escala de respuesta de cinco opciones. En la validación de este instrumento, presentada por Álvarez y colaboradores,²¹ el punto de corte de este instrumento fue de ≥ 85 . Los autores reportan algunas diferencias en el criterio de diagnóstico médico.

- En el caso de la anorexia nerviosa, se emplea el cuestionario de auto reporte, que contiene 40 preguntas, conocido como EAT-40 (Eating Attitudes Test), de Garner & Garfinkel.²² La validación en poblaciones mexicanas, reporta diferencias en los diagnósticos clínicos, debido a la utilización de distintos manuales diagnósticos. El punto de corte fue de ≥ 30 en estudiantes.²³

- En lo que se refiere a la detección de síndromes parciales, la clasificación de individuos con riesgo se realiza por medio de los cuestionarios antes citados, donde las puntuaciones son “sensiblemente menores a las de los TCA (síndrome completo), pero que caen en un rango patológico”.²⁴

Otra manera de identificar a los individuos con riesgo en síndromes parciales, es por medio de nuevos cuestionarios, donde las conductas anómalas se contemplan como indicadores clínicos, -clasificación cualitativa-, de tal manera, que tres o más indicadores, con una escala de respuesta de “frecuentemente” y “siempre”, (elevada frecuencia) indica un trastorno de la conducta alimentaria o síndrome parcial, clasificación que se analiza en el presente trabajo.⁸

En una investigación española, se llevó a cabo un análisis alternativo, donde las conductas alimentarias de riesgo se revisan por separado, asociándolas a alguna variable de interés, en este caso, con el ejercicio físico.²⁵

Por último, en un estudio efectuado en Australia, se buscó una alternativa de clasificación de personas que no cubrían todas las características de bulimia nerviosa (casos subclínicos), debido a que se suscitó una controversia en el diagnóstico, se empleó un análisis de conglomerados jerárquico, para encontrar una clasificación alternativa. Esta investigación se efectuó con una muestra representativa de mujeres jóvenes (250) que recurría a atracones, las cuales se siguieron durante un año.

Primero se realizó un análisis de componentes principales, para representar con menos dimensiones el conjunto de datos, después se efectuó el análisis de conglomerados jerárquico por el método de Ward ($kappa=0.70$). Se realizó la validez predictiva de la solución dada por los conglomerados o grupos, así como los esquemas de diagnóstico efectuadas por medio del DSM-IV.⁶ Se concluye que los grupos formados por medio del análisis de conglomerados, permite identificar grupos clínicos con similares características. Además se obtuvo una validez predictiva muy buena, mediante el seguimiento de las pacientes. Esta validez fue superior en la solución dada por los conglomerados que en el efectuado por medio del diagnóstico clínico.²⁶

No se cuenta con más investigaciones que utilicen análisis de conglomerados, esto puede deberse, entre otros motivos, a que los cuestionarios que indagan estos trastornos psiquiátricos han sido cuantificados por medio de respuestas categóricas.

En este trabajo el interés radica en encontrar individuos con riesgo, mediante las respuestas generadas de un cuestionario autoaplicado que contiene una batería de preguntas referentes a conductas y actitudes alimentarias de riesgo, con respuesta de tipo Likert (escala categórica), donde se formen grupos homogéneos y que no se sobrepongan.²⁷ El análisis de conglomerados permite maximizar la homogeneidad de los objetos dentro del conglomerado y maximizar la heterogeneidad entre los grupos.²⁸

Objetivo

El presente estudio tiene por objetivo, clasificar a individuos por medio de su patrón de respuestas a conductas alimentarias de riesgo, aplicando la técnica multivariada de análisis de conglomerados de k-medias, en una muestra de estudiantes adolescentes y adultos jóvenes del estado de Morelos, México.

Metodología

Población y muestra

El presente trabajo deriva de la segunda evaluación del estudio epidemiológico “Estilos de vida en adolescentes y adultos jóvenes y su relación con el desarrollo longitudinal de enfermedades crónicas”, realizado en el estado de Morelos con 2,584 adolescentes y adultos jóvenes. El marco muestral se conformó con las escuelas públicas de nivel medio, medio superior y superior del ciclo escolar de 2001-2002. La unidad de muestreo fueron las escuelas públicas, los centros educativos corresponden a las 72 zonas administrativas de los 33 municipios del estado de Morelos, pertenecientes a las áreas urbanas, semiurbano y rurales. El muestreo en la fase basal fue un muestreo sistemático por conteo de seis en seis, de tal manera que se seleccionó una escuela de cada seis en el listado de escuelas, las cuales se ordenaron por nombre de municipio, por clave del IEBEM (Instituto de Educación Básica del Estado de Morelos) y por nombre de la escuela.

Posteriormente, se realizó una reconsideración debida a que no había la cantidad suficiente de escuelas con una matrícula completa, las escuelas pequeñas se encontraban a grandes distancias y en zonas poco accesibles y esto podía derivar en una baja respuesta al proyecto.

Por lo anterior, se realizó una reelección de las escuelas, las cuales deberían tener las siguientes características:

No se cuenta con más investigaciones que utilicen análisis de conglomerados, esto puede deberse, entre otros motivos, a que los cuestionarios que indagan estos trastornos psiquiátricos han sido cuantificados por medio de respuestas categóricas.

En este trabajo el interés radica en encontrar individuos con riesgo, mediante las respuestas generadas de un cuestionario autoaplicado que contiene una batería de preguntas referentes a conductas y actitudes alimentarias de riesgo, con respuesta de tipo Likert (escala categórica), donde se formen grupos homogéneos y que no se sobrepongan.²⁷ El análisis de conglomerados permite maximizar la homogeneidad de los objetos dentro del conglomerado y maximizar la heterogeneidad entre los grupos.²⁸

Objetivo

El presente estudio tiene por objetivo, clasificar a individuos por medio de su patrón de respuestas a conductas alimentarias de riesgo, aplicando la técnica multivariada de análisis de conglomerados de k-medias, en una muestra de estudiantes adolescentes y adultos jóvenes del estado de Morelos, México.

Metodología

Población y muestra

El presente trabajo deriva de la segunda evaluación del estudio epidemiológico “Estilos de vida en adolescentes y adultos jóvenes y su relación con el desarrollo longitudinal de enfermedades crónicas”, realizado en el estado de Morelos con 2,584 adolescentes y adultos jóvenes. El marco muestral se conformó con las escuelas públicas de nivel medio, medio superior y superior del ciclo escolar de 2001-2002. La unidad de muestreo fueron las escuelas públicas, los centros educativos corresponden a las 72 zonas administrativas de los 33 municipios del estado de Morelos, pertenecientes a las áreas urbanas, semiurbano y rurales. El muestreo en la fase basal fue un muestreo sistemático por conteo de seis en seis, de tal manera que se seleccionó una escuela de cada seis en el listado de escuelas, las cuales se ordenaron por nombre de municipio, por clave del IEBEM (Instituto de Educación Básica del Estado de Morelos) y por nombre de la escuela.

Posteriormente, se realizó una reconsideración debida a que no había la cantidad suficiente de escuelas con una matrícula completa, las escuelas pequeñas se encontraban a grandes distancias y en zonas poco accesibles y esto podía derivar en una baja respuesta al proyecto.

Por lo anterior, se realizó una reelección de las escuelas, las cuales deberían tener las siguientes características:

- a) Escuela seleccionada con matrícula igual o mayor a 100 estudiantes.
- b) Que se seleccionara al menos una escuela por municipio.

Además de lo anterior por cuestiones logísticas del proyecto se consideró una mayor población femenina. Para esta segunda evaluación se tienen a los individuos que permanecieron en las escuelas seleccionadas para la primera muestra y que aceptaron nuevamente participar en el estudio.

Los estudiantes contestaron en forma escrita un cuestionario autoaplicable en sus escuelas donde se les pidió información socioeconómica, de conductas alimentarias de riesgo y percepción de la figura corporal. Además, se tomaron las mediciones antropométricas, de peso, talla, circunferencias y pliegues; así como análisis de sangre de donde se obtuvieron cuantificación de triglicéridos y glucosa. Cabe señalar que la obtención de la muestra sanguínea se realizaba después de contestado el cuestionario.

En el caso de estudiantes menores de edad, este estudio posee consentimiento informado por los padres de familia o tutores. En este estudio se cuenta con carta avalada por la comisión de ética y por la comisión de bioseguridad del Instituto Nacional de Salud Pública.²⁹

La muestra esta conformada por un 60.8% de mujeres y un 39.2% de hombres, de los cuales el 54.64% tienen edades entre 14 a 16 años; además el 15.3% de ambos géneros sobreestima su imagen corporal. Los jóvenes que poseen sobrepeso u obesidad son el 24%. Existe un importante porcentaje de adolescentes que comen mientras ven televisión 46.13%.

Se ha comprobado que la obesidad y la sobreestimación de la imagen corporal son dos factores que tienen una importante participación en el desarrollo de los trastornos de la conducta alimentaria.³⁰

La variable llamada *conductas alimentarias de riesgo* se obtuvo de un cuestionario de 13 preguntas (anexo) que enmarcan conductas y actitudes anormales en la alimentación (α de Cronbach de 0.54, ver anexo),³¹ y son:

- 1.- Si has estado a dieta ¿qué tan seguido?
- 2.- Durante los últimos 30 días ¿vomitaste o tomaste laxante para perder o no aumentar de peso?
- 3.- Durante los últimos 30 días ¿tomaste pastillas para perder o no aumentar de peso?
- 4.- Me aterroriza tener sobrepeso.
- 5.- Estoy preocupada por tratar de ser delgada.
- 6.- Durante los últimos 30 días ¿hiciste ejercicio para perder o no aumentar de peso?
- 7.- Siento que la comida controla mi vida.
- 8.- Preparo mis alimentos pero no los consumo.
- 9.- Corto mis alimentos en pedazos muy chicos.
- 10.- Guardo pequeños trozos de alimentos y los voy comiendo poco a poco.
- 11.- Siento que otros preferirían que yo comiera más.
- 12.- Disfruto comiendo carne.
- 13.- Disfruto comiendo en restaurantes

De estas preguntas, nueve de ellas derivan del cuestionario EAT-40 de Garner y Garfinkel (1979),²² con la modificación de que estas aseveraciones poseen una escala de respuesta tipo Likert de cinco opciones:

- 1).- "Nunca"
- 2).- "Rara vez"
- 3).- "Algunas veces"
- 4).- "Con frecuencia"
- 5).- "Siempre"

Las cuatro preguntas que restan (1, 2, 3 y 6), fueron elaborados para el estudio de adolescentes y poseen una escala de respuesta de seis opciones, que van de:

- 1).- "Nunca"
- 2).- "Una vez en el último mes"
- 3).- "Dos a tres veces en el último mes"
- 4).- "Una vez a la semana"
- 5).- "De cuatro a seis veces a la semana"
- 6).- "Diario"

Debido a que los dos patrones de respuesta son equivalentes, para homogenizar la escala se unió los incisos 2 y 3 de la escala de seis opciones, de tal manera que se obtiene para todas las preguntas cinco opciones de respuesta. Por último se cuida el sentido de las respuestas, es decir, que vayan de una menor a una mayor frecuencia. En dos preguntas, "disfruto comiendo carne" y "disfruto comiendo en restaurantes", este sentido es contrario, esto es, las personas que poseen TCA, por lo regular no desean comer alimentos que deban masticar más tiempo y no les agrada estar en compañía de otras personas en el momento de ingerir sus alimentos, por ello, el sentido, debe ser: 1)"siempre", 2)"con frecuencia", 3)"algunas veces", 4)"Rara vez" y 5)"nunca"; ya que las calificaciones altas permitirán agrupar a aquellas personas que tienen riesgo a los TCA.

Posteriormente se procede a analizar las variables mediante el método multivariado de conglomerados de k-medias (sin estandarizar variables), figura 1, este análisis permitió cubrir los siguientes objetivos:

- o Mediante el diagrama de perfil de las trece variables, se eligieron las variables que diferenciaron los grupos de riesgo y sin riesgo.
- o Desarrollo de una tipología o clasificación de acuerdo a la escala de respuesta, la primera con datos con escala de respuesta de cinco categorías ordenadas y la segunda con respuesta dicotómica o binaria.

*Distinción entre la escala ordinal y nominal*³²

Existen dos tipos de escalas de mediciones para las variables categóricas. Muchas escalas categóricas tienen orden natural, por ejemplo, respuesta a un medicamento (excelente, bueno y malo), las variables categóricas que tienen una escala ordenada se les llama variables ordinales. Las variables categóricas que no tienen escalas ordenadas se les llama variables nominales, ejemplo, género (masculino y femenino), tipo de música que se escucha (clásica, rock, jazz, country).

Para la elaboración de los conglomerados, en las variables con escala de respuesta ordinal, no se requiere ninguna modificación, ya que finalmente, este tipo de escala puede ser tratada como puntuaciones de una escala continua.¹¹

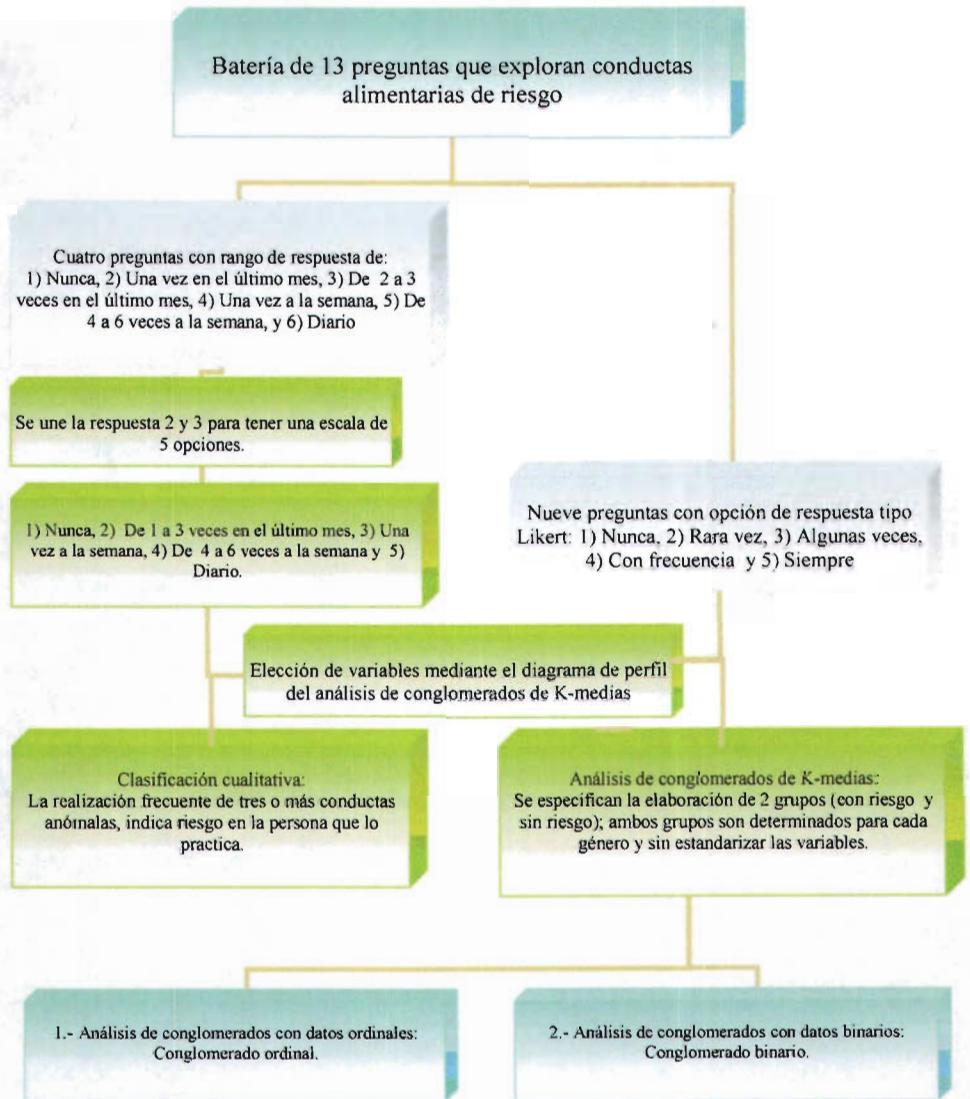
Sin embargo, para obtener los conglomerados con datos dicotómicos (0 y 1), se modificó la escala de Likert, de tal forma que para las respuestas, que significan: “nunca”, “rara vez” y “a veces”, o bien, “una vez en el último mes” y “dos a tres veces en el último mes”, tengan valor 0. En contraparte, el valor de las respuestas indicativas de: “con frecuencia”, “siempre”, “de cuatro a seis veces a la semana” y “diario” poseerán el valor de 1.

Para hacer la diferencia, entre estos dos conglomerados, con datos ordinales y con datos binarios, la denominación que se empleará en adelante, será la de conglomerado ordinal y conglomerado binario.

Finalmente, se comparan con una clasificación cualitativa que ha sido empleada por especialistas en el tema,⁸ y comenta que una persona que tiene riesgo, es aquella que realiza frecuentemente tres o más conductas anómalas.

Los conglomerados o agrupaciones efectuados, se graficaron mediante los dos componentes principales en base a la matriz de covarianzas, ya que son las mismas unidades de medición en todas las variables.

Figura 1. Análisis de las conductas alimentarias de riesgo por medio del análisis de conglomerados no jerárquico de k- medias.



Análisis estadístico

Análisis multivariado de conglomerados ²⁸

El análisis de conglomerados agrupa a los individuos y a los objetos en conglomerados, de tal forma que los objetos del mismo conglomerado son más parecidos entre sí que a los objetos de otros conglomerados. Lo que intenta es maximizar la homogeneidad de los objetos dentro de los conglomerados mientras que a la vez se maximiza la heterogeneidad entre los grupos.

El objetivo de cualquier agrupación es de importancia crucial. La pregunta natural a responder antes de formar agrupaciones es ¿qué tipo de similitud o disimilitud deseamos buscar? La manera de calcular la similitud entre dos objetos depende del tipo de variables originales.

El análisis de conglomerados, por lo tanto, intenta identificar las vectores de observaciones que son similares y agruparlos dentro de grupos, muchas técnicas usan un índice de similitud o de proximidad entre cada par de observaciones.

Tipos de datos y cómo manejarlos

Generalmente los algoritmos en el análisis de conglomerados operan en dos estructuras, la primera representa los n sujetos por los p variables (conductas anómalas). Estos datos pueden estar arreglados en una matriz de $n \times p$, donde los renglones corresponden a los sujetos y las columnas a los atributos. Cuando la f -ésima medición del i -ésimo individuo, es denotado por x_{if} (donde $i=1, \dots, n$ y $f=1, \dots, p$) la matriz es:

$$\begin{array}{l} \text{p variables o atributos} \\ \left[\begin{array}{cccc} x_{11} & \cdots & x_{1f} & \cdots x_{1p} \\ \vdots & & \vdots & \vdots \\ x_{i1} & \cdots & x_{if} & \cdots x_{ip} \\ \vdots & & \vdots & \vdots \\ x_{ni} & \cdots & x_{nf} & \cdots x_{np} \end{array} \right] \\ \text{n objetos} \end{array}$$

La segunda estructura es la colección de las proximidades de todos los pares de objetos. Estas proximidades deben estar en una tabla de $n \times n$. Se consideran dos tipos de proximidades, una llamada disimilitud, la cual mide qué tan lejos están dos objetos uno del otro; y la similitud, la cual cuantifica qué tan parecidos son dichos objetos.

En las variables de escala de intervalo, se necesitan calcular las distancias entre los objetos, para cuantificar su grado de desemejanza. Para ello es necesario tener distancias para cada par de objetos i y j . La elección más popular es la *distancia euclidiana*, definida como la distancia

euclidiana entre dos puntos $x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}, x_j = \begin{pmatrix} x_{j1} \\ \vdots \\ x_{jp} \end{pmatrix}$

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2}$$

La fórmula anterior corresponde a la distancia geométrica real entre los puntos con coordenadas x_{i1}, \dots, x_{ip} y x_{j1}, \dots, x_{jp} .

Otra distancia muy utilizada es la *distancia Manhattan* o *city block*, definida como

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

Una generalización de la *distancia euclidiana* y *Manhattan* es la distancia *Minkowski*,

$$d(i, j) = \left(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q \right)^{1/q}$$

Donde q es cualquier número real igual o mayor a 1.

Estas distancias satisfacen los siguientes requerimientos matemáticos que describen una función de distancia: ³⁴

$$D1. -d(i, j) \geq 0,$$

$$D2. -d(i, i) = 0$$

$$D3. -d(i, j) = d(j, i)$$

$$D4. -d(i, j) \leq d(i, h) + d(h, j)$$

Para todos los objetos i, j y h . La condición D1 simplemente indica que las distancias son números no negativos; la D2 indica que la distancia de un objeto a si mismo es igual a cero. El axioma D3 indica la simetría de la función de distancia. La desigualdad del triángulo, D4, indica que ir directamente de i a j es más corto que hacer una desviación hacia el objeto h .

Cabe hacer notar que $d(i, j) = 0$, no necesariamente implica que $i=j$, ya que puede suceder que dos diferentes objetos posean distancia como las variables bajo estudio.

La matriz de la distancia es siempre simétrica, y las entradas de la diagonal son ceros.

Disimilitud

La medición de la disimilitud, se puede obtener de varias maneras dependiendo del tipo de variable. La matriz $n \times n$ es:

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & 0 & \end{bmatrix}$$

Donde $d(i,j) = d(j,i)$ cuantifica la diferencia o disimilitud entre los objetos i y j . Esta diferencia satisface los axiomas D1, D2 y D3, aunque ninguna de estas propiedades es realmente esencial, ya que existen métodos de conglomerados que no requiere de algunos de ellos. La disimilitud se puede obtener de varias maneras, en variables binarias, nominales, de intervalo, etc. Así mismo, estas disimilitudes también pueden ser rangos subjetivos de que tan diferente es un objeto de otro, desde el punto de vista de uno o más observadores.

Similitud

Usando el coeficiente de disimilitud $d(i,j)$, es posible trabajar con un coeficiente de similitud $s(i,j)$. Este coeficiente nos permite indicar qué tan cerca está un objeto o sujeto de otro. Frecuentemente se asumen las siguientes condiciones:

- S1. - $0 \leq s(i,j) \leq 1$
- S2. - $s(i,i) = 1$
- S3. - $s(i,j) = s(j,i)$

Para todos los sujetos i y j . Los números $s(i,j)$ pueden ser arreglos en la matriz $n \times n$, la cual es llamada como matriz de similitud. Las matrices de similitudes y disimilitudes son referidas como matrices de proximidad. La similitud puede calcularse de varias maneras, una de ellas es mediante la caracterización de los atributos. Supóngase que los datos consisten de una matriz de similitud, pero que se desea aplicar los algoritmos de conglomerados a la matriz de disimilitud. Entonces es necesario transformar los datos. Un valor grande de similitud $s(i,j)$ entre i y j , será un valor pequeño en la disimilitud $d(i,j)$. Esta transformación es:

$$d(i,j) = 1 - s(i,j)$$

Las variables binarias tienen solo dos posibles resultados o estados. Cuando la variable f es binaria, los objetos i pueden tener los valores $x_{if} = 0$ ó $x_{if} = 1$. Frecuentemente el valor de 1 es indicativo de que cierto atributo está presente, el caso contrario, 0 es que está ausente ese atributo.

Objeto i	Objeto j		Total
	1	0	
1	a	b	$a+b$
0	c	d	$c+d$
Total	$a+c$	$b+d$	p

Donde a es el número de variables que son igual a 1 en ambos sujetos. La casilla b es el número de variables f la cual $x_{if} = 1$ y $x_{jf} = 0$; la casilla c es el número de variables f la cual $x_{if} = 0$ y $x_{jf} = 1$. La d es el número de variables que son igual a cero en ambos sujetos. El total de número de variables es $a+b+c+d=p$.

En algunas ocasiones en las variables binarias no importa la codificación o el peso de 1 ó 0, ya que ambos valores son iguales, aquí no hay preferencia por uno o por otro valor, estas variables se llaman variables *binarias simétricas*.

La situación cambia drásticamente, cuando se trabaja con variables *binarias asimétricas*, como en el presente trabajo, en las cuales los dos resultados no son igualmente importantes. En estas variables se debe codificar al resultado más importante como 1 y el otro como 0. Entonces el tener dos unos (1-1), llamado también como puntuación positiva, será considerado como el más significativo que los dos ceros (0-0) o puntuación negativa. Por lo tanto, los coeficientes deben ser aplicados en la casilla a , que es el número de puntuaciones positivas, lugar con más peso que en la casilla d , lugar con puntuaciones negativas. Los coeficientes calculados en estas variables no toman en cuenta a la casilla d .

El coeficiente más utilizado es el de Jaccard (1908),³⁵ también se le llama *coeficiente-S*. Aquí se le asigna doble peso a la casilla a ó $b+c$.

$$S_{rs} = \frac{a}{a+b+c}, \text{ o bien}$$

$$d_{rs} = 1 - S_{rs} = \frac{b+c}{a+b+c}, \text{ que es el coeficiente de disimilitud.}$$

Finalmente, y como conclusión las variables discretas que poseen categorías ordenadas como la escala de Likert, se les puede tratar como variable de escala de intervalo;³⁰ o bien, se pueden analizar como variables binarias o dicotómicas.¹¹

Los dos enfoques que se efectúan en el análisis de conglomerados para agrupar los vectores de observación son:

A).- El agrupamiento jerárquico: Aquí se encuentran los métodos aglomerativos y métodos divisibles. En los métodos aglomerativos se empieza con n grupos, uno por cada observación y se finaliza con un solo agrupamiento conteniendo todas las observaciones. A cada paso una observación o un grupo de observaciones son contenidos dentro de otro agrupamiento. Se puede revertir este proceso, esto es, empezar con una sola agrupación conteniendo todas las observaciones y finalizar con “ n ” grupos, como sucede en los métodos divisibles (anexo).

B).- Agrupación no jerárquico: En este apartado se localizan los métodos de optimización. Simplemente se dividen las observaciones en g grupos. Éste puede estar determinado por una k división inicial o por centros de grupos, donde se corre el algoritmo para un rango de k valores. El método que se revisa en este trabajo es el de k -medias.

Métodos de optimización³⁶

Las técnicas de optimización difieren de los métodos jerárquicos en que los datos no necesariamente forman jerarquías. La idea básica detrás de este método, es que se encuentra el índice, $f(n,g)$ asociado a cada partición de n -individuos, para la formación de g grupos, el cual, es indicativo de la “calidad de agrupación o conglomerado”. Muchos criterios de agrupación han sido sugeridos, sin embargo los más usados se originan de las siguientes tres matrices.

$$T = \frac{1}{n} \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})'$$

$$W = \frac{1}{n-g} \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

$$B = \sum_{i=1}^g n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

Donde g es el número de grupos, x_{ij} es el vector de los valores de la variable para la j -ésima observación en el i -ésimo grupo, la \bar{x} es la media del vector de todas las n observaciones, \bar{x}_i es el vector de la media de las observaciones en el grupo i y n_i es el número de observaciones en el grupo i .

Estas tres matrices $p \times p$ (donde p es el número de variables) representan respectivamente el total de la dispersión, la dispersión dentro del grupo (W) y la dispersión entre los grupos (B), y donde se satisface la siguiente ecuación:

$$T = W + B \quad \dots\dots\dots \quad \text{©}$$

Para $p=1$, esta ecuación representa la relación entre números escalares; simplemente es la división del total de la suma de cuadrados para una variable, dentro y entre la suma de cuadrados de los grupos, esto es, algo similar al análisis de varianza de una vía.

En este caso un criterio natural para agrupar, puede ser elegir la partición correspondiente al mínimo valor de la suma de cuadrados dentro del grupo o equivalentemente el máximo valor entre los grupos.

Para $p > 1$, el criterio de agrupamiento de la ecuación © no posee un corte claro, por ello se sugieren varias alternativas.

Minimización de la traza (W)

Una extensión obvia de la minimización de la suma de cuadrados, dentro del grupo, sugiere que para el caso $p=1$, cuando los datos no son univariados, es la minimización de la suma de cuadrados dentro de los grupos, donde se abarcan a todas las variables, esto es, minimizar la traza (W). Esto es equivalente a minimizar la suma de las distancias *euclidianas* al cuadrado entre individuos y entre las medias de los conglomerados, por ello

$$E = \sum_{i=1}^n d_{i,c(i)}^2$$

Donde $d_{i,c(i)}$ es la distancia *euclidiana* del individuo i , a la media del conglomerado al cual es asignado. Minimizar la traza W es equivalente a maximizar la traza (B). Este criterio esta explícito en el método de k -medias y en el método de Ward (método jerárquico).

Minimización del determinante de (W)

En un análisis multivariado de varianza, una de las pruebas para ver las diferencias en vectores de medias grupales, esta basado en la relación de determinantes de las matrices de dispersión dentro y en el total de los grupos. De tal manera que grandes valores de $det(T)/det(W)$, indican que la media de los vectores de los grupos son diferentes. Por ello uno de los criterios es maximizar esta relación.

Todas las particiones de n individuos en g grupos, T quedan igual, la maximización de $det(T)/det(W)$ es equivalente a la minimización de $det(W)$.

Maximización de la traza (BW^{-1})

Otro criterio es la maximización de la traza de la matriz obtenida del producto de la matriz de dispersión entre los grupos y la inversa de la matriz de dispersión dentro de los grupos. Esta función es usada también en el contexto del análisis multivariado de varianza y es equivalente a lo que se le llama la generalización de la distancia de *Mahalanobis* en más de dos grupos. La matriz BW^{-1} , también se encuentra en el análisis de varianza canónica.

Ambos traza (BW^{-1}) y $\det(T)/\det(W)$ se pueden expresar en términos de eigenvalores, λ_i , de BW^{-1} , esto es

$$\text{traza}(BW^{-1}) = \sum_{i=1}^p \lambda_i$$

$$\frac{\det(T)}{\det(W)} = \prod_{i=1}^p (1 + \lambda_i)$$

Optimización de los criterios de agrupamiento

Una vez que el criterio de agrupamiento numérico se ha seleccionado, es necesario ahora saber cómo elegir la partición de g grupos de los datos la cual permita su optimización.

Desafortunadamente el problema en la práctica no ha sido simple. Algunos ejemplos ilustran el problema:

$$N(15,3)=2,375,101$$

$$N(20,4)=45,232,115,901$$

$$N(100,5)=10^{68}$$

Donde $N(n,g)$ es el número de las distintas particiones de n individuos dentro de g grupos vacíos. Una expresión general para $N(n,g)$ es

$$N(n, g) = \frac{1}{g!} \sum_{i=0}^g (-1)^{g-i} \binom{g}{i} i^n$$

Lo impráctico que resulta cada posible partición, nos indica el desarrollo de algoritmos diseñados para investigar un valor óptimo de un criterio de conglomerado y reordenar las particiones existentes y elegir una nueva que permita una mejor solución, estos son los llamados algoritmos “cuesta arriba”, aunque en el caso de los criterios se requiere de la minimización son llamados “cuesta abajo”, los pasos esenciales de estos algoritmos son:

- Encontrar una partición inicial de los individuos dentro de un número requerido de grupos.
- Calcular el cambio en el criterio de agrupamiento producido en cada movimiento individual de un grupo a otro.
- Realizar el cambio para mejorar el valor del criterio de conglomerado.
- Repetir el paso (b) y (c) hasta que se tenga la mejor solución de grupos.

Una configuración del conglomerado inicial, se puede especificar en base a un conocimiento previo del problema. Otra posibilidad puede ser elegir aleatoriamente dicha partición.

Método K medias (MacQueen, 1967):³⁷

Primero se seleccionan g observaciones para servir como “semillas”. Éstas son después reemplazadas por los centroides (vectores medios) de los grupos. Hay varias maneras que uno puede escoger los centros: seleccionar g observaciones aleatoriamente (puede ser separados por una distancia mínima especificada), elegir los primeros g puntos en el grupo de datos (nuevamente sujetos a una distancia mínima requerida), seleccionar los g puntos que son mutuamente más lejanos, encontrar los g puntos de máxima densidad, o especificar los g puntos espaciados con regularidad en un patrón de red (éstos no pueden ser puntos de datos actuales).

Para estos métodos de elección de semillas, el número de grupos, g , debe ser especificado previamente. Alternativamente una mínima distancia entre semillas puede ser especificada y entonces todos las observaciones que satisfacen este criterio son elegidos como semillas.

Después de que las semillas son seleccionadas, cada punto que sobra en el grupo de datos es asignado al conglomerado con semilla más cercana, basado en la distancia Euclidiana. Tan pronto como un conglomerado tiene más de un miembro, la semilla de dicho conglomerado es reemplazada por el centroide.

Después todas las observaciones son asignadas a los grupos, cada observación es examinada para ver a qué centroide de otro grupo le queda más cercano que el centro de su propio grupo. De esta manera la observación se cambia para un nuevo grupo y los dos centroides de los grupos son nuevamente calculados. Este proceso se continúa hasta que todas las observaciones están acomodadas. El procedimiento de k medias es muy sensible a la selección inicial de semillas.³⁸

Elaboración de conglomerados de acuerdo al paquete estadístico JMP y Stata

Para la obtención del *conglomerado ordinal*, se utilizó el paquete estadístico JMP 5.0,³¹ el procedimiento del método de k -medias es el siguiente:

Este método primero selecciona un grupo de n puntos llamados *semillas de conglomerados*, como una suposición de los promedios de los conglomerados. Cada observación es asignada a la semilla más cercana que forma un grupo de conglomerados temporales. Las semillas son entonces reemplazadas, por los promedios de los conglomerados, los puntos son reasignados, el proceso continúa hasta que no ocurren cambios en los grupos. Cuando este proceso finaliza, se muestran tablas que sintetizan las características de los grupos formados. Este método se suele utilizar con datos que van de 200 a 100,000 observaciones. Con pocos datos, los resultados pueden ser muy sensibles al orden de las observaciones de la base de datos.

En el caso de la elaboración del *conglomerado binario*, se utilizó el paquete estadístico Stata 7.0.¹⁴ Este paquete hace lo siguiente:

La elaboración para k -medias es un procedimiento iterativo que parte los datos en k grupos o centros. Las observaciones son asignadas al grupo con el centro más cercano. Se calcula la media de las observaciones para cada grupo y así consecutivamente. Estos pasos continúan hasta que todas las observaciones se encuentran en los grupos más compatibles.

La media para un grupo de variables binarias, es la proporción de unos que posee ese grupo de observaciones. Se puede ajustar la medición de la similitud a una media binaria (proporción).

En datos binarios, el promedio de grupo es interpretado como una proporción. En el presente trabajo se utiliza la similitud calculada mediante el coeficiente de Jaccard.

Resultados

Los resultados que se obtienen corresponden a los siguientes análisis estadísticos:

1. Análisis descriptivo de los datos, mediante el paquete Stata 7.0
2. Empleo de análisis de conglomerados de k-medias. Este apartado comprende la elección de las variables y la formación de grupos de riesgo y no riesgo.
3. Comparación de la clasificación cualitativa contra el conglomerado ordinal y binario. En este apartado se calcula el índice de validez, mediante el cálculo de sensibilidad y especificidad, así como el cálculo de confiabilidad por medio del valor de concordancia de kappa.¹²

1.- Análisis descriptivo

En el análisis descriptivo correspondiente a las trece variables, tabla I, se observa que la mayoría de individuos, poseen frecuencias nulas o bajas en estas conductas y actitudes anómalas. Se aprecia además que la variable de consumo de pastillas presenta solo dos respuestas, la de nunca y una vez en el último mes. La escala de respuesta de la variable de vómito autoinducido, presenta una muy baja proporción de individuos que respondieron en las opciones 3 y 4 (“dos a tres veces en el último mes” y “una vez a la semana”).

Tabla I. Descripción de cada variable

a).- Si has estado a dieta ¿Qué tan seguido?

dieta	Freq.	Percent	Cum.
1	2463	95.32	95.32
3	3	0.12	95.43
4	103	3.99	99.42
5	15	0.58	100.00
Total	2584	100.00	

En datos binarios, el promedio de grupo es interpretado como una proporción. En el presente trabajo se utiliza la similitud calculada mediante el coeficiente de Jaccard.

Resultados

Los resultados que se obtienen corresponden a los siguientes análisis estadísticos:

1. Análisis descriptivo de los datos, mediante el paquete Stata 7.0
2. Empleo de análisis de conglomerados de k-medias. Este apartado comprende la elección de las variables y la formación de grupos de riesgo y no riesgo.
3. Comparación de la clasificación cualitativa contra el conglomerado ordinal y binario. En este apartado se calcula el índice de validez, mediante el cálculo de sensibilidad y especificidad, así como el cálculo de confiabilidad por medio del valor de concordancia de kappa.¹²

1.- Análisis descriptivo

En el análisis descriptivo correspondiente a las trece variables, tabla I, se observa que la mayoría de individuos, poseen frecuencias nulas o bajas en estas conductas y actitudes anómalas. Se aprecia además que la variable de consumo de pastillas presenta solo dos respuestas, la de nunca y una vez en el último mes. La escala de respuesta de la variable de vómito autoinducido, presenta una muy baja proporción de individuos que respondieron en las opciones 3 y 4 (“dos a tres veces en el último mes” y “una vez a la semana”).

Tabla I. Descripción de cada variable

a).- Si has estado a dieta ¿Qué tan seguido?

dieta	Freq.	Percent	Cum.
1	2463	95.32	95.32
3	3	0.12	95.43
4	103	3.99	99.42
5	15	0.58	100.00
Total	2584	100.00	

b).- Durante los últimos 30 días ¿Vomitaste o tomaste laxante para perder o no aumentar de peso?

vom	Freq.	Percent	Cum.
1	2545	98.49	98.49
2	11	0.43	98.92
3	4	0.15	99.07
4	2	0.08	99.15
5	22	0.85	100.00
Total	2584	100.00	

c).- Durante los últimos 30 días ¿Tomaste pastillas para perder o no aumentar de peso?

pas	Freq.	Percent	Cum.
1	2569	99.42	99.42
2	15	0.58	100.00
Total	2584	100.00	

d).- Me aterroriza tener sobre peso.

	Freq.	Percent	Cum.
1	2346	90.79	90.79
2	123	4.76	95.55
3	26	1.01	96.56
4	74	2.86	99.42
5	15	0.58	100.00
Total	2584	100.00	

e).- Estoy preocupado por tratar de ser delgado.

	Freq.	Percent	Cum.
1	2354	91.10	91.10
2	118	4.57	95.67
3	43	1.66	97.33
4	57	2.21	99.54
5	12	0.46	100.00
Total	2584	100.00	

f).- Durante los últimos 30 días ¿Hiciste ejercicio para perder o no aumentar de peso?

.	Freq.	Percent	Cum.
1	653	25.27	25.27
2	1797	69.54	94.81
3	123	4.76	99.57
4	11	0.43	100.00
Total	2584	100.00	

g).- Siento que la comida controla mi vida.

	Freq.	Percent	Cum.
1	902	34.91	34.91
2	1484	57.43	92.34
3	136	5.26	97.60
4	62	2.40	100.00
Total	2584	100.00	

h).- Preparo mis alimentos pero no los consumo

	Freq.	Percent	Cum.
1	1349	52.21	52.21
2	963	37.27	89.47
3	210	8.13	97.60
4	51	1.97	99.57
5	11	0.43	100.00
Total	2584	100.00	

i).- Corto mis alimentos en pedazos muy chicos

	Freq.	Percent	Cum.
1	1086	42.03	42.03
2	1235	47.79	89.82
3	245	9.48	99.30
4	17	0.66	99.96
5	1	0.04	100.00
Total	2584	100.00	

j).- Guardo pequeños trozos de alimentos y los voy comiendo poco a poco.

	Freq.	Percent	Cum.
1	327	12.65	12.65
2	1483	57.39	70.05
3	699	27.05	97.10
4	72	2.79	99.88
5	3	0.12	100.00
Total	2584	100.00	

k).- Siento que otros preferirían que yo comiera más.

	Freq.	Percent	Cum.
1	854	33.05	33.05
2	928	35.91	68.96
3	577	22.33	91.29
4	222	8.59	99.88
5	3	0.12	100.00
Total	2584	100.00	

l).- Disfrute comiendo carne

carne	Freq.	Percent	Cum.
1	125	4.84	4.84
2	515	19.93	24.77
3	1179	45.63	70.39
4	749	28.99	99.38
5	16	0.62	100.00
Total	2584	100.00	

m).- Disfrute comiendo en restaurante

rest	Freq.	Percent	Cum.
1	6	0.23	0.23
2	149	5.77	6.00
3	1136	43.96	49.96
4	1235	47.79	97.76
5	58	2.24	100.00
Total	2584	100.00	

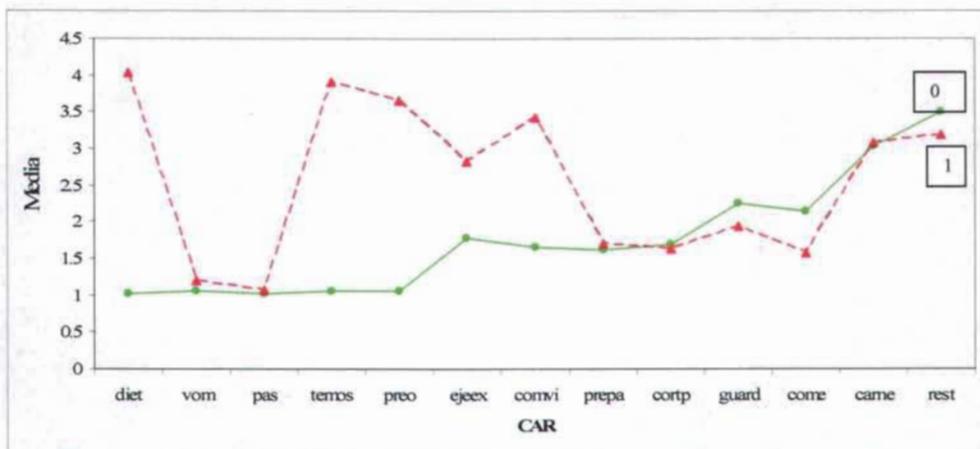
2.1.- Elección de variables

Con la finalidad de analizar únicamente las variables que diferencian ambos grupos, el de riesgo y de no riesgo, se realiza el análisis de conglomerados de k-medias, de tal manera que permita observar en el diagrama de perfil qué variables deberán de usarse para los conglomerados ordinal y binario.

En la figura 2 y en la figura 3, se observa que en ambos géneros el patrón de distribución de las 13 variables es muy similar, las variables dieta, temor al sobrepeso, preocupación por estar delgado, ejercicio extenuante, el control de la vida por la comida, diferencian perfectamente ambos grupos, -riesgo y no riesgo-. Las variables indicativas de vómito autoinducido, consumo de pastillas para bajar de peso, preparación de los alimentos y no consumirlos, cortar en pequeños pedazos los alimentos, guardar los alimentos y comerlos poco a poco, el sentimiento de creer que otros desearían que comieran mas, además del gusto por comer carne y gusto por comer en restaurantes, no permiten clasificar adecuadamente ambos grupos; al contrario, las últimas 4 variables invierten el grupo de riesgo.

Por lo tanto, en la determinación de los conglomerados, se utilizaran sólo cinco variables, dieta, temor al sobre peso, preocupación por estar delgado, ejercicio extenuante y la comida controla su vida, las cuales corresponden a conductas no purgativas (α de Cronbach de 0.884).

Figura 2. Valores promedio por grupo de riesgo y no riesgo de las 13 conductas anómalas en los hombres de la muestra del estado de Morelos, México 2001-2002.



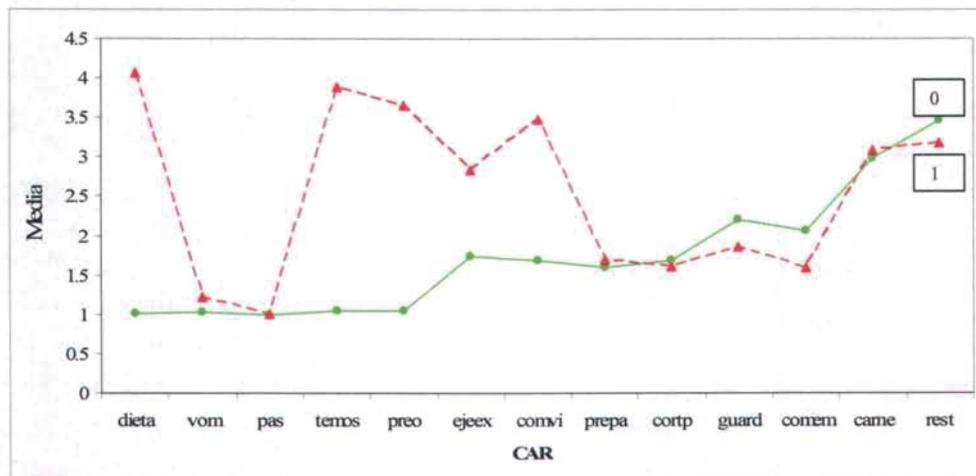
Cluster	Cantidad	Max Dist	Prior Dist
0	969	5.286	3.684
1	43	3.360	4.447
	1,012		

Prevalencia = $43/1,012 = 0.0424 = 4.24\%$

Medias de los grupos

Grupo	diet	vom	pas	temos	preo	ejeex	comvi	prepa	cortp	guard	come	carne	rest
0	1.016	1.041	1.005	1.054	1.049	1.763	1.646	1.609	1.683	2.242	2.137	3.025	3.49
1	4.023	1.186	1.069	3.906	3.651	2.813	3.418	1.697	1.627	1.93	1.581	3.069	3.186

Figura 3. Valores promedios por grupo de riesgo y no riesgo de las 13 conductas anómalas en las mujeres de la muestra del estado de Morelos, México 2001-2002.



Cluster	Cantidad	Max Dist	Prior Dist
0	1500	4.999	3.551
1	72	4.249	4.321
	1,572		

$$\text{Prevalencia} = 72/1,572 = 0.0458 = 5.6\%$$

Medias de los grupos

Grupo	dieta	vom	pas	temos	preo	ejeex	comvi	prepa	cortp	guard	comem	carne	rest
0	1.005	1.032	1.004	1.046	1.046	1.75	1.688	1.605	1.697	2.202	2.06	2.987	3.462
1	4.069	1.236	1.013	3.902	3.666	2.847	3.486	1.708	1.625	1.861	1.597	3.097	3.18

Clasificación cualitativa

La clasificación cualitativa, describe que la realización frecuente de tres o más conductas anómalas, indica riesgo en la persona que lo practica.⁸ Para reproducir esta metodología, y con el afán de comparar los conglomerados con esta clasificación, se analizan las cinco variables anteriormente descritas y se transforman a variables dicotómicas, los valores de 0 corresponden a aquellas respuestas, indicativas de baja frecuencia: "nunca", "rara vez" y "a veces". Los valores de alta frecuencia, que equivalen a valores 1, corresponden a las respuestas, "frecuentemente" y "siempre", tabla II.

Posteriormente se suman, y se obtiene una escala categórica, cuando el individuo, presenta mas de 3 conductas alimentarias en alta frecuencia, entonces, se encuentra en el grupo de riesgo, tabla III.

Tabla II. Frecuencia de las cinco CAR de manera dicotómica.

a) .- Si has estado a dieta ¿Qué tan seguido?

die2	Freq.	Percent	Cum.
0	2466	95.43	95.43
1	118	4.57	100.00
Total	2584	100.00	

b) .- Me aterroriza tener sobre peso.

tem2	Freq.	Percent	Cum.
0	2495	96.56	96.56
1	89	3.44	100.00
Total	2584	100.00	

c) .- Estoy preocupado por tratar de ser delgado

pre2	Freq.	Percent	Cum.
0	2515	97.33	97.33
1	69	2.67	100.00
Total	2584	100.00	

d) .- Durante los últimos 30 días ¿Hiciste ejercicio para perder o no aumentar de peso?

eje2	Freq.	Percent	Cum.
0	2573	99.57	99.57
1	11	0.43	100.00
Total	2584	100.00	

e).- Siento que la comida controla mi vida.

comv2	Freq.	Percent	Cum.
0	2522	97.60	97.60
1	62	2.40	100.00
Total	2584	100.00	

Tabla III. Suma de conductas

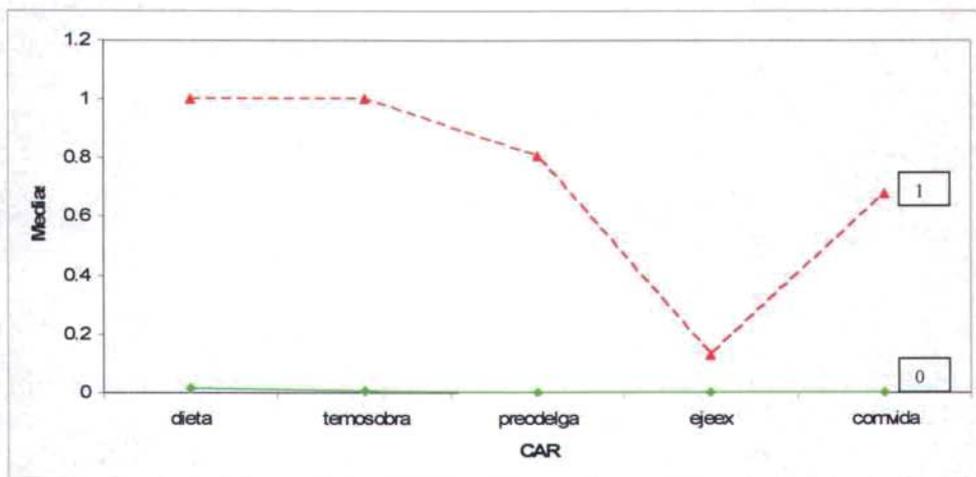
Clasificación cual. = die2+tem2+pre2+eje2+comv2

unikel	Freq.	Percent	Cum.
0	2442	94.50	94.50
1	51	1.97	96.48
2	9	0.35	96.83
3	32	1.24	98.07
4	42	1.63	99.69
5	8	0.31	100.00
Total	2584	100.00	

La prevalencia que se obtiene de individuos con riesgo mediante esta clasificación es de 3.24% en las mujeres y en los hombres de 3.06%; prevalencias muy semejantes, tablas de las figuras 4 y 5. La distribución de las conductas es igual en ambos géneros, donde la realización de ejercicio extenuante es la más baja, en el grupo de riesgo. Contrariamente la conducta que mejor diferencia ambos grupos es la variable de realización de distas para reducir peso corporal.

Se determinan componentes principales, con la finalidad de graficar los grupos formados, figura 6, los grupos formados en general, se dividen favorablemente, aunque algunos individuos que pertenecen al grupo de no riesgo se combinan con el grupo de riesgo, en los valores de 4.75 y 5.25 del primer componente principal.

Figura 4. Valores promedio por grupo de riesgo y no riesgo de las cinco conductas anómalas en los hombres de la muestra del estado de Morelos, México 2001-2002. Clasificación cualitativa.



Medias de los grupos

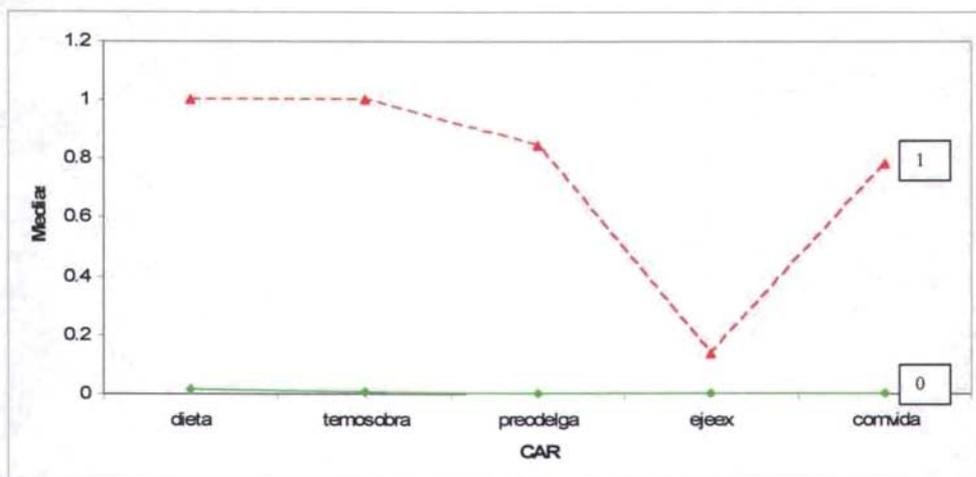
Grupos	Cantidad	dieta	temsobra	preodelga	ejeex	comvida
0	981	0.016	0.003	0	0	0
1	31	1	1	0.806	0.129	0.677
	1,012					

Prevalencia = $31/1,012 = 0.0306 = 3.06\%$

Desviaciones estándar de los grupos

Grupos	Dieta	temsobra	preodelga	ejeex	comvida
0	0.015	0.003	0	0	0
1	0	0	0.031	0.101	0.073

Figura 5. Valores promedios por grupo de riesgo y no riesgo de las cinco conductas anómalas en las mujeres de la muestra del estado de Morelos, México 2001-2002. Clasificación cualitativa.



Medias de los grupos

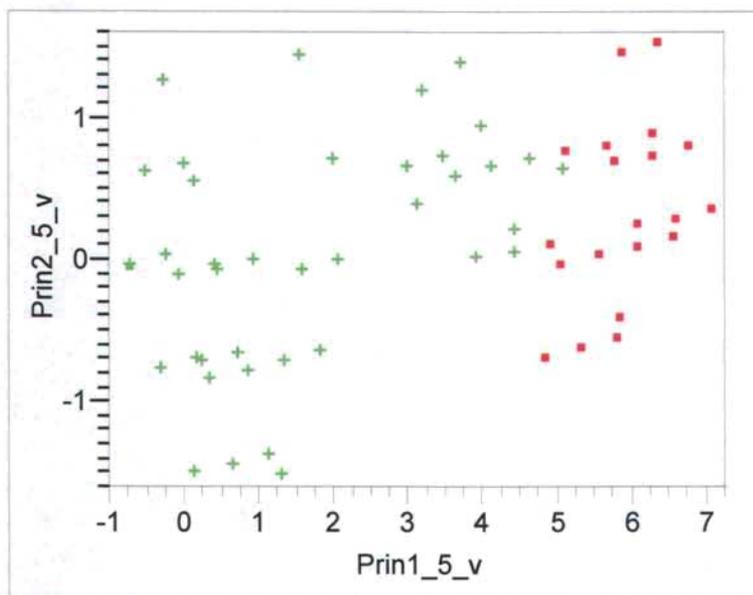
Grupos	Cantidad	dieta	temosobra	preodelga	ejeex	comvida
0	1,521	0.013	0.003	0.001	0	0.001
1	51	1	1	0.843	0.137	0.784
	1,572					

Prevalencia = $51/1,572 = 0.0324 = 3.24\%$

Desviaciones estándar de los grupos

Grupos	Dieta	temosobra	preodelga	ejeex	comvida
0	0.013	0.003	0.001	0	0.001
1	0	0	0.021	0.104	0.037

Figura 6. Grafico de los dos componentes principales para evidenciar la clasificación cualitativa.



- Individuos con riesgo = 82 individuos
- + Individuos sin riesgo = 2,502 individuos

2.2.- Formación de grupos

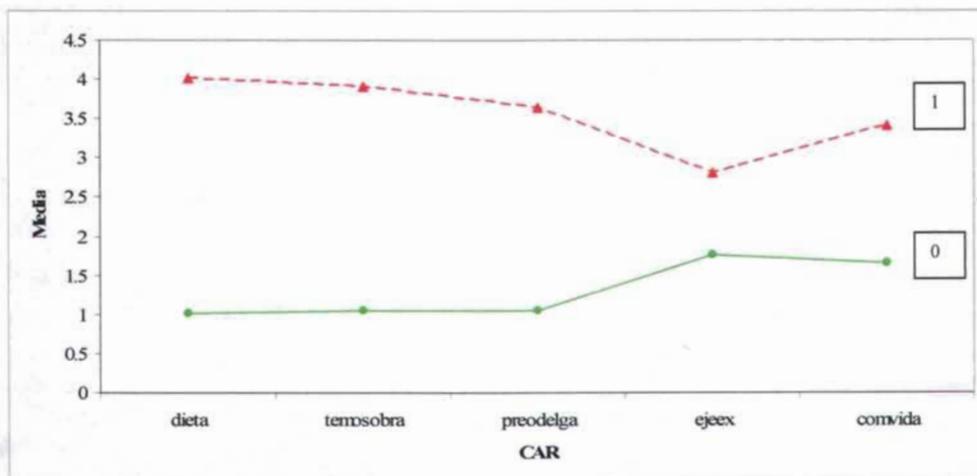
- ❖ **Método 1: Conglomerados ordinales, obtenidos del análisis de conglomerados de k-medias, considerando la escala de respuesta ordinal.**

Los conglomerados ordinales que se obtienen, figuras 7 y 8, se aprecia que para ambos géneros, la distribución de las variables es igual, nuevamente, se aprecia que la variable de frecuencia de ejercicio para reducir peso corporal, es la variable mas baja en el grupo de riesgo, y esta misma es la más elevada en el grupo de no riesgo.

Los cuadros correspondientes a estas figuras, indican que la prevalencia en las CAR es mayor en las mujeres que en los hombres (4.6% vs. 4.2%), aunque nuevamente esta diferencia no es muy pronunciada. Cuando se estratifica por grupo de edad, 20 casos corresponden a edades entre los 14 y 16 años de los 43 casos totales en los varones; mientras que en las mujeres para este mismo grupo de edad, corresponde a 35 de los 72 casos.

Estos resultados indican que no es tan despreciable la prevalencia de este tipo de conductas en los varones. Sería interesante analizar estos datos de manera longitudinal, para ver si este patrón se repite.

Figura 7. Valores promedio por grupo de riesgo y no riesgo de las cinco conductas anómalas en los hombres de la muestra del estado de Morelos, México 2001-2002. Conglomerado ordinal.



Medias de los grupos

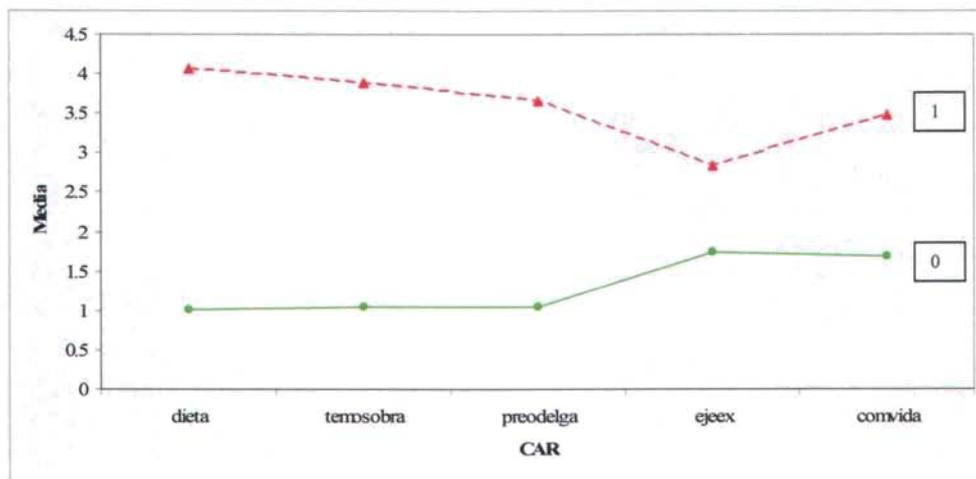
Grupos	Cantidad	dieta	temosobra	preodelga	ejeex	Comvida
0	969	1.016	1.054	1.049	1.763	1.646
1	43	4.023	3.906	3.651	2.813	3.418
	1,012					

Prevalencia = $43/1,012 = 0.0424 = 4.2\%$

Desviaciones estándar de los grupos

Grupos	dieta	temosobra	preodelga	ejeex	Comvida
0	0.256	0.227	0.217	0.466	0.552
1	0.152	0.569	0.752	0.587	0.626

Figura 8. Valores promedios por grupo de riesgo y no riesgo de las cinco conductas anómalas en las mujeres de la muestra del estado de Morelos, México 2001-2002. Conglomerado ordinal.



Medias de los grupos

Grupos	Cantidad	dieta	temsobra	preodelga	ejeex	Comvida
0	1500	1.005	1.046	1.046	1.75	1.688
1	72	4.069	3.902	3.666	2.847	3.486
	1,572					

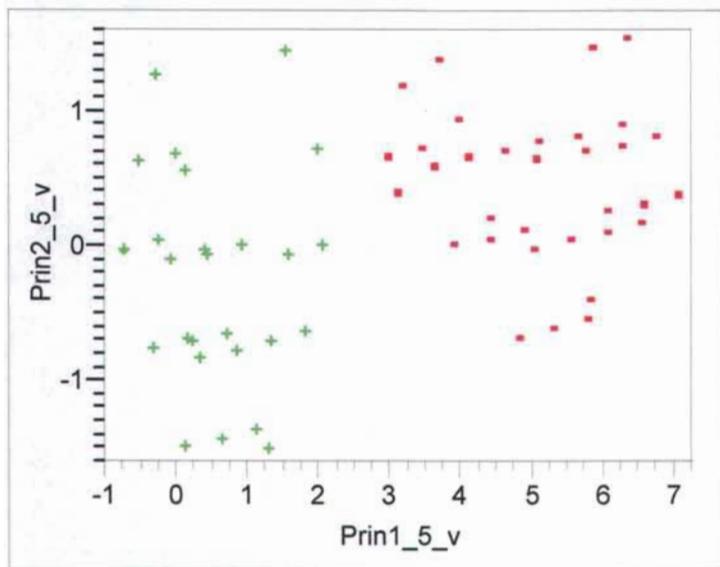
Prevalencia = $72/1,572 = 0.0458 \approx 4.6\%$

Desviaciones estándar de los grupos

Grupos	dieta	temsobra	preodelga	ejeex	Comvida
0	0.146	0.210	0.215	0.478	0.537
1	0.387	0.608	0.712	0.573	0.627

Al graficar los dos componentes principales, para apreciar los grupos, figura 9, se observa que el primer componente, diferencia claramente el grupo de riesgo y no riesgo, de tal manera que los valores elevados en este componente indica los elementos agrupados en el grupo de riesgo (valores mayores de 2.75). Al contrario, los valores bajos son indicativos de elementos agrupados como cluster de no riesgo.

Figura 9. Gráfico de los dos componentes principales para evidenciar el conglomerado ordinal.



- Individuos con riesgo = 115 individuos
- + Individuos sin riesgo = 2,469 individuos

Concordancia (correspondencia) entre clasificación cualitativa y conglomerado ordinal

Al calcular el índice de concordancia o correspondencia (anexo), se observa que las dos clasificaciones estarían de acuerdo al clasificar a los individuos con riesgo en un 92.66%. De hecho, ellos están de acuerdo en el 98.7% de los individuos enfermos. Se tiene un valor de kappa de 0.83, esto indica una buena concordancia entre ambas clasificaciones. Una baja reproducibilidad hace que disminuya la validez de la clasificación, sin embargo una alta reproducibilidad no es sinónimo de validez, tabla IV.

Tabla IV. Concordancia entre la clasificación cualitativa y la conglomerado ordinal.

De acuerdo	Se esperaría un acuerdo	Kappa	Std. Err.	Z	Prob>Z
98.72%	92.66%	0.8260	0.0194	42.64	0.0000

El índice de kappa normalmente se usa para determinar el grado de confiabilidad del instrumento que se utilizó, sin embargo, para medir la validez se calcula la sensibilidad y la especificidad.¹² La tabla V, permite observar que tienen una muy buena clasificación con el conglomerado ordinal, debido a que la sensibilidad y especificidad son elevados, sin embargo estos resultados deben ser tomados con cautela, ya que normalmente esta determinación se emplea con más frecuencia en un instrumento que posee más puntos de corte, y donde se determina un punto de corte con alta sensibilidad y especificidad. Para la Característica de Operación Receptora (ROC), donde se grafica la sensibilidad vs. 1-especificidad, el área bajo la curva es un resumen de la precisión del diagnóstico, esto es, el conglomerado ordinal tiene un 99.3% de probabilidad de distinguir correctamente a un sujeto normal de un sujeto enfermo basado en esta clasificación (anexo). Además se da un ejemplo numérico de este procedimiento.

Tabla V. Cálculo de la sensibilidad y la especificidad

Punto de corte	Sensibilidad	Especificidad	Correctamente Clasif	Area bajo ROC
(= 1)	100.00%	98.68%	98.72%	0.993

Ejemplo numérico de esta metodología³⁶

Considere los siguientes datos:

Individuo	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

En este procedimiento se usará el método de la minimización de traza(W). El primer paso es encontrar una partición inicial, usando la distancia Euclidiana, en donde las medias de los grupos son:

	Grupo 1	Grupo2
Individuo	1	4
Vector de la media	[1.0,1.0]	[5.0,7.0]

Ahora los individuos serán redistribuidos a sus grupos:

Paso	Grupo1		Grupo2	
	Individuo	Vector de la media	Individuo	Vector de la media
1	1	[1.0, 1.0]	4	[5.0, 7.0]
2	1,2	[1.2, 1.5]	4	[5.0, 7.0]
3	1,2,3	[1.8, 2.3]	4	[5.0, 7.0]
4	1,2,3	[1.8, 2.3]	4,5	[4.2, 6.0]
5	1,2,3	[1.8, 2.3]	4,5,6	[4.3, 5.7]
6	1,2,3	[1.8, 2.3]	4,5,6,7	[4.1, 5.4]

Los dos grupos poseen las siguientes características:

Grupo 1 Individuos 1, 2 y 3
 Vector de la media = [1.8, 2.3]
 Traza (W1):
 $Var1 = 0.64+0.09+1.44 = 2.17$ (distancias de la media de la var1 a c/obs.)
 $Var2 = 1.69+0.09+2.89 = 4.67$ (distancias de la media de la var2 a c/obs.)
 $tr(W1) = 2.17+4.67 = 6.84$

Grupo 2 Individuos 4, 5, 6 y 7
 Vector de la media = [4.1, 5.4]
 Traza (W2):
 $Var1 = 0.81+0.36+0.16+0.36 = 1.69$
 $Var2 = 2.56+0.16+0.16+0.81 = 3.69$
 $tr(W2) = 1.69+3.69 = 5.38$

$$tr(W) = tr(W1)+tr(W2) = 6.84+5.38 = 12.22$$

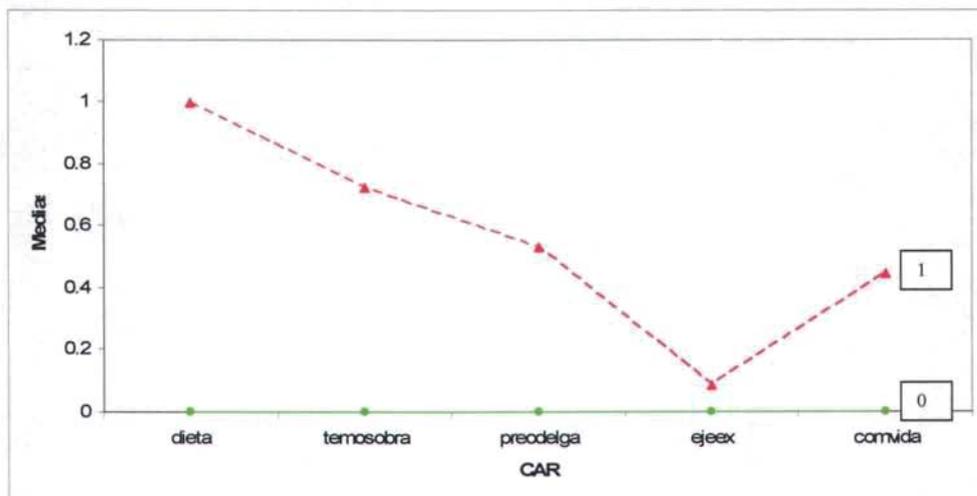
Ahora si se mueve el individuo 3 al Segundo grupo, se tiene la $tr(W1)=0.63$, $tr(W2)=7.90$ y $tr(W)=8.53$, si el movimiento del individuo permite disminuir el criterio de agrupamiento, el cambio se realiza y continua el proceso iterativo.

❖ **Método 2: Conglomerados binarios, obtenido del análisis de conglomerados de k-medias, considerando la escala de respuesta binaria.**

Una vez que se han transformado los patrones de respuesta, a datos binarios de manera que: “nunca”, “rara vez” y “a veces”, “una vez en el último mes” y “dos a tres veces en el último mes”, valen 0. En contraparte, el valor de las respuestas indicativas de: “con frecuencia”, “siempre”, “de cuatro a seis veces a la semana” y “diario” poseen valor de 1. Posteriormente se determina el coeficiente de semejanza para cada par de observaciones.

La distribución nuevamente se acerca mucho a lo anteriormente descrito en las anteriores clasificaciones, figura 10 y figura 11. La prevalencia es exactamente la misma para hombres y mujeres (cuadros correspondientes). Pareciera que en esta muestra de estudiantes, esencialmente las actitudes evidencian el riesgo de este tipo de enfermedades, esto es, las variables temor al sobrepeso, preocupación por estar delgada y el sentir que la comida controla la vida, en comparación con las conductas referentes a la realización de dietas reductivas, así como el hacer ejercicio para bajar de peso. Sería interesante analizar también actitudes referentes a las conductas purgativas, es decir, analizar por ejemplo, el deseo de controlar el peso corporal mediante vómito autoinducido.

Figura 10. Valores promedio por grupo de riesgo y no riesgo de las cinco conductas anómalas en los hombres de la muestra del estado de Morelos, México 2001-2002. Conglomerado binario.



Medias de los grupos

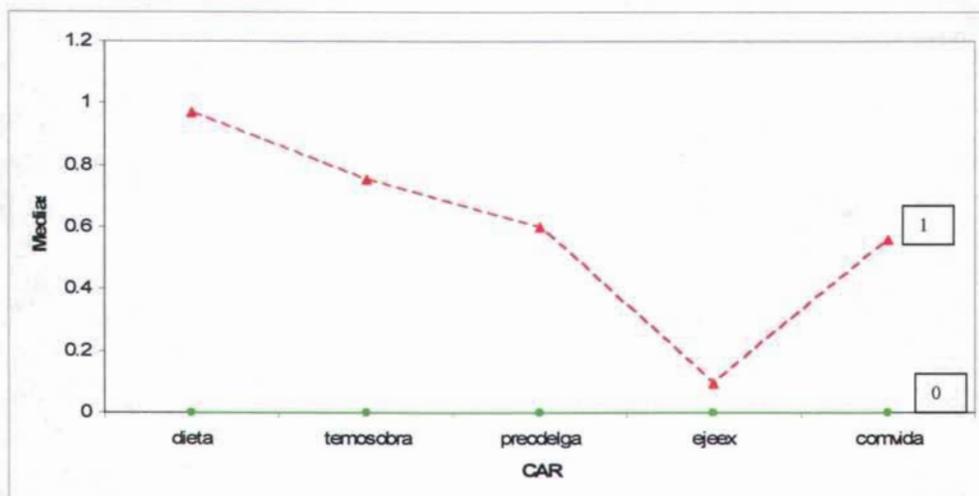
Grupos	Cantidad	dieta	temosobra	preodelga	ejeex	Comvida
0	965	0	0	0	0	0
1	47	1	0.723	0.532	0.085	0.447
	1,012					

Prevalencia = $47/1,012 = 0.0464 = 4.64\%$

Desviaciones estándar de los grupos

Grupos	dieta	temosobra	preodelga	ejeex	Comvida
0	0	0	0	0	0
1	0	0.205	0.254	0.033	0.253

Figura 11. Valores promedio por grupo de riesgo y no riesgo de las cinco conductas anómalas en las mujeres de la muestra del estado de Morelos, México 2001-2002. Conglomerado binario.



Medias de los grupos

Grupos	Cantidad	dieta	temosobra	preodelga	ejeex	Comvida
0	1,499	0	0	0	0	0
1	73	0.972	0.753	0.602	0.095	0.561
	1,572					

Prevalencia = $73/1,572 = 0.0464 = 4.64\%$

Desviaciones estándar de los grupos

Grupos	dieta	temosobra	preodelga	ejeex	Comvida
0	0	0	0	0	0
1	0.027	0.181	0.241	0.087	0.249

A continuación se proporciona un ejemplo con algunos valores de las cinco conductas anómalas, tabla VI, así como el valor del conglomerado binario en 12 individuos. La tabla VII y VIII, corresponde a la matriz de similitudes y disimilitudes, respectivamente, calculada en base al

coeficiente de Jaccard, además se presentan los promedios por grupos (0 vs. 0.541 y 0.832 vs. 0.356).

Tabla VI. Datos correspondientes a 12 individuos de la muestra de adolescentes

folio	sexo2	di2	te2	pr2	ej2	co2	Conglo_bina rio
1	2	0	0	0	0	0	0
9	2	0	0	0	0	0	0
11	2	0	0	0	0	0	0
13	2	0	0	0	0	0	0
16	2	0	0	0	0	0	0
112	2	1	0	0	0	0	1
162	2	1	1	1	0	1	1
561	2	1	0	0	0	0	1
567	2	1	1	1	0	0	1
799	2	1	1	1	0	1	1
827	2	1	1	1	1	1	1
1064	2	1	0	0	0	0	1
1428	2	1	1	1	0	1	1
1619	2	1	1	1	0	1	1

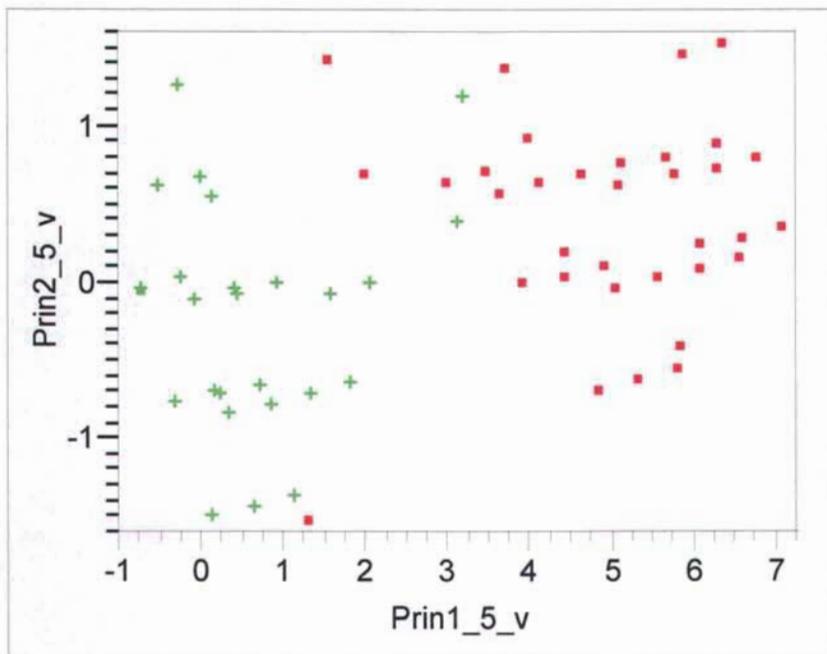
Tabla VII. Matriz de similitudes

	1	9	11	13	16	112	162	561	567	799	827	1064	1428	1619
1	1.000													
9	0.000	1.000												
11	0.000	0.000	1.000											
13	0.000	0.000	0.000	1.000										
16	0.000	0.000	0.000	0.000	1.000									
112	0.000	0.000	0.000	0.000	0.000	1.000								
162	0.000	0.000	0.000	0.000	0.000	0.250	1.000							
561	0.000	0.000	0.000	0.000	0.000	1.000	0.250	1.000						
567	0.000	0.000	0.000	0.000	0.000	0.333	0.750	0.333	1.000					
799	0.000	0.000	0.000	0.000	0.000	0.250	1.000	0.250	0.750	1.000				
827	0.000	0.000	0.000	0.000	0.000	0.200	0.800	0.200	0.600	0.800	1.000			
1064	0.000	0.000	0.000	0.000	0.000	1.000	0.250	1.000	0.333	0.250	0.200	1.000		
1428	0.000	0.000	0.000	0.000	0.000	0.250	1.000	0.250	0.750	1.000	0.800	0.250	1.000	
1619	0.000	0.000	0.000	0.000	0.000	0.250	1.000	0.250	0.750	1.000	0.800	0.250	0.000	1.000
	0.000	0.000	0.000	0.000	9.000	3.533	5.05	2.283	3.183	3.05	1.8	0.5	0	
	0	0	0	0	0	0.441	0.721	0.381	0.636	0.7625	0.6	0.250	0	0

$$3.7915 / 7 = 0.541$$

0

Figura 12. Gráfico correspondiente a los dos componentes principales para evidenciar el conglomerado binario.



- Individuos con riesgo = 120 individuos
- + Individuos sin riesgo = 2,464 individuos

Concordancia de la clasificación cualitativa contra clasificación del conglomerado binario

Al calcular el índice de concordancia, tabla IX, se tienen que se esperaba que las dos clasificaciones, la cualitativa con el conglomerado binario, estarían de acuerdo al clasificar a los individuos con riesgo en un 92.48%. De hecho, estas clasificaciones están de acuerdo en el 98.5% de los individuos enfermos. Se tiene una kappa de 0.80, esto es, se obtiene una muy buena concordancia entre ambas clasificaciones. Se obtiene también elevados valores de sensibilidad y especificidad, así como valores muy buenos para el área bajo la curva ROC, tabla X.

Tabla IX. Concordancia de los grupos formados por el conglomerado binario contra clasificación cualitativa

De acuerdo	Se esperaría un acuerdo	Kappa	Std. Err.	Z	Prob>Z
98.53%	92.48%	0.8045	0.0193	41.70	0.0000

Tabla X. Cálculo de la sensibilidad y la especificidad

Punto de corte	Sensibilidad	Especificidad	Correctamente Clasificado	Area bajo ROC
(= 1)	100.00%	98.48%	98.53%	0.992

❖ *Comparación de las tres clasificaciones: cualitativa, conglomerado ordinal y conglomerado binario.*

La prevalencia de los dos grupos formados, el de riesgo y el de no riesgo, en los conglomerados ordinales y binarios, son muy semejantes, tabla XI; en cambio en la clasificación cualitativa, el porcentaje de personas con riesgo disminuye.

La comparación correspondiente al conglomerado elaborado con datos ordinales contra la clasificación cualitativa, tabla XII, indica que en los hombres y las mujeres, 31 y 51, respectivamente, fueron clasificados de riesgo en ambos análisis, sin embargo, 12 y 21, en ese mismo orden, solo se encuentran en riesgo en el conglomerado ordinal, al revisar a estos individuos, se tienen que estas personas, poseen un patrón de respuestas, con elevada frecuencia en una o dos conductas anómalas, por ello, no aparecen en la clasificación sugerida por Unikel y colaboradores, sin embargo, una persona que frecuentemente hace dietas para mantener un peso corporal "bajo", o que la comida controla su vida, puede ser una persona que puede estar ideando la manera de incorporar otros métodos para mantenerse "delgado" o bien, puede llevar al extremo, una o mas conductas.

Cabe aclarar, que además, estas personas también presentan cuatro conductas en con frecuencia intermedia, es decir, presentan respuestas en el cuestionario de "algunas veces", que añadidas con una o dos conductas en elevada frecuencia, son personas catalogadas como de riesgo por el conglomerado ordinal.

La comparación entre la clasificación de conglomerado binario contra clasificación cualitativa, tabla XIII, se observa que 38 individuos, 16 y 22, hombres y mujeres respectivamente, son clasificados como de riesgo solo por el conglomerado binario, se observa que estas personas poseen al igual que la comparación anterior, una o dos conductas en una frecuencia elevada, y cuatro o tres conductas en una escala intermedia.

En la tabla XIV, se describe la comparación entre el conglomerado ordinal y el binario, se aprecia que 43 (91.4%) hombres han sido clasificados dentro del grupo de riesgo, en ambos conglomerados. En el conglomerado binario se tienen además 4 individuos agregados en el grupo de riesgo. Al revisar los datos de estos individuos, se observa que la variable de dieta, tiene una frecuencia elevada, mientras que el resto de las otras conductas la frecuencia es baja. En el caso de las mujeres, 70 mujeres (95%), son clasificadas por los dos métodos de análisis, como grupo de riesgo. Las diferencias encontradas en los conglomerados, ordinales y binarios, son 2 y 3 individuos respectivamente. Las dos mujeres contempladas en el grupo de riesgo solo por el ordinal, poseen las cinco conductas y actitudes en una frecuencia intermedia, es decir de 3 (“algunas veces”). Por otro lado, analizando ahora las mujeres que aparecen en el grupo de riesgo para el binario, se observa que dos de ellas, poseen una frecuencia elevada en la variable de dieta, pero baja en las demás variables. La tercera, posee una frecuencia elevada en la variable de “la vida es controlada por la comida”, y baja en el resto de las variables.

Es decir, en general, los individuos catalogados como de riesgo, suelen presentar más de 3 variables con frecuencias elevadas, excepto por las 9 personas descritas anteriormente.

Por último, 12 hombres, que están contemplados en el grupo de riesgo en ambos conglomerados, pero que no se encuentran en el grupo de riesgo en la agrupación cualitativa, poseen uno o dos conductas en una frecuencia elevada, dos o tres de las demás conductas se encuentran en frecuencia intermedia.

De forma semejante, las mujeres que se encuentran como en el grupo de riesgo solo en ambos conglomerados, son 19, ellas poseen uno a dos conductas en frecuencia elevada y dos o mas frecuencias intermedias en el resto de estas conductas.

No hay personas que estén en el grupo de riesgo en los dos conglomerados y que en la clasificación cualitativa indique lo contrario, es decir, que en esta clasificación no se encuentren como de riesgo.

Al graficar los dos componentes principales contra todas las clasificaciones obtenidas, figura 13, se observa que los grupos formados por los conglomerados, caracterizan al grupo de riesgo de manera semejante a como lo realiza la clasificación cualitativa, aunque hay sus excepciones. Nuevamente son notorias las observaciones que se encuentran en el grupo de riesgo solo para el conglomerado binario, donde de estas 7 observaciones, solo un individuo presenta muy frecuentemente la idea de que la comida controla su vida (anteriormente descrito en el conglomerado binario).

Tabla XI. Tabla descriptiva de las clasificaciones analizadas

Género	Tipo de clasificación						Total
	Clasificación cualitativa		Conglomerado ordinal		Conglomerado binario		
	Sin riesgo	Con riesgo	Sin riesgo	Con riesgo	Sin riesgo	Con riesgo	
Hombre	981 (96.94%)	31 (3.06)	969 (95.75%)	43 (4.25%)	965 (95.36%)	47 (4.64%)	1,012 (39.16%)
Mujer	1,521 (96.76%)	51 (3.24%)	1,500 (95.42%)	72 (4.58%)	1,499 (95.36%)	73 (4.64%)	1,572 (60.84%)
Total	2,502 (96.83%)	82 (3.17%)	2,469 (95.55%)	115 (4.45%)	2,464 (95.36%)	120 (4.64%)	2,584 (100.0%)

Tabla XII. Tabla comparativa entre la clasificación cualitativa y el conglomerado ordinal

			Clasificación		
Clasificación	Género	Grupo	Conglomerado ordinal		
			Sin riesgo	Con riesgo	Total
Clasificación cualitativa	Hombres	Sin riesgo	969 98.78 100%	2 1.22% 27.91%	981 100% 96.94%
		Con riesgo	0 0% 0%	31 100% 72%	31 100% 3.06%
		Total	969 95.75% 100%	43 4.25% 100%	1,012 100%
	Mujeres	Sin riesgo	1,500 98.62% 100%	21 1.38% 29.17%	1,521 100% 96.76%
		Con riesgo	0 0%	51 100% 70.83%	51 100% 3.24%
		Total	1,500 95.42% 100%	72 4.58% 100%	1,572 100%

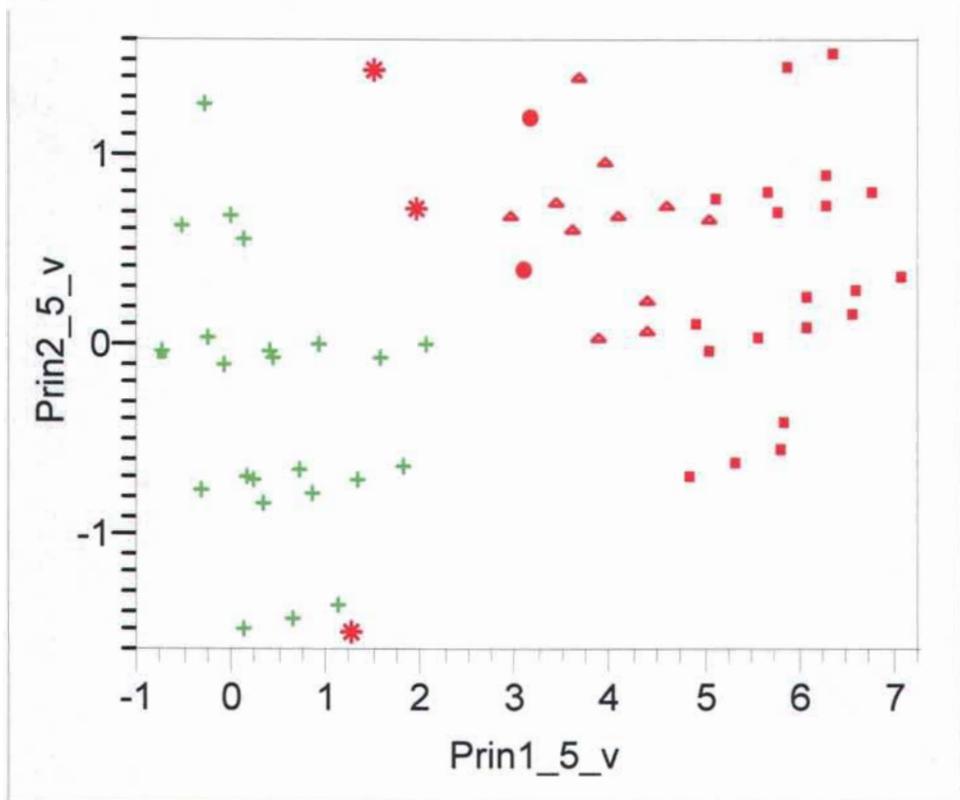
Tabla XIII. Tabla comparativa entre la clasificación cualitativa y conglomerado binario

			Clasificación		
Clasificación	Género	Grupo	Conglomerado binario		
			Sin riesgo	Con riesgo	Total
Clasificación cualitativa	Hombres	Sin riesgo	965 98.37% 100%	16 1.63% 34.04%	981 100% 96.94%
		Con riesgo	0 0%	31 100% 65.96%	31 100% 65.96%
		Total	965 95.36% 100%	47 4.64% 100%	1,012 100%
	Mujeres	Sin riesgo	1,499 98.55% 100%	22 1.45% 30.14%	1,521 100% 96.76%
		Con riesgo	0 0%	51 100% 69.86%	51 100% 3.24%
		Total	1,499 95.36% 100%	73 4.64% 100%	1,572 100%

Tabla XIV. Tabla comparativa entre el conglomerado ordinal y el conglomerado binario

			Clasificación		
Clasificación	Género	Grupo	Conglomerado binario		
			Sin riesgo	Con riesgo	Total
Conglomerado ordinal	Hombres	Sin riesgo	965 99.59% 100%	█ 0.41% 8.51%	969 100% 95.75%
		Con riesgo	0 0% 0%	43 100% 91.49%	43 100% 4.25%
		Total	965 95.36% 100%	47 4.64% 100%	1,012 100%
	Mujeres	Sin riesgo	1,497 99.80% 99.87%	█ 0.20% 4.11%	1,500 100% 95.42%
		Con riesgo	█ 2.78% 0.13%	70 97.22% 95.89%	72 100% 4.59%
		Total	1,499 95.36% 100%	73 4.64% 100%	1,572 100%

Figura 13. Gráfico correspondiente a los dos componentes principales para evidenciar la clasificación cualitativa y los conglomerados ordinal y binario.



Simbología:

- Individuos con riesgo (en las tres clasificaciones) = 82 individuos
- Individuos con riesgo solo en el conglomerado ordinal = 2 individuos
- * Individuos con riesgo solo en el conglomerado binario = 7 individuos
- △ Individuos con riesgo en los conglomerados ordinal y binario = 31 individuos
- + Individuos sin riesgo

Discusión

Varios estudios tienen evidencia de que un gran número de individuos entre seis y diecisiete años, no están satisfechos con su peso corporal. Un porcentaje relativamente pequeño, de ellos, posee trastornos de la conducta alimentaria, y una proporción mayor puede exhibir signos de síndrome parcial o “a riesgo”.³⁹

En el presente estudio se analiza el síndrome parcial en ambos géneros, en el apartado de resultados, la prevalencia en el grupo de riesgo para hombres y mujeres es muy semejante, contradiciendo este resultado con el reportado en otros estudios.⁴⁰ Sin embargo, en un estudio realizado en estudiantes mexicanos, la figura ideal promedio en los hombres es una figura mesomorfa o musculosa, además se evidenció la presencia de subestimación de la imagen corporal, elemento decisivo en el desencadenamiento de conductas de este tipo en las mujeres.⁴¹ En la literatura mundial, se constata que la frecuencia de TCA va en aumento en ambos géneros.⁴²

En el análisis de la medición basal, se obtuvieron prevalencias igualmente muy semejantes entre hombres y mujeres; además se determinó que en estas últimas la obesidad y la sobreestimación de la imagen corporal son dos elementos muy importantes, en el desarrollo de CAR, (análisis de regresión logística múltiple), resultados coherentes con la literatura.⁴³

Analizando el cuestionario empleado, se tiene que tres de sus preguntas, requerían que los encuestados recordaran la frecuencia por un lapso de 30 días, originando con ello un sesgo de memoria; además, la batería de preguntas es nueva y compleja en su análisis, debido a lo anterior no se puede comparar con otros estudios previos. Otro defecto que posee, es que no se contemplan los atracones, y los sentimientos de pérdida de control sobre la ingesta del alimento, elementos sumamente importantes en la conducta bulímica.⁶ Sin embargo, las cinco variables analizadas, poseen un alfa de Cronbach, indicativo de una alta confiabilidad.

Por otra parte la muestra no es representativa, ya que el muestreo presenta algunos errores que no permiten extrapolar resultados a poblaciones semejantes. Conjuntamente a lo anterior, la muestra posee más mujeres que hombres, ya que originariamente se deseaba realizar un estudio para la población femenina, por ello se tienen un sesgo por género.

La formación de dos grupos, -de riesgo y sin riesgo-, obedece a que se probó el análisis para formar dos y tres grupos (éste último no reportado), de los cuales, la obtención de dos grupos, fue más consistente, ya que se delimitaron mejor los grupos. De alguna manera el grupo de riesgo está conformado casi en su totalidad por individuos que realizan frecuentemente dietas para reducir peso, dietas que por cierto son recomendadas casi en su totalidad por amistades de los entrevistados.

En el apartado de elección de las variables, se trabajó con cinco de las variables originales, ya que delimitaron muy bien los grupos; de hecho, cuando se analizan en el análisis de conglomerados en las trece variables, las personas que se encuentran con riesgo, son las mismas que se encuentran en el conglomerado de cinco variables, por lo tanto, para obtener el conglomerado binario se utilizan nuevamente estas cinco conductas. La clasificación cualitativa que sirve como estándar, se deriva nuevamente de las variables anteriores.

Las prevalencias son muy parecidas en dos de las tres clasificaciones analizadas en este trabajo.

En el caso del conglomerado ordinal, los casos (de riesgo) encontrados en los grupos de edad de 14 a 16 años son muy parecidos a los encontrados en las edades de 17 a 24 años. Además el resultado obtenido en este mismo conglomerado, es más consistente gráficamente (componentes principales) y en la determinación de su confiabilidad (índice de concordancia), así como en su validez (sensibilidad y especificidad). Sin embargo, en general, ambas clasificaciones, el conglomerado ordinal y binario, dividen adecuadamente a los individuos con riesgo. Cabe aclarar, que debido a que no se cuenta con una muestra psiquiátrica, el "gold standard" o estándar de oro, fue un razonamiento que a través de la experiencia ha permitido diferenciar individuos con riesgo, sin embargo se sugiere que en análisis posteriores, este instrumento sea evaluado con una muestra diagnosticada como enferma mediante el análisis discriminante.

Cuando se analizan las variables como variables dicotómicas, se determinó además el coeficiente de Russell, donde la confiabilidad se ve disminuida ($\kappa=0.756$). En cambio, cuando se forman los grupos en base al coeficiente de Hamman, es incluso mucho mejor que el elaborado por el coeficiente de Jaccard ($\kappa=0.963$), estos datos no están reportados en los resultados. Por último, la consideración de este análisis es interesante, ya que en ocasiones, se tienen datos dicotómicos y no se dispone de más información.

Al revisar nuevamente las clasificaciones sin datos "extremos", esto es sin considerar aquellas personas que se encuentran en los gráficos de componentes principales en los bordes (folios, 246, 508, 1,540 y 1,744), se aprecia que los conglomerados son consistentes. Al revisar las conductas de estas personas, no se perciben como casos atípicos ya que las conductas y actitudes son muy parecidas al resto de los individuos, por ello no se descartaron estas observaciones.

Es preciso comentar que uno de los inconvenientes del análisis multivariado de conglomerados es que es un análisis descriptivo y no inferencial, ya que este análisis no tiene bases estadísticas sobre las cuales deducir inferencias estadísticas para una población a partir de una muestra, utilizándose fundamentalmente como una técnica exploratoria. Las soluciones no son únicas, en la medida en que la pertenencia al conglomerado para cualquier número de soluciones depende de muchos elementos del procedimiento y se pueden obtener muchas soluciones diferentes variando uno o más de estos elementos.²⁸ A pesar de lo anterior se calculó la validez y la confiabilidad de los conglomerados bajo reserva, sin embargo se debe hacer notar que estas mediciones son más convenientes una vez que se ha efectuado un análisis discriminante.

Por otro lado, la clasificación cualitativa es mucho más exigente en considerar sanos y no sanos, ya que, en algunas ocasiones el individuo que no presentó riesgo en esta clasificación, debido a que la persona presenta una o dos conductas muy frecuentes, se clasificó de riesgo en el conglomerado ordinal. Cuando se analizan casos clínicos, se pueden llegar a presentar casos en donde el paciente reporte, la presencia de una o dos de estas conductas, en una elevada frecuencia y donde el resto de las demás actitudes y conductas sea en menor frecuencia. La clasificación menos exigente, en este contexto fue el conglomerado binario.

El próximo paso a seguir es la validación del cuestionario, ya que se utilizó en la 1ª, 2ª y se pretende aplicar en una 3ª evaluación. Los dos análisis cuantitativos mostraron ser una posible

opción para determinar que individuos están en riesgo, mediante un patrón de respuesta a cuestionarios autoaplicados, sin embargo, se necesita estar consciente de que las preguntas pueden no ser tan claras para el encuestado, además de los inconvenientes que se presten al momento de la encuesta (horarios, ruido, etc.), y de que finalmente, las personas no deseen ser honestas por alguna causa.

En este estudio se analizaron conductas y actitudes de alimentación que son anómalas, puede ser interesante analizar a futuro, qué permite al individuo desarrollar una u otra conducta, ¿qué hace que conductas purgativas sean menos frecuentes que las conductas no purgativas? (datos no reportados), ¿es más terrible tener una conducta que una actitud?, pero si una actitud, finalmente nos permite hacer lo que hacemos en la vida cotidiana, ¿las actitudes pueden ser igualmente mortales?

Para finalizar, el fenómeno de conductas alimentarias de riesgo es multicausal y complejo, ya que esta involucrado el contexto social, en donde los medios de difusión venden una figura delgada como exitosa. Los cambios físicos y psicológicos en la adolescencia también juegan un papel importante; por ello, convendría realizar estudios longitudinales, para permitirnos entender un poco más de este gran problema psiquiátrico.

Conclusiones

Algunas alternativas de análisis que pueden ayudar a los investigadores del área de salud mental, es sin duda alguna los métodos multivariados; en el presente trabajo, se enfatiza el método de análisis de conglomerados para obtener el algoritmo de clasificación de individuos, mediante un conjunto de variables categóricas. Las propuestas cuantitativas permitirán dar posibles soluciones a un tema tan complicado como lo es la medición de conductas y actitudes en las personas. Con este tipo de análisis se pretende disminuir en lo posible los juicios tendenciosos de los investigadores.

Las investigaciones encaminadas a la comprensión de este fenómeno permitirán detectar, confirmar y predecir, individuos con algún problema de esta índole.

Por último, el conocimiento más amplio de esta enfermedad permitirá poder establecer programas preventivos adecuados a poblaciones vulnerables.

opción para determinar que individuos están en riesgo, mediante un patrón de respuesta a cuestionarios autoaplicados, sin embargo, se necesita estar consciente de que las preguntas pueden no ser tan claras para el encuestado, además de los inconvenientes que se presten al momento de la encuesta (horarios, ruido, etc.), y de que finalmente, las personas no deseen ser honestas por alguna causa.

En este estudio se analizaron conductas y actitudes de alimentación que son anómalas, puede ser interesante analizar a futuro, qué permite al individuo desarrollar una u otra conducta, ¿qué hace que conductas purgativas sean menos frecuentes que las conductas no purgativas? (datos no reportados), ¿es más terrible tener una conducta que una actitud?, pero si una actitud, finalmente nos permite hacer lo que hacemos en la vida cotidiana, ¿las actitudes pueden ser igualmente mortales?

Para finalizar, el fenómeno de conductas alimentarias de riesgo es multicausal y complejo, ya que esta involucrado el contexto social, en donde los medios de difusión venden una figura delgada como exitosa. Los cambios físicos y psicológicos en la adolescencia también juegan un papel importante; por ello, convendría realizar estudios longitudinales, para permitirnos entender un poco más de este gran problema psiquiátrico.

Conclusiones

Algunas alternativas de análisis que pueden ayudar a los investigadores del área de salud mental, es sin duda alguna los métodos multivariados; en el presente trabajo, se enfatiza el método de análisis de conglomerados para obtener el algoritmo de clasificación de individuos, mediante un conjunto de variables categóricas. Las propuestas cuantitativas permitirán dar posibles soluciones a un tema tan complicado como lo es la medición de conductas y actitudes en las personas. Con este tipo de análisis se pretende disminuir en lo posible los juicios tendenciosos de los investigadores.

Las investigaciones encaminadas a la comprensión de este fenómeno permitirán detectar, confirmar y predecir, individuos con algún problema de esta índole.

Por último, el conocimiento más amplio de esta enfermedad permitirá poder establecer programas preventivos adecuados a poblaciones vulnerables.

Anexo

*Tipos de variables*⁴⁴

Variables

Las variables son características medibles en las unidades de estudio. Las variables se clasifican:

Variables $\left\{ \begin{array}{l} 1.- \text{ Cualitativas} \\ 2.- \text{ Cuantitativas} \end{array} \right.$

Las variables cualitativas, se dividen a su vez, en nominales y ordinales, las variables cuantitativas se divide en variables de intervalo y de relación.

Tipos de escalas

Nominal. Para cada unidad de estudio se determina la pertinencia a una entre varias categorías. Dichas categorías son mutuamente excluyentes y exhaustivas y no es posible establecer relaciones de orden entre las categorías. En esta escala solo puede decirse a qué clase pertenece cada unidad de estudio, por ejemplo, género, donde las categorías son: femenino y masculino.

Ordinal. En este caso se considera que existe un grado de intensidad de la propiedad medida, por lo que las categorías guardan un orden. Al igual que la escala nominal se determina la pertenencia de las unidades a categorías. Ejemplo, la escala de Likert, donde su respuesta va de nunca, rara vez, algunas veces, con frecuencia y siempre.

Intervalo. Se asignan números para indicar la intensidad de una característica, con unidad de medida y origen arbitrarios, que se elige en conveniencias prácticas, de tal manera que se mantiene la igualdad de los intervalos. Además de estar ordenados, se permite hacer operaciones aritméticas. Por ejemplo, temperatura, el cociente intelectual, aquí el cero no indica la ausencia de la característica y como se indica anteriormente, el cero puede moverse de manera arbitraria.

De relación o razón. Se asignan números para señalar la intensidad de una característica con unidad de medida arbitraria, pero origen fijo, lo que mantiene la igualdad de las relaciones o proporciones, por ejemplo, el peso, el ingreso, crecimiento bacteriano.

Confiabilidad (α de Cronbach) ³¹

Se determina la confiabilidad, por medio del alfa de Cronbach, esto es, este índice calcula las correlaciones inter ítem o covarianzas de todos los pares de variables a analizar. Es decir, cuantifica la confiabilidad de varias variables que poseen una escala de Likert en sus respuestas. La confiabilidad α representa la correlación esperada de una prueba con una forma alternativa conteniendo el mismo número de ítems.

$$\text{alfa} = \frac{k \sum \frac{c}{v}}{1 + (k-1) \sum \frac{c}{v}}$$

Donde, k es el número de ítems en la escala, c es el promedio de la covarianza entre ítems, v es el promedio de la varianza entre ítems. Si los ítems están estandarizados, la fórmula es:

$$\text{alfa} = \frac{k(r)}{1 + (k-1)r}$$

Donde r es el promedio de correlación entre ítems. Un coeficiente grande de alfa es indicativo de una muy buena confiabilidad de los ítems. El coeficiente se puede aproximar a 1.0.

Análisis de cluster, métodos jerárquicos ³⁸

Agrupamiento jerárquico

Los métodos jerárquicos y otros algoritmos de agrupamiento procuran encontrar un “buen grupo” en los datos, usando una técnica computacional eficiente. Este método no es adecuado en bases de datos muy grandes. El número de maneras de dividir un grupo de n ítem dentro de g grupos esta dado por:

$$N(n, g) = \frac{1}{g!} \sum_{k=1}^g \binom{n}{k} (-1)^{g-k} k^n$$

Esto puede ser aproximadamente $g^n / g!$. Por ejemplo, $N(25, 10) \cong 2.8 \times 10^{18}$. El algoritmo para el agrupamiento jerárquico involucra un proceso secuencial. En cada paso de la aglomeración jerárquica, una observación o grupo de observaciones es unido a otro grupo. En este proceso el número de grupos disminuye y el grupo o cluster crece en dimensión.

Tipos de métodos

Métodos aglomerativos

En el método de *encadenamiento simple*, la distancia entre dos grupos A y B está definido por una mínima distancia entre el punto A y el B :

$$D(A, B) = \min \{ d(x_i, x_j), \text{ para } x_i \in A \text{ y } x_j \in B \}$$

Donde $d(x_i, x_j)$ es la distancia Euclídeana o en otras distancias entre los vectores x_i y x_j . Esta aproximación es también llamada "método del vecino más cercano". La representación se denomina dendograma o gráfico en forma de árbol. El proceso continúa hasta que los objetos se encuentran en un solo conglomerado.

La técnica de *encadenamiento completo* (vecino más lejano), es la distancia entre dos grupos A y B definida como la máxima distancia entre dichos puntos.

$$D(A, B) = \max \{ d(x_i, x_j) \text{ para } x_i \in A \text{ y } x_j \in B \}$$

El *encadenamiento medio o promedio*, está determinado por la distancia promedio de todos los individuos de un conglomerado con todos los individuos de otro.

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(x_i, x_j),$$

Donde la suma es sobre todos las x_i en A y todos x_j en B . A cada paso, se unen dos grupos con la distancia más pequeña.

En el *método del centroide*, la distancia entre los dos conglomerados es la distancia (normalmente Euclídea simple o cuadrada) entre sus centroides. Los centroides de los grupos son los valores medios de las observaciones de las variables en el valor teórico del conglomerado. En este método cada vez que se agrupan a los individuos, se calcula un nuevo centroide. Los centroides de los grupos cambian a medida que se fusionan conglomerados.

$$D(A, B) = d(\bar{x}_A, \bar{x}_B),$$

Donde \bar{x}_A y \bar{x}_B son los vectores de las medias de los vectores de observación en A y en los vectores de observación en B , respectivamente.

Definición, de \bar{y}_A y \bar{y}_B , son de la misma manera
$$\bar{x}_A = \sum_{i=1}^{n_A} \frac{x_i}{n_A}$$

Después de que dos grupos A y B son unidos, el nuevo centroide AB está determinado de la siguiente manera:

$$\bar{x}_{AB} = \frac{n_A \bar{x}_A + n_B \bar{x}_B}{n_A + n_B}$$

El método de Ward, la distancia entre dos conglomerados es la suma de los cuadrados entre dos conglomerados sumados por todas las variables. En cada paso del proceso de aglomeración, se minimiza la suma de los cuadrados dentro del conglomerado para todas las particiones (el conjunto completo de conglomerados disjuntos o separados) obtenida mediante la combinación de dos conglomerados en un paso previo. Este proceso tiende a combinar los conglomerados con un número reducido de observaciones.

Si A y B son los grupos obtenidos por la combinación de los grupos A y B, entonces la suma de las distancias dentro del grupo son:

$$SSE_A = \sum_{i=1}^{n_A} (x_i - \bar{x}_A)' (x_i - \bar{x}_A),$$

$$SSE_B = \sum_{i=1}^{n_B} (x_i - \bar{x}_B)' (x_i - \bar{x}_B),$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (x_i - \bar{x}_{AB})' (x_i - \bar{x}_{AB})$$

Métodos divisibles

Los métodos divisibles comienzan con un grupo completo de observaciones consideradas como un solo cluster o grupo, dentro del cual, se dividen dos grupos. Uno o el otro de estos dos grupos, se divide en subgrupos y así consecutivamente. La elección de que grupo se divide y cómo se divide puede estar basado en variables consideradas a un tiempo o bien que todas las variables sean consideradas simultáneamente. Las técnicas que se emplean aquí son las monotéticas y polítéticas.

Un ejemplo de una técnica monotética es cuando se toman variables binarias, con valores de 0 y 1. Las variables entonces exploran para encontrar la dicotomía maximiza algunos criterios de disimilaridad. Un criterio común es:

$$\sum_{\substack{j=1 \\ j \neq k}}^d \chi^2_{jk}, \text{ donde } k=1, \dots, d; \text{ y } \chi^2_{jk}, \text{ es el coeficiente de asociación de la ji-cuadrada entre}$$

variables x_j y x_k , calculados de una tabla de 2 X 2 de una distribución marginal. La división es hecha en que k, puede maximizar esto.

Un ejemplo de la técnica polítética, es cuando los objetos son gradualmente transformados de un grupo principal a un subgrupo, seleccionando en cada paso de transferencia el objeto cuya disimilaridad del grupo principal, y tenga una menor disimilaridad al subgrupo. Esta transferencia se detiene cuando todos los miembros del grupo principal son muy similares, en comparación con

el subgrupo. El coeficiente de disimilaridad para datos binarios, es: Sea x_{Aj} la proporción de objetos en el grupo A, donde 1 esta codificado para la variable j y x_{Bj} , la proporción pareada para el grupo B. Entonces la disimilaridad entre los grupos A y B es:

$$\sum_j \left[(x_{Aj} - x_{Bj})^2 \sum_{k \neq j} \chi^2_{jk} \right]$$

Donde la χ_{jk} es calculado para el grupo combinado A+B. Los autores tentativamente sugieren que los datos cuantitativos se pueden extender al siguiente coeficiente.

$$\sum_j \left[(\bar{x}_{Aj} - \bar{x}_{Bj})^2 \sum_{k \neq j} r^2_{jk} \right]$$

*Eigenvalores y eigenvectores*⁴⁵

Los eigen valores y los eigen vectores de una matriz, son funciones especiales de los elementos de esta última sumamente importante.

1.- Los eigenvalores (o raíces características o latentes) de Σ son las raíces de la ecuación polinomial dada por:

$$\Sigma - \lambda I = 0$$

2.- A cada eigenvalor de Σ le corresponde un vector diferente de cero llamado eigenvector (vector característico o latente) que satisface $\Sigma c_i = \lambda_i c_i$, para $i=1,2,\dots, p$.

3.- Si Σ es una matriz simétrica de números reales, entonces sus eigenvalores y eigenvectores también constituirán en números reales.

4.- Los eigenvalores de Σ se denotan por $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$

Análisis de componentes principales

El análisis de componentes principales comprende es un procedimiento matemático que transforma un conjunto de variables correlacionadas de respuesta en un conjunto menor de variables no correlacionadas llamadas componentes principales. Este análisis es quizá el más útil para cribar datos multivariados. Los objetivos del análisis de componentes principales son:

- 1) Reducir la dimensionalidad del conjunto de datos.
- 2) Identificar nuevas variables significativas subyacentes.

En el proceso de este análisis se forman nuevas variables llamadas componentes principales, en orden decreciente de importancia, de modo que no estén correlacionadas, la primera componente principal explique tanto de la variabilidad en los datos como sea posible. Y que cada componente subsiguiente tome en cuenta tanto de la variabilidad restante como sea posible.

La primera componente principal se define por $y_1 = a_1'(x - \mu)$, en donde a_1 se elige de modo que la varianza de $a_1'(x - \mu)$, se maximice sobre todos los vectores a_1 que satisfagan $a_1'a_1 = 1$.

Se puede demostrar que el valor máximo de la varianza de $a_1'(x - \mu)$ entre todos los vectores a_1 que satisfacen $a_1'a_1 = 1$ es igual a λ_1 , el eigenvalor más grande de Σ , y que este máximo ocurre cuando a_1 es un eigenvector de Σ , correspondiente al eigenvalor λ_1 y que satisface $a_1'a_1 = 1$.

La segunda componente principal se define por $y_2 = a_2'(x - \mu)$, en donde a_2 se elige de modo que la varianza de $a_2'(x - \mu)$ sea un máximo entre todas esas combinaciones lineales de x que no están correlacionadas con la primera variable componente principal y tenga $a_2'a_2 = 1$.

Se puede demostrar que el valor máximo de la varianza $a_2'(x - \mu)$ entre todas las combinaciones lineales de x que tienen $a_2'a_2 = 1$ y que no están correlacionadas con y_1 es igual a λ_2 , el segundo eigenvalor mas grande de Σ y que este máximo ocurre cuando a_2 es un eigenvector de Σ correspondiente a este segundo eigenvalor más grande λ_2 , y que satisface $a_2'a_2 = 1$. De manera semejante, se puede definir componentes principales adicionales.

Con la finalidad de saber más al respecto de estas cinco variables, así como de graficar los conglomerados, que se elaborarán, se analizan dichas preguntas mediante el análisis de componentes principales en la matriz de varianzas y covarianzas; no obstante sólo por comparar las dos matrices, se corre también en la matriz de correlaciones, sin embargo ésta última no tiene ninguna utilidad en este trabajo.

En el análisis de componentes principales, los eigenvalores mayores de uno, corresponde solo al primer componente y explican el 72.28 % de la variabilidad total de los datos.

El primer componente principal, se refiere a la preocupación o al temor que se tiene al sobrepeso, en una constante lucha por mantenerse delgado(a), inclusive, la ideal de que la vida es controlada por lo que se ingiere. Además de estas ideas, se tiene el comportamiento de la realización de dietas para disminuir el peso corporal. En el segundo componente se tienen los valores elevados del ejercicio extenuante.

Principal Components: on Covariances

Eigenvalue	1.3850	0.2552	0.1981	0.0631	0.0146
Percent	72.2824	13.3213	10.3404	3.2958	0.7601
Cum Percent	72.2824	85.6038	95.9441	99.2399	100.0000

Eigenvecto					
dieta2	0.51451	0.20435	-0.32852	-0.76164	0.07415
temosobr2	0.52503	0.04435	-0.20143	0.38220	-0.73193
preodelg2	0.48434	0.06970	-0.21843	0.50591	0.67594
ejeex2	0.21758	0.65494	0.72359	0.01082	0.00228
comvida2	0.42154	-0.72281	0.52934	-0.13329	0.04330

**Multivariate
Principal Components / Factor Analysis
Principal Components: on Correlations**

Eigenvalue	3.4934	0.8050	0.5101	0.1538	0.0378
Percent	69.8673	16.1000	10.2011	3.0759	0.7556
Cum Percent	69.8673	85.9673	96.1685	99.2444	100.0000

Eigenvectors

dieta2	0.48883	-0.03526	-0.36825	0.78494	0.08981
temosobr2	0.51689	-0.10663	-0.19293	-0.33044	-0.75832
preodelg2	0.50968	-0.09696	-0.24069	-0.50835	0.64380
ejeex2	0.28287	0.92304	0.26042	-0.01279	0.00234
comvida2	0.39253	-0.35494	0.83751	0.12690	0.04909

Medidas de asociación ⁴⁶

En el proceso de investigación en las ciencias conductuales, frecuentemente se desea saber si dos series de puntuaciones están relacionadas y, si es así, el grado de su relación. Las medidas de relación que se presentan son para datos no paramétricos. En el caso paramétrico, la medida usual de correlación es el coeficiente de correlación producto-momento r de Pearson.

Confiabilidad de un instrumento

Considérese un grupo de N objetos o sujetos, cada uno de los cuales va a ser asignado a una de m categorías. Los datos de las asignaciones pueden ser colocados en una tabla $N \times m$:

Objeto	Categoría						
	1	2	...	j	...		m
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	S_1
2	n_{21}						S_2
.	\vdots						\vdots
i	n_{i1}						S_i
.	\vdots						\vdots
N	n_{N1}		...	n_{Nj}	...	n_{Nm}	S_N
	C_1	C_2	...	C_j	...	C_m	

Donde n_{ij} es el número de evaluadores que asigna el i -ésimo objeto a la j -ésima categoría. Ya que cada evaluador clasifica cada objeto, la suma de frecuencias en cada fila es igual a k . Sea C_j el número de veces que un objeto es asignado a la j -ésima categoría, lo cual es simplemente la columna sumada de frecuencias:

$$C_j = \sum_{i=1}^N n_{ij}$$

Índice de concordancia

Para encontrar el índice de concordancia o reproducibilidad entre la clasificación anteriormente descrita (cualitativa) y los grupos que se formaran mediante el análisis de cluster, con variables ordinales y binarias, se calcula el coeficiente de Kappa.¹³ Este coeficiente es la razón de la proporción de veces que los evaluadores están de acuerdo (corregida para acuerdo aleatorio), a la proporción máxima de veces que los evaluadores podrían concordar (corregida para acuerdo aleatorio), además cabe mencionar que este índice se utiliza para medir confiabilidad en variables categóricas.

$$K = \frac{P(A) - P(E)}{1 - P(E)} \dots\dots\dots (1)$$

Donde P(A) es la proporción de veces que los k evaluadores concordaran por azar. Si existe completo acuerdo entre los evaluadores, entonces K=1; mientras que si no existe acuerdo entre los evaluadores (diferentes del acuerdo que se esperaría que ocurriera por azar), entonces k=0.

Para encontrar P (E) notamos que la proporción de sujetos asignados a la j-ésima categoría es $P_j = \frac{C_j}{Nk}$. Si los evaluadores hacen sus asignaciones al azar, la proporción esperada de acuerdo para cada categoría sería P_j^2 , y el acuerdo total esperado a través de todas las categorías sería

$$P(E) = \sum_{j=1}^m P_j^2$$

La extensión del acuerdo entre los evaluadores concerniente al i-ésimo sujeto es la proporción del número de pares para los cuales existe acuerdo, a los posibles pares de asignaciones. Para el i-ésimo sujeto esto es:

$$S_i = \frac{\sum_{j=1}^m \binom{n_{ij}}{2}}{\binom{k}{2}} = \frac{1}{k(k-1)} \sum_{j=1}^m n_{ij} (n_{ij} - 1)$$

Para obtener la proporción total de acuerdo, encontramos el acuerdo de estas proporciones a través de todos los objetos evaluados:

$$P(A) = \frac{1}{N} \sum_{i=1}^N S_i = \left[\frac{1}{Nk(k-1)} \sum_{i=1}^N \sum_{j=1}^m n_{ij}^2 \right] - \frac{1}{k-1}$$

Los valores de P (E) y P (A) van a entonces combinarse usando la ecuación 1 para encontrar el estadístico kappa.

En un cuadro de dos por dos se tendría que:

La concordancia observada $P(A)=a+b/ (a+b+c+d)$

La concordancia esperada $P(E)= (a+b)(a+c) + (c+d)(b+d) / (a+b+c+d)^2$

Sensibilidad y especificidad

La sensibilidad y la especificidad son dos índices tradicionales de validez, cuando las definiciones de exposición y resultado son categóricas. El estudio de exposición o categorización de resultado es contrastado con un método utilizado (“gold standard”), el cual se asume que representa “la verdad”. La sensibilidad se refiere a la habilidad de la prueba para identificar correctamente quienes poseen la enfermedad.

Contrariamente, la especificidad se refiere a la habilidad que tiene la prueba para identificar correctamente quienes no poseen la enfermedad (sanos).

Resultado del estudio	Resultado del gold estándar		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	N

Sensibilidad = $a/(a+c)$

Especificidad = $d/(b+d)$

Correctamente clasificado = $a+d / (a+b+c+d)$

ROC (Característica de Operación Receptora):

Esta área se calcula por la regla trapezoidal, correspondiente a la probabilidad de que un par de sujetos aleatoriamente seleccionados normales o enfermos, la clasificación en cuestión, permitirá identificar a los sujetos normales dado la clasificación que se esta probando.

Cuestionario

8.79) Si has estado a dieta ¿Qué tan seguido?

- | | |
|-----------------------------|---------------------------------|
| 1) Una vez en el último mes | 2) 2 a 3 veces en el último mes |
| 3) Una vez a la semana | 4) 4 a 6 veces a la semana |
| 5) Diario | 6) Nunca |

8.81) Durante los últimos 30 días ¿Hiciste ejercicio para perder o no aumentar de peso?

- | | |
|-----------------------------|---------------------------------|
| 1) Una vez en el último mes | 2) 2 a 3 veces en el último mes |
| 3) Una vez a la semana | 4) 4 a 6 veces a la semana |
| 5) Diario | 6) Nunca |

8.82) Durante los últimos 30 días ¿Vomitaste o tomaste laxante para perder o no aumentar de peso?

- | | |
|-----------------------------|---------------------------------|
| 1) Una vez en el último mes | 2) 2 a 3 veces en el último mes |
| 3) Una vez a la semana | 4) 4 a 6 veces a la semana |
| 5) Diario | 6) Nunca |

8.83) Durante los últimos 30 días ¿Tomaste pastillas para perder o no aumentar de peso?

- | | |
|-----------------------------|---------------------------------|
| 1) Una vez en el último mes | 2) 2 a 3 veces en el último mes |
| 3) Una vez a la semana | 4) 4 a 6 veces a la semana |
| 5) Diario | 6) Nunca |

8.84) Preparo los alimentos pero no los consumo

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.85) Me aterroriza tener sobrepeso

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.86) Corto mis alimentos en pedazos muy chicos

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.87) Guardo pequeños trozos de alimentos y los voy comiendo poco a poco

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.88) Siento que otros preferirían que yo comiera más

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.89) Estoy preocupado por tratar de ser delgado (a)

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.91) Disfruto comiendo carne

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.92) Disfruto comiendo en restaurantes

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

8.93) Siento que la comida controla mi vida

- | | | |
|-------------------|-------------|------------------|
| 1) Nunca | 2) Rara vez | 3) Algunas veces |
| 4) Con frecuencia | 5) Siempre | |

Bibliografía

1. Kessler R. Special Theme Mental Health. *Psychiatric Epidemiology: selected recent advances and future directions* 2000; 471-482.
2. Kohn M., Golden NH. Eating disorders in children in adolescents: Epidemiology, diagnosis and treatment. *Paediatr Drugs* 2001;3:91-99.
3. Medina-Mora MA., Borges G., Lara C., Benjet C., Blanco J., Fleiz C. Villatoro J. Prevalencia de trastornos mentales y uso de servicios: Resultados de la Encuesta Nacional de Epidemiología Psiquiátrica en México. *Salud Mental* 2003;vol.26, No.4:1-16.
4. Berger U, Schilke C, Strauss B. Weight concerns and dieting among 8 to 12-year-old children. *Psychother Psychosom med psychol* 2005;55(7):331-338.
5. Pope H, Gruber A. Mangweth B, Bureau B, deCol C, Jouvent R, Hudson J. Body image perception among men in three countries. *Am J Psychiatry* 2000; 157:1297-1301.
6. American Psychiatric Association. Trastornos de la conducta alimentaria. En manual diagnóstico y estadístico de los trastornos mentales, 4º ed (DSM-IV). Barcelona:Masson, 1995:553-564.
7. Unikel C. Factores de riesgo en los trastornos de la conducta alimentaria. Tesis para obtener el grado de doctor en Psicología. UNAM, facultad de Psicología, División de Estudios de Postgrado 2003.
8. Unikel C, Villatoro JA, Fleiz C, Alcantar EN, Hernández SA. Conductas alimentarias de riesgo en adolescentes mexicanos. Datos en población estudiantil del Distrito Federal. *Rev Invest Clin Méx* 2000; 52(2):140-147.
9. Martínez NF. Antó JM, castellanos PL, Gili M, Maset P, Navarro V. *Salud Pública*. Madrid:McGraw-Hill-Interamericana 1998.
10. Rose G. *The strategy of preventive medicine*. Oxford university Press. Oxford. 1992.
11. Hand. *Discrimination and Classification*. John Wiley & Sons LTD. 1981. USA.
12. Szklo M, Nieto FJ. *Epidemiology*. Aspen Publishers, Inc. 2000.
13. Unikel-Santoncini C., Bojorquez-Chapela I., Carreño-García S. Validación de un cuestionario breve para medir conductas alimentarias de riesgo. *Salud Pública de México* 2004;vol.46, no.6:509-515.
14. Stata 7, reference A-G. *Statistics, Graphics, data management*. 1994;Cap. Cluster, 224-281.

-
15. Aldenderfer M, Blashfield R. Cluster analysis. Sage Publications. The international Professional Publishers, Series/number 07-044. 1991.
 16. Magnusson. Teoría de los test: Psicología diferencial, Psicología aplicada, orientación vocacional. 2da. Edición-México: Editorial Trillas 1995.
 17. Likert R. "A technique for the measurement of attitudes". Archives of Psychology 1932; 140:385-402.
 18. García-García E., Vázquez-Velázquez V., López-Alvarenga JC., Arcilla-Martínez D. Validez interna y utilidad diagnóstica del Eating Disorders Inventory en mujeres mexicanas. Salud Pública de México 2003; vol.45, no.3:206-210.
 19. Walsh T., Garner DM. Handbook of treatment for eating disorders. Diagnostic Issues. Guilford Press, second edition 1997; 25-33.
 20. Smith MC, Thelen MH. Development and validation of a test for bulimia. Journal of consulting and clinical psychology 1984;52(5): 863-872.
 21. Álvarez G., Mancilla JM., Vázquez R. Revista Psicología Contemporánea 2000; vol.7, No.1:74-85.
 22. Garner DM., Garfinkel PE. The Eating Attitudes Test: an index of the symptoms of anorexia nervosa. Psychological Medicine 1979; 9:273-279.
 23. Álvarez G., Vázquez R., Mancilla JM., Gómez-Peresmitré G. Evaluación de las propiedades psicométricas del test de actitudes alimentarias (EAT-40) en mujeres mexicanas. Revista Mexicana de Psicología 2002; vol. 19, No.1: 47:56.
 24. Vaz FJ., Peñas EM. Estudio diferencial de las formas completas y subclínicas de bulimia nerviosa. Actas Esp Psiquiatr 1999;27(6):359-365.
 25. Rodríguez A., Martínez JM., Novalbos JP:, Ruiz MA., Jiménez D. Ejercicio físico y hábitos alimentarios: un estudio en adolescentes de Cádiz. Rev Esp Salud Pública 1999; 73:81-87.
 26. Hay PJ, Fairburn CG, Doll HA. The classification of bulimic eating disorders: a community-based cluster analysis study. Psychological Medicine 1996; 26:801-812.
 27. Everitt B, Duna G. Applied multivariate data analysis. Oxford University Press, Inc., NY. 2001.
 28. Hair A. Análisis multivariado. Prentice Hall. 2000.

-
29. Lazcano-Ponce E., Hernández B., Cruz-Valdez A., Allen B., Díaz R., Hernández C., Hernández-Ávila. Chronic disease risk factors among healthy adolescents attending public schools in the state of Morelos, México. *Achieves of Medical Research* 2003; 34;222-236.
30. Espina A, Asunción-Ortego M, Ochoa de Alda I, Alemán A, Juaniz M. Body shape and eating disorders in a simple of students in the basque country: a pilot study. *Psychology in Spain* 2002;vol.6,no.1:3-11.
31. JMP User's Guide. Versión 5 of JMP. SAS Institute Inc. Cary,NC,USA. 1989.
32. Agresti A. An introduction to categorical data analysis. John Wiley & Sons, Inc. 1996.
33. Kaufman L, Rousseeuw P. Finding groups in data. Wiley series in probability and mathematical statistics.1990.
34. Manual S-Plus 6. Guide to statistics, of Windows. 2001; Vol.2
35. Jaccard P. Nouvelles recherches sur la distribution florale. *Bull Soc Vaud Sci Nat.* 1908;44,223-270.
36. Everitt B. Cluster analysis. Copublished in the Americas by Halsted Press, an imprint of John Wiley & Sons Inc. 1993.
37. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symposium.* 1967; 1:281-297.
38. Rencher A. Methods of multivariate analysis. John Wiley & Sons, Inc. 2002.
39. Thompson JK., Smolak L. Body image, eating disorders, and obesity in youth. APA 2001.
40. Gómez Perezmitré G. Detección de anomalías de la conducta alimentaria en estudiantes universitarios: obesidad, bulimia y anorexia nerviosa. *Revista Mexicana de Psicología* 1995; 10:17-27.
41. Gómez-Peresmitré G. Trastornos de la alimentación: Factores de riesgo en muestras del género masculino. *Revista de psicología Contemporánea* 1997; vol.7, número 1.

-
42. Dea JA, Abraham S. Improving the body image, eating attitudes, and behaviors of young male and female adolescents: A new educational approach that focuses on self-esteem. *Int J Eat Disord* 2000;28:43-57.
43. Bulik CM, Sullivan PF, Kendler KS. An empirical study of the classification of eating disorders. *Am J Psychiatry* 2000; 157:6.
44. Méndez I., Namihira D., Moreno L, Sosa C. El protocolo de la investigación. Lineamientos para su elaboración y análisis. 2da. Edición, Editorial Trillas 2004.
45. Johnson D. Métodos multivariados aplicados al análisis de datos. Internacional Thomson Editores, S.A. de C.V. 2000.
46. Siegel S, Castellan NJ. Estadística no paramétrica aplicada a las ciencias de la conducta. 4ta. Edición México. Editorial Trillas 1995.