

UNIVERSIDAD DE SALAMANCA  
DEPARTAMENTO DE ESTADÍSTICA

---



TESIS DOCTORAL

ANÁLISIS DE LAS RELACIONES  
ENTRE HÁBITOS DE VIDA,  
PERCEPCIÓN DEL ENTORNO Y APOYO  
SOCIAL, UTILIZANDO  
HERRAMIENTAS DE BIG DATA

ADRIANA ISABEL ANDRADE SÁNCHEZ

2017

ANÁLISIS DE LAS RELACIONES ENTRE HÁBITOS DE VIDA,  
PERCEPCIÓN DEL ENTORNO Y APOYO SOCIAL, UTILIZANDO  
HERRAMIENTAS DE BIG DATA

Memoria que para optar al Grado de Doctor,  
por el Departamento de Estadística de la  
Universidad de Salamanca, presenta:

*Adriana Isabel Andrade Sánchez*

Salamanca

2017



**VNiVERSIDAD  
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

DEPARTAMENTO DE ESTADÍSTICA

---

**M. PURIFICACIÓN GALINDO VILLARDÓN**

*Catedrática de la Universidad de Salamanca*

CERTIFICA:

Que **Doña Adriana Isabel Andrade Sánchez**, ha realizado en el Departamento de Estadística de la Universidad de Salamanca, bajo su dirección, el trabajo que para optar al grado de Doctor, presente con el título: *“Análisis de las Relaciones entre Hábitos de Vida, Percepción del Entorno y Apoyo Social, Utilizando Herramientas de Big Data”*; y para que conste, firma el presente certificado en Salamanca, en julio de 2017.

# Agradecimientos

*“Los realmente grandes te hacen sentir  
que tú también puedes ser grande”*

Mark Twain

Gracias a mi directora la *Dra. Purificación Galindo Villardón* por compartir no sólo conocimientos, sino por transmitir su pasión por el trabajo y por hacerme parte de este gran equipo. Gracias Puri por tu confianza entregada más de una vez y por hacerme soñar en grande.

Gracias a la *Dra. Ciria M. Salazar C.*, quien es una mujer de una sola pieza, a quien admiro y respeto. Por ser esa amiga, en quien no sólo confías sino que aprendes cada minuto que estás cerca de ella. Sin tu apoyo esta tesis no sería ahora una realidad.

Al *Dr. José E. Del Río Valdivia*, gracias por darme ese voto de confianza y por incluirme a su gran equipo de trabajo. Por su apoyo incondicional como jefe, colega y amigo.

Y nunca menos importante, a mi *mamá* que nunca me ha dejado sola, que me acompaña a cada paso que doy y quien ha sido mi mayor fortaleza estos años. Un ejemplo de lucha constante y de congruencia en sus actos. Gracias mamá.

A mi *papá* quien me enseñó el maravilloso mundo de la lectura, quien me formó para no caerme ante la adversidad y a levantarme si algo me derrumba en algún momento. Espero puedas estar orgulloso de mí.

A mi *hermano*, por esa tenacidad demostrada, quien lucha por lo que desea sin dejar que nada ni nadie lo detenga.

Al *Dr. Rodolfo Gallardo Rosales*, quien me ayudó a dar el primer paso del resto de mi vida. Quien me enseñó que las barreras y las fronteras sólo están en la mente.

A mis compañeros de la FCE, *Jesús, Brandy, Paty Viera, Paty López, Moni, Julio, Norma, Martitha, Ross, Lenin, Pedro, Huber, Tehua*, por su paciencia y apoyo. Ustedes son pieza clave para que yo haya finalizado este proceso.

A mis compañeritos de mesa, *Mitzi, Maribel, Nerea, Ana, Greibin, Carmen, María, Natalia*, por esas charlas de café y ese ánimo inyectado a todos los días. Por adoptarme y hacerme parte.

A mis amigos y a mi familia, que siempre están al pendiente de mí, que todos los días tienen un saludo, una caricia en forma de mensaje y una oración pidiendo a Dios que no me abandone.

¡Gracias!

*“How easily men could make  
things much better than they are  
- if the only tried together!”*

*Winston Churchill, 1909.*

**UBUNTU**

*A mis abuelos*

*A mis padres*

*A mi hermano*

# Índice

INTRODUCCIÓN .....	1
I ANTECEDENTES .....	5
I.1 APOYO SOCIAL .....	13
II OBJETIVOS E HIPÓTESIS .....	17
III MATERIAL Y MÉTODO.....	20
III.1 Población de estudio.....	21
III.1.1 Bachillerato .....	21
III.2 Sobremuestreo.....	23
III.3 Variables y constructos incluidos en el estudio.....	25
III.3.1 Peso y talla.....	27
IV INSTRUMENTOS .....	28
IV.1.1 Cuestionario de Apoyo Social Funcional (DUKE - UNC - 11)	29
IV.1.2 Esquema de Goldthorpe .....	31
IV.1.3 Índice de la Calidad de la Dieta Mediterránea (KIDMED) ...	33
IV.1.4 Cuestionario Internacional de Actividad Física (IPAQ - Corto)	
.....	35
IV.1.5 Test de Fagerström para la Dependencia de la Nicotina	
(FTND) .....	39
IV.1.6 Test de alcoholismo (CAGE) .....	41
IV.1.7 APGAR - Familiar.....	43

IV.1.8	Escala de Satisfacción con la Vida para Niños (SWLS - C) ..	46
IV.1.9	Calidad de Vida Relacionada con la Salud en Niños y Adolescentes (KIDSCREEN - 10) .....	48
IV.1.10	General Health Questionnaire (GHQ - 12).....	51
IV.1.11	Cuestionario sobre la Forma Corporal (BSQ - 8C).....	54
IV.1.12	Figuras de Stunkard .....	57
V	MÉTODOS .....	60
V.1	Sobremuestreo.....	61
V.2	Análisis de Clúster.....	65
V.2.1	Antecedentes.....	65
V.2.2	TwoStep Cluster (Clúster bietápico).....	69
V.3	Descomposición de matrices.....	83
V.4	Análisis de Segmentación .....	88
V.4.1	Antecedentes.....	88
V.4.2	Algoritmo TAID .....	93
VI	RESULTADOS.....	99
VI.1	Descripción general de la población .....	100
VI.1.1	Exploración de las relaciones entre variables sociodemográficas, hábitos de vida y percepción del entorno .....	107

VI.2 Inspección de las agrupaciones naturales de los jóvenes según sus variables socioeconómicas, hábitos de vida y la percepción de su entorno .....	110
VI.3 Búsqueda de las agrupaciones naturales de los jóvenes a través de la población recalibrada .....	114
VI.4 Exploración de las variables y constructos que tienen alta influencia en la percepción del Apoyo Social en los jóvenes ...	118
VI.5 La percepción de Apoyo Social como una respuesta a la edad de los jóvenes, su sexo, clase social, entorno y hábitos de vida. Un análisis a través de los métodos de segmentación. ....	121
CONCLUSIONES .....	129
BIBLIOGRAFÍA .....	131
ANEXO .....	164
1.1 Instrumento de estudio.....	165
1.2 Formas de evaluar el instrumento .....	174

## Índice de tablas

Tabla 1. Conformación de la muestra .....	24
Tabla 2. Instrumentos incluidos en el estudio.....	26
Tabla 3. Puntos de corte internacional para el IMC por sexo .....	27
Tabla 4. Puntos de corte para el DUKE-UNC-11 .....	30
Tabla 5. Clasificación de las clases sociales según el Esquema de Goldthorpe.....	32
Tabla 6. Puntos de corte para la clasificación de la dieta según el Índice de KIDMED.....	35
Tabla 7. Clasificaciones a las respuestas del IPAQ .....	38
Tabla 8. Clasificación a las respuestas del Test de Tolerancia de Fagerström.....	41
Tabla 9. Dictamen a las respuestas del test de CAGE.....	43
Tabla 10. Valores del Test de CAGE .....	43
Tabla 11. Definiciones de los componentes del APGAR Familiar .....	44
Tabla 12. Valores del Test APGAR-Familiar.....	45
Tabla 13. Puntos de corte para la SWLS-C .....	48
Tabla 14. Puntos de corte para el GHQ-12 .....	53
Tabla 15. Valores del Test BSQ-8.....	56
Tabla 16. Categorización del índice de discrepancia .....	59
Tabla 17. Resumen de los métodos aplicados en esta investigación ...	98

Tabla 18. Resultados de las variables sociodemográficas .....	101
Tabla 19. Resultados de los cuestionarios sobre estilos de vida .....	103
Tabla 20. Resultado de los cuestionarios que miden la percepción del entorno .....	105
Tabla 21. Asociaciones de Spearman .....	107
Tabla 22. Tabla de contingencia entre el constructo respuesta y la edad de los adolescentes.....	122
Tabla 23. Coeficientes de predictividad e índices de Catanova de todos los predictores .....	122
Tabla 24. Puntuaciones de cada categoría del predictor seleccionado.	123
Tabla 25. Clasificación de las categorías del predictor.....	124
Tabla 26. Perfiles de los nodos terminales del árbol final .....	127

## Índice de figuras

Figura 1. Organigramas del Sistema Educativo Español y Mexicano ....	22
Figura 2. Clasificaciones a las respuestas del IPAQ-Corto .....	39
Figura 3. Dimensiones de los cuestionarios KIDSCREEN .....	49
Figura 4. Proceso para llevar a cabo un análisis de clúster .....	66
Figura 5. Diferentes tipos de métodos de análisis de clúster .....	68
Figura 6. Gráfica de formación del árbol de características (CF) .....	73
Figura 7. Ejemplo de descripción de las partes del árbol CF .....	74
Figura 8. Descripción del BIRCH .....	74
Figura 9. Ejemplo gráfico del cálculo de las similitudes entre clústeres .....	81
Figura 10. Diagrama de flujo correspondiente al algoritmo de Kass (1980) .....	90
Figura 11. Esquema de los algoritmos BSA-I y BSA-II .....	91
Figura 12. Algoritmo TAID .....	97
Figura 13. Conjunto de gráficas de las variables sociodemográficas ..	102
Figura 14. Conjunto de gráficas de los constructos sobre estilos de vida .....	104
Figura 15. Conjunto de gráficas de los constructos sobre la percepción del entorno .....	106
Figura 16. Matriz de asociaciones de Spearman .....	109
Figura 17. Comparación de clústeres .....	112

Figura 18. Comparación de clústeres basados en la muestra .....	116
Figura 19. Leverage para la selección de variables .....	118
Figura 20. ACNS comparando el AS percibido y la funcionalidad de la familia .....	123
Figura 21. Primer nivel del árbol de segmentación .....	124
Figura 22. Árbol ternario de segmentación considerando hasta que ningún predictor tenga un resultado significativo .....	125
Figura 23. Árbol ternario de segmentación final.....	128

---

# INTRODUCCIÓN

---

La metodología abordada en los artículos que buscan caracterizar los hábitos de vida de los jóvenes implica en general un análisis descriptivo básico como la presentación de tablas de frecuencias o de dos entradas, gráficos de barras o pastel; para la comparación de dos variables utilizan proporciones (odd ratios), razones de momios, correlación de Pearson, test Chi-Cuadrado o T de Student; para incluir más variables utilizan la prueba de Scheffé o la regresión logística simple o múltiple para averiguar la capacidad predictiva de algunas variables incluidas en los modelos. Sin aventurarnos a señalar que sean las únicas técnicas aplicadas sí podemos asegurar que son las más comunes hasta ahora. La mayoría de investigaciones relacionadas con adolescentes y jóvenes en el ámbito de las ciencias sociales llevan un análisis cualitativo profundo, donde las técnicas provenientes del Big Data no han sido prácticamente utilizadas.

Las técnicas de análisis de datos tienen cada día más usuarios en razón de que aportan muchos elementos informativos a investigadores de muy diversas áreas de conocimiento (Castro, 2005). Con el progreso tanto del hardware como del software, se genera y recoge diariamente una gran cantidad de datos, los cuales son significativos sólo cuando se puede obtener información oculta dentro de ellos (Ming-Yi, Jar-Wen & Lien-Fu, 2010). La era de la información de hoy, desafía a los desarrolladores a crear rápidamente aplicaciones que transformen la sobrecarga de datos en información crítica que los trabajadores necesitan para tomar decisiones más inteligentes y rápidas (IBM Knowledge Center, 2001). De esta forma, en la investigación actual es necesario el análisis de bases de datos extensas y complejas, por lo que la minería de datos nace para descubrir patrones interesantes de grandes cantidades de datos dentro de fuentes limitadas (la memoria de una computadora y el tiempo de ejecución) (Ming-Yi et al., 2010).

Aunque las técnicas del Big Data fueron desarrolladas para estudiar grandes cantidades de datos en tiempo real, esto no significa que dichas herramientas no puedan ser aplicadas a contextos con una menor cantidad de datos cuyo procesamiento, en un menor tiempo resulta ventajoso para los investigadores. De esta forma se considera que la metodología desarrollada en

el ámbito del Big Data presenta una respuesta a las limitantes de los métodos tradicionales y ofrece una variedad de herramientas que es posible utilizar en el entorno de las Ciencias Sociales, donde a través de su uso se puedan caracterizar poblaciones vulnerables, hábitos de vida y características del desarrollo social.

Con la finalidad de obtener resultados a través de los cuales se puedan desarrollar y crear estrategias de actuación para mejorar el entorno de los adolescentes, el objetivo principal de esta investigación es analizar las relaciones entre hábitos de vida, percepción y apoyo social, utilizando herramientas de Big Data.

Para lograr este objetivo, esta tesis se divide en seis capítulos. El primero donde se presentan los *antecedentes*, en el cual se comienza con la exploración general del estado del arte explicando la importancia de estudiar a los jóvenes y adolescentes como una población en riesgo, presentando las características del apoyo social como una de las posibles soluciones otorgadas desde campo de las ciencias sociales.

Para continuar se presenta el capítulo de *objetivos e hipótesis*, donde se establecen tanto el objetivo general como los específicos. Asimismo, se exponen tanto la hipótesis general de trabajo como las hipótesis específicas.

En el capítulo tres, *material y método*, se describe la población de estudio y la construcción de la muestra de trabajo, a través de la cual se realizaron los análisis presentes en esta investigación. De igual forma se exponen brevemente las diferencias entre el sistema educativo español y mexicano para contextualizar a los individuos con los que se trabajó. Así también se presentan las variables y constructos presentes en el estudio.

Los *instrumentos* utilizados para medir los constructos de la presente investigación, se presentan en el capítulo cuatro. Para cada uno de ellos se detalla su origen, objetivo, forma de medición y puntos de corte para la clasificación de los jóvenes preuniversitarios.

El capítulo cinco refiere a los *métodos* utilizados para analizar las respuestas de los jóvenes preuniversitarios colimenses a cada uno de los cuestionarios aplicados. En la búsqueda de las relaciones entre el apoyo social percibido, las variables sociodemográficas, estilos de vida y la percepción del entorno se detalla el proceso para realizar el sobremuestreo estratificado, el TwoStep como método de Clúster, el procedimiento CUR para la descomposición de matrices y el análisis de segmentación a través del algoritmo TAID.

Los *resultados* se presentan en el capítulo seis, el cual comienza con la descripción de la población; continuando con la agrupación de los individuos a través del TwoStep Cluster, primero con la población completa y después con la muestra calculada previamente; en seguida se presenta la descomposición de la matriz; cerrando el capítulo con el análisis de segmentación.

Finalmente se presenta las *conclusiones* de esta tesis doctoral, seguida de la *bibliografía* utilizada, e incluyendo, en el *anexo*, el instrumento utilizado con la forma de evaluar cada cuestionario para la posible reproducción del estudio en otro contexto.

---

I

**ANTECEDENTES**

---

*Porque sumar evidencias permitirá promover y orientar políticas públicas para modificar las múltiples situaciones de carencia e inequidad que adolescentes y jóvenes enfrentan en el presente en la región [América Latina y el Caribe].*

*Porque conocer y analizar las situaciones, ya es dar los primeros pasos para cambiarlas.*

(Born, Minujín & Lombardía, UNICEF, 2015)

La adolescencia como el “periodo de la vida humana que sigue a la niñez y precede a la juventud” (RAE, 2001a) es señalada por la Organización Mundial de la Salud (OMS, 2015) entre los 10 y 19 años como una de las etapas de transición más importantes en la vida del ser humano, caracterizada por un ritmo acelerado de crecimiento y riesgos. Distinguida por drásticos cambios físicos y la afección psicológica de los mismos (Salazar, 2012), así como por innumerables problemas y tensiones, es considerada una etapa de confusión normativa, tormentosa, estresante y de oscilaciones y oposiciones (Hall, 1904 y Correrell, 1996, citados por Musitu & Cava, 2003) que provoca ambivalencias y contradicciones en el proceso de búsqueda del equilibrio consigo mismo y con la sociedad a la que desea incorporarse (Rodrigo et al., 2004).

Esta población representa aproximadamente una sexta parte de la población mundial (1200 millones de personas) (OMS, 2017). En 2010, en América Latina y el Caribe, las y los adolescentes de 10 y 19 años sumaban 110 millones, representando casi una quinta parte (18.7%) de la población. En México, en 2015 residían más de 22 millones de hombres y mujeres entre 10 y 19 años de edad según cifras del Consejo Nacional de Población (CONAPO, 2017).

La edad de 15 a 18 años particulariza la visión del fenómeno psico-social, sus características tan específicas marcan una diferencia con las demás

poblaciones (Salazar, 2012). Más allá de la maduración física y sexual, esas experiencias incluyen la transición hacia la independencia social y económica, el desarrollo de la identidad, la adquisición de las aptitudes necesarias para establecer relaciones y asumir funciones adultas junto con la capacidad de razonamiento abstracto (OMS, 2015). Constituyen un grupo de edad que modela el perfil demográfico de América Latina y el Caribe, y ofrece muchas oportunidades para el avance social, político, cultural y económico (Born, Minujín & Lombartdía, 2015). Los adolescentes pueden considerarse una “materia” moldeable y receptiva abierta a las influencias de los modelos sociales y de los entornos de vida que frecuentan, considerando que se encuentran en una etapa decisiva en la adquisición y consolidación de los estilos de vida, ya que se consolidan algunas tendencias de comportamientos adquiridos en la infancia y se incorporan nuevas provenientes de dichos entornos (Rodrigo et al., 2004).

La adolescencia y juventud (como periodos que se superponen) son etapas de oportunidad y cambios en las cuales se desarrollan las capacidades para aprender, experimentar, utilizar el pensamiento crítico, expresar su libertad creativa y participar en procesos sociales y políticos, por lo que asegurar el pleno desarrollo de estas capacidades debe ser prioridad común a todas las sociedades (Born et al., 2015); constituyendo también una etapa de riesgos considerables, durante la cual el contexto social puede tener una influencia determinante (OMS, 2015; OPS, 2016). Se calcula que en 2015 murieron 1.2 millones de adolescentes en el mundo, en su mayoría por causas prevenibles o tratables, siendo las lesiones por accidente de tránsito la principal causa de mortalidad (OMS, 2017). En México la cifra asciende a 11,578 casos en este mismo rango de edad donde los accidentes, homicidio y suicidio son las causantes del 53.44% de fallecimientos (INEGI, 2017a).

Dada la magnitud de los cambios por los que atraviesan y la todavía escasa experiencia vital de los adolescentes, muchos eventos vitales pueden adquirir un carácter estresante y tener diversos efectos psicológicos (Barra, Cerna, Kramm & Véliz, 2006). La mitad de todos los trastornos de salud mental en la edad adulta empiezan a manifestarse a los 14 años, pero la mayoría de

los casos no se detectan ni son tratados (OMS, 2017). De acuerdo con Born et al. (2015) alrededor del 2010, 17.4% de los estudiantes de 13 a 15 años de más de veinte países de América Latina y el Caribe pensó seriamente suicidarse. En México según datos del Instituto Nacional de Estadística y Geografía (INEGI, 2017b) para los jóvenes entre 15 y 24 años de edad, las lesiones autoinfligidas intencionalmente son la tercera causa de mortalidad. Donde la violencia, la pobreza, la humillación y el sentimiento de desvalorización son factores que pueden aumentar el riesgo de padecer problemas de salud mental (OMS, 2017).

Por otro lado, el consumo de sustancias nocivas entre los adolescentes es una preocupación de la OMS (2017), de acuerdo con la Encuesta Nacional de Adicciones (ENA) (INPRF, INSP, SS, Medina-Mora et al., 2011), el abuso de sustancias psicotrópicas, constituye uno de los problemas de salud pública de nuestra época, que se presenta tanto a nivel nacional como internacional, afectando, sin distinción de género e incidiendo principalmente en niños y adolescentes, de cualquier estrato social y de todas las regiones de México. El consumo de alcohol en esta edad reduce el autocontrol y aumenta los comportamientos de riesgo, como las relaciones sexuales no protegidas o comportamientos peligrosos en la carretera; siendo una de las causas de lesiones, violencia y muertes prematuras (OMS, 2017). Aquellos que empiezan a consumir alcohol o cualquier otra sustancia psicoactiva en las primeras etapas de la adolescencia (12 a 14 años de edad) tienen más probabilidades de desarrollar abuso o dependencia a largo plazo, que aquellos que empiezan en edad adulta; cuánto más temprano se inicia el consumo de alcohol, mayor es el riesgo de que, con el paso del tiempo, tenga diversos efectos nocivos para la salud (CICAD, SSM & OEA, 2015). Poco más de la mitad de la población mexicana (55.2%) reporta haber iniciado el consumo de alcohol a los 17 años o menos según la ENA (INPRF, INSP, SS, Medina-Mora et al., 2011), donde además entre la población adolescente el consumo de bebidas alcohólicas aumentó significativamente, el 14.5% de los jóvenes entre 12 y 17 años tienen un consumo alto; 0.2% un consumo diario; y el 4.1% presentan dependencia alcohólica.

“El tabaco es el único producto legal que mata hasta la mitad de sus consumidores” (OPS, 2016, p. 5). El consumo del tabaco es un factor de riesgo pues es responsable de casi dos tercios de las muertes a nivel mundial, asociada a las defunciones por enfermedades no transmisibles; el tabaco es responsable de 16% de las muertes por enfermedades cardiovasculares, 25% de las defunciones por cáncer y 52% de los decesos por enfermedades crónicas respiratorias (OPS, 2016). La gran mayoría de personas que consumen tabaco hoy en día comenzaron a hacerlo cuando eran adolescentes (OMS, 2017). La ENA (INPRF, INSP, SS, Reynales-Shigematsu, et al., 2011) reporta una prevalencia de consumo activo de tabaco en 17.3 millones de mexicanos, de los cuales 1.7 millones son adolescentes entre 12 y 17 años, donde el 2% fuman todos los días (en promedio 4.1 cigarros al día), iniciando su consumo en promedio a los 14.1 años, siendo similar el porcentaje entre hombres y mujeres.

El consumo de drogas entre los jóvenes de 15 a 19 años también es una fuente de preocupación importante a nivel mundial (OMS, 2017). Considerada como una conducta vinculada al proceso de desarrollo, ya que la mayor parte del consumo de drogas comienza en la adolescencia (UNODC, 2015) donde las causas más comunes de defunción bajo la influencia de alguna droga fueron ocasionadas por arma de fuego, accidentes de tránsito y asfixia (ENA, 2011). Se calcula que en el mundo un total de 246 millones de personas entre 15 y 64 años usó una droga ilícita en 2013 (UNODC, 2015). En México en general las prevalencias de consumo de drogas ilegales entre adolescentes (12 a 17 años) es de 1.5%, donde la marihuana es la más prevalente seguida de la cocaína y los inhalables; entre las drogas de inicio más reportadas están el alcohol (49.1%), seguida del tabaco (18.4%) y la marihuana (17.8%) (INPRF, INSP, SS, Villatoro-Velázquez et al., 2011).

En cuanto a actividades de riesgo se refiere, la mayor atención e inversión en la primera infancia se ha traducido en una generación de adolescentes más saludables en términos generales, no obstante, muchos factores atentan contra la supervivencia y la salud de los adolescentes y jóvenes, entre los que se destacan los problemas nutricionales, como el sobrepeso (Born et al., 2015). El número de adolescentes con exceso de peso y

obesidad está aumentando en los países de ingresos bajos y altos; en el lado opuesto muchos niños y niñas de países en desarrollo padecen desnutrición cuando llegan a la adolescencia, lo que los hace más propensos a contraer enfermedades y morir a una edad temprana (OMS, 2017). En México el aumento de la masa grasa, que se manifiesta en exceso de peso es el problema nutricional más frecuente en la población escolar, adolescente y adulta, de acuerdo a los resultados de la Encuesta Nacional de Salud y Nutrición (ENSANUT), en adolescentes de entre 12 y 19 años la prevalencia de sobrepeso es de 22.4% y obesidad de 13.9%; además se observan bajas proporciones de consumidores regulares de grupos de los alimentos recomendables (agua sola, leguminosas, lácteos, carnes, huevo, frutas y verduras) y elevadas de adolescentes que consumen grupo no recomendables (bebidas no lácteas endulzadas, botanas, dulces, postres, carnes procesadas y comida rápida) (INSP & SS, 2016).

De igual forma, se ha observado que la inactividad física es el cuarto factor de riesgo en lo que respecta a la mortalidad mundial (6% de las muertes registradas en todo el mundo), además se estima que la inactividad física es la causa principal de aproximadamente un 21 a 25% de los cánceres de mama y de colon, el 27% de los casos de diabetes y aproximadamente el 30% de la carga de cardiopatía isquémica (OMS, 2013a). Con el fin de mejorar las funciones cardiorrespiratorias y musculares y la salud ósea y de reducir el riesgo de enfermedades no transmisibles, se recomienda que los niños y jóvenes de 5 a 17 años inviertan como mínimo 60 minutos diarios de actividad física de intensidad moderada a vigorosa (OMS, 2013b). Desarrollar en la adolescencia unos buenos hábitos de alimentación sana y ejercicio físico es fundamental para gozar de una buena salud en la edad adulta, sin embargo, los datos reflejan que menos de uno de cada cuatro adolescentes sigue las directrices recomendadas sobre actividad física (OMS, 2017). En México la prevalencia de actividad física suficiente (420 minutos de actividad física moderada/vigorosa los 7 días de la semana) en adolescentes, aumentó de 56.7% a 60.5% de 2012 a 2016 (INSP & SS, 2016).

Es indudable que se han realizado avances que favorecen a las y los adolescentes en América Latina y el Caribe sin embargo, aún quedan deudas pendientes (Born et al., 2015). Dada la importancia que representa este grupo poblacional para el desarrollo de México, se hace necesario un abordaje holístico, a través del cual se haga frente a los riesgos a los que se encuentran expuestos (Santos-Preciado et al., 2003). Resumiendo lo señalado anteriormente, en México, los adolescentes entre 15 y 18 años son una población vulnerable con altos índices de sobrepeso y obesidad, dietas poco balanceadas y trastornos alimentarios, hipoactividad física y con consumo de sustancias dañinas (alcohol, tabaco y drogas), suicidios, muertes prematuras por enfermedades degenerativas, decesos violentos, entre otros problemas particulares de esta edad. El bienestar y la calidad de vida, son esenciales para la identificación del problema y la solución del mismo (Salazar, 2012).

Las variables que han sido tradicionalmente estudiadas en las investigaciones sobre estilos de vida en la adolescencia (hábitos alimenticios, ejercicio físico, deporte, consumo de tabaco, alcohol, drogas ilegales, etc.), más recientemente se han incorporado otras variables que apunta más directamente a la protección de la salud y a la prevención del riesgo, incorporando indicadores de competencias socio-cognitivas, tanto individuales como relacionales, que podrían ir asociadas a los estilos de vida adoptados en la adolescencia (percepción de salud, bienestar psíquico, satisfacción con la imagen corporal, entorno familiar y escolar, entre otros) (Rodrigo et al., 2004). “Las investigaciones sobre la juventud en México se han ido incorporando paulatinamente a las distintas discusiones teórico-metodológicas generadas por las distintas disciplinas sociales, a través de las cuales se reconoce a los jóvenes como actores sociales” (Mendoza, 2011, p. 193).

En esta tendencia, a nivel nacional en México, los análisis con la perspectiva hacia los adolescentes y jóvenes son llevados a cabo por el Instituto Mexicano de la Juventud (IMJUVE), quienes a través de la Encuesta Nacional de la Juventud presenta cifras sobre temas de hogar y dinámica familiar (arreglo residencial); salud (acceso a servicios de salud pública, tabaco, alcohol y actividad física); sexualidad (inicio de actividad sexual, conocimiento sobre

prevención de enfermedades de transmisión sexual, embarazos); actividad laboral (condición de actividad, satisfacción, inactividad, emprendimiento); y relaciones sociales (acceso y uso de internet, actividades de tiempo libre, noviazgo, religión, participación en organizaciones y asociaciones, confianza en instituciones) (IMJUVE, 2010).

Este mismo instituto lleva a cabo la Encuesta Nacional de Valores en Juventud donde además de nivel de estudios (de ellos y sus padres), actividad laboral y actividad sexual, se busca conocer percepción sobre la educación; empleo; factores importantes en la vida de los jóvenes (familia, trabajo, pareja, dinero, escuela, amigos, religión, política, etcétera); creencias (violencia, familia, educación, aborto, sexualidad, adopción); satisfacción y retos para el futuro; visión del país y sociedad; acceso y consumo de información; participación en la política; percepción de seguridad; medio ambiente; y ocio y tiempo libre (IMJUVE & UNAM, 2012).

“Respecto a las aportaciones en torno al conocimiento de la juventud [en México], sobresalen los trabajos de García (1995), Payá (1998) y Reguillo (1998) en cuanto a la organización juvenil. Por otra parte, son relevantes las ideas de Feixa (1998) en relación con las culturas juveniles. Aunado a ello, son fundamentales las contribuciones de autores como Brito (2002), Reguillo (2000) y Valenzuela (1997, 2009) sobre las identidades juveniles” (Mendoza, 2011, p. 214).

En México, las discusiones en torno a la juventud han sido muy prolíficas, entre algunas de las investigaciones de los últimos años enfocada en adolescentes y sus hábitos de vida se encuentran las que tienen relación con factores de riesgo asociados al sobrepeso y obesidad (Salazar, Manzo & Vargas, 2010; Trujillo-Hernández et al., 2010; Trujano, De García, Nava & Limón, 2014), suicidios (Palacios, Sánchez & Andrade, 2010; Sánchez-Cervantes, Serrano-González & Márquez-Caraveo, 2015), conductas de riesgo (Huitrón-Bravo et al., 2011; Páramo, 2011), entorno familiar y sus repercusiones (Ruiz-Risueño, Ruiz-Juan & Zamarripa, 2012), habilidades para la vida y la prevención con consumos nocivos (Pérez, 2012), bienestar subjetivo y estilos de

afrontamiento de estrés (Verdugo-Lucero et al., 2013), actividad física (Cortés, Castro-Pozo & Salazar, 2015; Hidalgo-Rasmussen, Ramírez-López & Hidalgo-San Martín, 2013; Salazar et al., 2016) entre muchas otras.

Salazar (2012) y García (2015) amplían el análisis del contexto de los adolescentes mexicanos incluyendo variables no sólo sociodemográficas (género, edad, clase social) sino también de hábitos de vida (actividad física, patrón de alimentación, consumos nocivos), percepción del entorno (tipo de familia, satisfacción con la vida, calidad de vida relacionada con la salud, salud mental) y de autoconcepto (preocupación por la figura e imagen corporal).

Tener un contexto general de los hábitos de vida y entorno en el que se desarrollan los adolescentes como primer paso para ofrecer soluciones a los problemas que aquejan a este grupo poblacional es uno de nuestros objetivos. Por lo que en la búsqueda de resultados comparables con investigaciones realizadas previamente y sobre todo en soluciones capaces de impactar a esta población se analizaron a los adolescentes preuniversitarios de acuerdo a sus hábitos de vida y la percepción de su entorno.

## **I.1 APOYO SOCIAL**

Los adolescentes dependen de su familia, su comunidad, su escuela, sus servicios de salud y su lugar de trabajo para adquirir toda una serie de competencias importantes que pueden ayudarles a hacer frente a las presiones que experimentan haciendo una transición satisfactoria de la infancia a la edad adulta (Wilkinson & Marmot, 2003). Fomentar relaciones de atención y cariño entre padres e hijos en una etapa temprana de la vida, propia para el desarrollo de aptitudes para la vida, reduce el acceso al alcohol y armas de fuego, mismo que puede contribuir a la prevención de lesiones y defunciones como consecuencias de la violencia (OMS, 2017). Por lo que la funcionalidad de la familia representa un factor determinante en la conservación de la salud o en la aparición de la enfermedad entre sus miembros (Maddaleno, 1986).

Como consecuencia a las críticas al modelo médico tradicional, donde se defiende que el origen intrapsíquico de los trastornos mentales, resaltando exclusivamente el papel de los factores biológicos en el desarrollo de las patologías, el interés de los investigadores se ha dirigido hacia el estudio de los factores psicosociales potencialmente capaces de influir en la salud, bien ejerciendo un efecto beneficioso, bien mostrando propiedades inductoras del estrés (Barrón, 1996). La importancia de ubicar al individuo como un ser social inserto en múltiples relaciones que complejizan tanto los problemas como las soluciones de las situaciones, que el individuo enfrenta de manera cotidiana, es cada vez más reconocido (Estrada et al., 2009).

“La hipótesis de amortiguamiento sugiere que el apoyo social puede moderar los impactos de los eventos de la vida sobre la salud mental” (Thoits, 1982, p. 145). Diversos estudios han señalado que la red de apoyo tiene un importante impacto sobre la salud (Asher, Hymel & Renshaw, 1984). Aunque existe poca información sobre el mecanismo a través del cual el Apoyo Social (AS) ejerce su efecto beneficioso sobre la salud, se considera la existencia de dos posibles vías de acción, una sobre el papel modificador, donde el AS mitiga el efecto de los estresores sociales y la enfermedad; y el otro donde sugieren que la ausencia de AS puede actuar como estresor en sí mismo, influyendo directamente sobre la salud (Bellón, Delgado, Luna, y Lardelli, 1996a).

“El AS puede proteger a las personas en crisis de una amplia variedad de estados patológicos: desde el bajo peso al nacimiento hasta la muerte, desde la artritis hasta la tuberculosis hasta la depresión, el alcoholismo y el síndrome de ruptura social. Además, el AS puede reducir la cantidad de medicación requerida, acelerar la recuperación y facilitar el cumplimiento de los regímenes médicos prescritos” Cobb (1976, p. 300).

De esta forma Bellón et al. (1996a) consideran sustancial la relación donde el AS ejerce un efecto beneficioso sobre la salud, dado que “su manejo posibilita la intervención preventiva, terapéutica y rehabilitadora de diferentes problemas de salud, por ejemplo, fortaleciendo las relaciones intrafamiliares o potenciando grupos de autoayuda” (p. 154).

Existen variadas definiciones y aproximaciones al concepto de AS, sin llegar a una consensuada, por la cantidad de acciones que abarca el término, y la complejidad de los múltiples aspectos interrelacionados (Estrada et al., 2009). Gracia (1997) considera que el concepto de AS es un metaconcepto por las múltiples dimensiones que integra (individuos, grupos, sociedad, características individuales y sus relaciones). Thoits (1982) señalaba como había sido conceptualizado y operacionalizado de forma inadecuada debido a que las relaciones entre los acontecimientos, el AS y la perturbación psicológica no se habían aclarado teóricamente, pasando por alto la posibilidad que el AS fuera un factor etiológico<sup>1</sup>.

Entre las definiciones más representativas se encuentra Cobb (1976) quien explica el AS como la información que lleva al sujeto a creer que es cuidado y amado, estimado y miembro de una red de obligaciones mutuas. Lin, Dean & Ensel (1979) señalan que se trata del “apoyo accesible a un individuo a través de lazos sociales con otros sujetos, grupos, una comunidad más grande” (p. 109). Thoits (1982) lo define como el grado en que las necesidades sociales básicas de la persona son satisfechas a través de la interacción con otros, entendiendo por necesidades básicas la afiliación, el afecto, la pertenencia, la seguridad y la aprobación”.

Lin, Dean & Ensel (2013) redefinen el concepto adoptado previamente y lo refieren como el conjunto de provisiones expresivas o instrumentales - percibidas o recibidas- proporcionadas por la comunidad, las redes sociales y las personas de confianza. Felton & Shinn (1992) resaltan el hecho de que se trata de un concepto que en cualquier caso refiere a transacciones entre personas; de esta forma, Barrón (1996) tiene una de las definiciones más completas y lo denomina como el conjunto de aspectos cualitativos o funcionales del apoyo, siendo estas funciones las desempeñadas por lo que se conoce como red de apoyo o red social.

---

<sup>1</sup> La etiología es el estudio sobre las causas de las enfermedades (RAE, 2001b).

Broadhead et al. (1983) consideran que la calidad del AS es mejor predictor de la salud que el apoyo estructural. Gottlieb (1988) citado por (Barrón, 1996) señala la relevancia de estudiar el AS dado el impresionante volumen de evidencia a favor de los efectos positivos que sobre la salud y el bienestar juega el AS moderando el impacto de otras fuerzas que actúan sobre la salud; la disminución de costos debido a que las intervenciones basadas en AS son mucho más eficaces; su validez ecológica, siendo el círculo informal el lugar donde se busca en primera instancia ayuda; además de ser especialmente adecuadas para un elevado porcentaje de problemas de salud mental que implican ruptura o un desajuste en las relaciones sociales habituales; siendo no sólo las personas que sufren directamente los problemas los que requieren ayuda, sino también los que se relacionan con ellos.

Aceptada la relevancia del tema, su importancia en la investigación actual y su destacado papel en las intervenciones, es evidente la necesidad de adentrarse en el estudio del AS en los jóvenes. Ya que siguiendo a la OMS (2017) “fomentar comportamientos saludables durante la adolescencia y adoptar medidas para proteger mejor a los jóvenes contra los riesgos sanitarios es fundamental para la prevención de problemas de salud en la edad adulta, así como para la salud futura de los países y su capacidad para desarrollarse y prosperar” (parr.3). De esta forma el AS es el enfoque a través del cual se analizan los resultados obtenidos.

---

**II**

**OBJETIVOS E**

**HIPÓTESIS**

---

Mejorar las vidas de adolescentes y jóvenes exige comprenderlos en toda su amplitud y diversidad, se hace necesario conocer de cerca los problemas y necesidades de los adolescentes para poder establecer con ellos un diálogo sincero, realista y fructífero (Born et al., 2015). Este proyecto de investigación no busca ofrecer conclusiones definitivas sino puntos de partida, ampliando las opciones de análisis tradicional a los datos. Mostrando como técnicas ampliamente utilizadas en otros contextos en el ámbito del Big Data pueden aplicarse perfectamente en el análisis de fenómenos sociales como los que involucran los hábitos de vida de los adolescentes y jóvenes, aunque no se trate de bases de datos tan extensas ni sean analizadas en tiempo real.

Con la finalidad de que a través de los resultados obtenidos se puedan desarrollar y crear estrategias de actuación para mejorar el entorno de los adolescentes, el objetivo principal de esta investigación es:

*Analizar las relaciones entre hábitos de vida, percepción del entorno y apoyo social, utilizando herramientas de Big Data.*

Donde los objetivos específicos son:

1. Caracterizar a los jóvenes estudiantes preuniversitarios con métodos tradicionales de análisis descriptivo, de acuerdo a sus características sociodemográficas (sexo, edad y clase social), estilos de vida (IMC, patrón de alimentación, actividad física, adicción al tabaco, consumo de bebidas alcohólicas) y entorno en el que se desarrollan (función familiar, satisfacción con la vida, calidad de vida relacionada con la salud, salud mental, preocupación por la figura e imagen corporal).
2. Detectar las variables que tienen alta influencia en la percepción de apoyo social en los jóvenes preuniversitarios colimenses.
3. Organizar y agrupar de manera eficiente los resultados para que su representación pueda caracterizar a los jóvenes, donde el resultado tenga una interpretación teórico-práctica como herramienta que permita proponer intervenciones eficaces posteriores.

Esta investigación está planteada a partir de la siguiente hipótesis general:

*La metodología utilizada en Big Data ofrece una variedad de herramientas que es posible utilizar en el entorno de las Ciencias Sociales, mostrando una respuesta a las limitantes que presentan los métodos tradicionales.*

La cual busca comprobarse a través de las siguientes hipótesis específicas:

1. El uso de métodos tradicionales para describir las relaciones entre hábitos de vida, entorno y apoyo social de los jóvenes y adolescentes, permiten el análisis cualitativo profundo a través de relaciones bivariantes de las características analizadas. Sin embargo, los cálculos se complican cuando se trabaja con grandes cantidades de datos y las conclusiones generadas a través de ellas no permiten crear estrategias eficaces de actuación que permitan hacer intervenciones para mejorar el entorno.
2. Los métodos supervisados y los no supervisados del Data Mining (árboles de clasificación, técnicas factoriales y métodos de clúster) permiten determinar las variables sociodemográficas, de hábitos de vida y de percepción que tienen alta influencia en apoyo social observado por los jóvenes preuniversitarios.
3. El algoritmo TAID como metodología propuesta para el análisis de grandes bases de datos con variables de tipo mixto permite organizar de manera eficiente las características de los jóvenes, donde sus resultados tienen una interpretación teórico-práctica para la propuesta de intervenciones eficaces posteriores.

---

**III**

**MATERIAL Y**

**MÉTODO**

---

### III.1 Población de estudio

Se trata de un estudio transversal, con un diseño cuantitativo descriptivo y un análisis empírico. Esta investigación está basada en un censo poblacional de los alumnos de primer año de bachillerato de las escuelas subsidiadas por el Estado y la Federación del Estado de Colima, México. Fueron excluidos del estudio los jóvenes que presentaban alguna patología cardíaca y quienes manifestaron gestación en cualquier estadio. La población de estudio final está conformada por 1,267 adolescentes de doce escuelas del estado de Colima, México 589 (46.5%) hombres y 678 (53.5%) mujeres, con una media de 16.5 años de edad ( $\pm 0.8$ ).

En la recolección de datos todos los jóvenes fueron enterados del anonimato de sus respuestas, así como el uso de los datos obtenidos, exclusivamente para tareas de investigación científica.

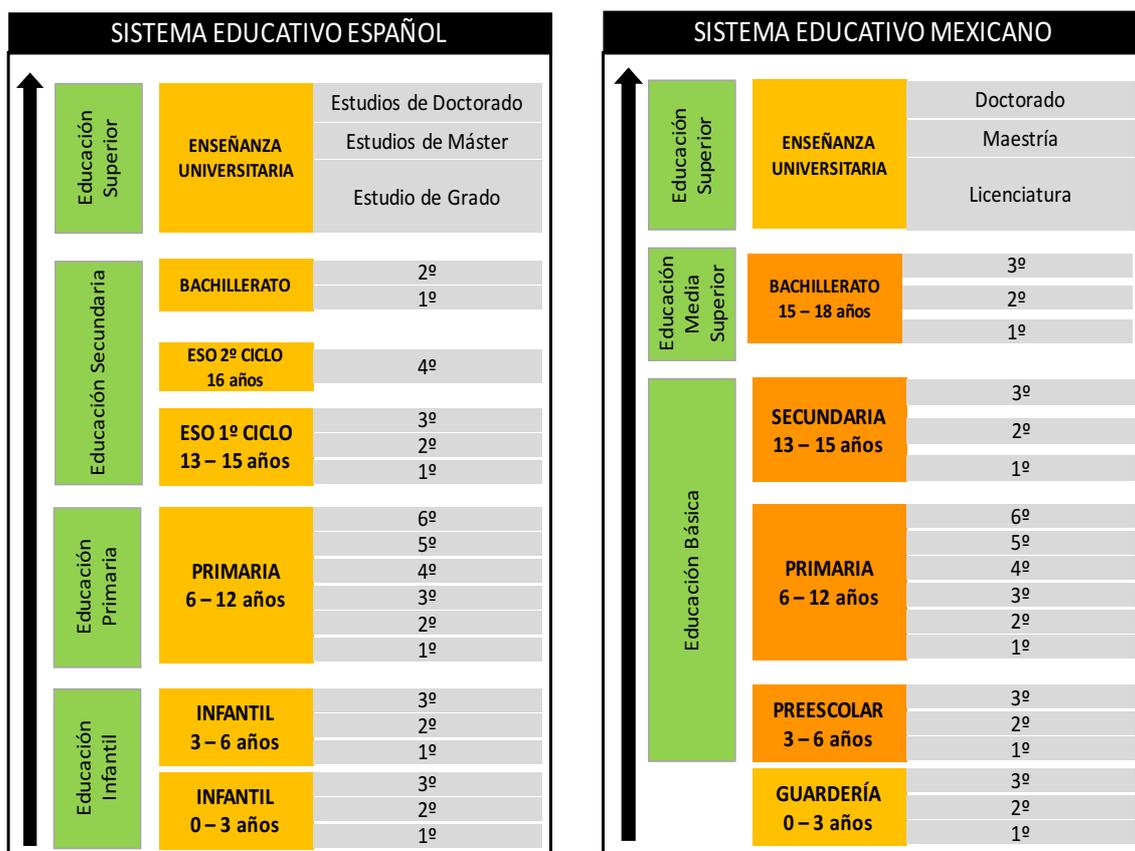
#### III.1.1 Bachillerato

Educación Media Superior, preparatoria, estudios preuniversitarios o bachillerato son expresiones sinónimas utilizadas en México para describir los tres años de estudio que preceden a los estudios de Licenciatura en el Sistema Educativo mexicano. De acuerdo con el artículo 3º de la Constitución Política de los Estados Unidos Mexicanos, en el Título Primero, Capítulo I De los Derechos Humanos y sus Garantías, el acceso hasta este nivel es gratuito y obligatoria la educación en México.

“Artículo 3. Toda persona tiene derecho a recibir educación. El Estado -Federación, Estados, Ciudad de México y Municipios-, impartirá educación preescolar, primaria, secundaria y media superior. La educación preescolar, primaria y secundaria conforman la educación básica; esta y la media superior serán obligatorias” (Diario Oficial de la Federación, 2016).

Los dos sistemas educativos, el mexicano y el español, son muy similares en cuanto a estructura, ya que deben pasar mínimo 15 años de educación previa a los estudios universitarios. Como puede verse en la figura 1, lo que en México se conoce como estudios preuniversitarios o bachillerato, equivale al segundo CICLO del ESO (Educación Secundaria Obligatoria) más el Bachillerato en España.

Figura 1. Organigramas del Sistema Educativo Español y Mexicano



Fuente: Elaboración propia.

Hasta hace unos años la Educación Media Superior en México no era obligatoria, no fue hasta el año 2011 que el Congreso de la Unión elevó a rango constitucional la obligatoriedad de este nivel educativo. Por tal motivo es que la oferta de este nivel educativo, aunque dependiente directamente de la Secretaría de Educación Pública, era ofertada en gran parte por las Universidades, públicas y privadas, como un paso preparatorio a la Educación Superior. En algunas poblaciones donde el sistema universitario no alcanzaba a

la población, el Gobierno Federal otorgaba opciones para cursar este nivel educativo, donde en su mayoría ofertaban, además de la educación preuniversitaria, la enseñanza de algún oficio que permitiera a los jóvenes trabajar al egresar de sus escuelas.

“A diferencia de las opciones universitarias [bachilleratos auspiciados por universidades locales], las preparaciones tecnológicas, agropecuarias y a distancia son pagadas por un fideicomiso nacional, dirigido a la capacitación intermedia de recursos humanos que posiblemente no alcanzarán la formación universitaria. El respaldo económico principalmente se ha convertido en la causal de elección de estas opciones profesionales, regularmente el nivel socioeconómico es bajo, los estudiantes estudian y trabajan, la figura materna es la cabeza de familia y el nivel académico de los padres y madres es mínimo” (Salazar C., 2012, p. 106).

### III.2 Sobremuestreo

Las características especiales de los datos que conllevan una distribución de frecuencias que carga en una de las colas de la distribución, puede ser la razón por la que el análisis de la muestra completa, con técnicas que presuponen una distribución mas homogénea, nos lleve a conclusiones poco interesantes en la práctica. Por esta razón y tras consultar la literatura que aborda una problemática similar en otros campos de la Ciencia (especialmente en estudios de Credit Scoring), se optó por llevar a cabo un submuestreo estratificado previo al análisis de los datos. A través del método de sobremuestreo estratificado (Siddiqi, 2006) que se detallarán más adelante, se obtiene una muestra de trabajo, construida a partir de la población que interesa analizar. Es decir los jóvenes estudiantes con EAS para evitar que, al realizar un análisis las características de estos queden ocultas por la gran proporción de adolescentes con ASF. De esta forma, para obtener estimaciones

razonablemente precisas de las propiedades de estos individuos, se toman en cuenta todos los estudiantes del grupo con EAS y al menos, el doble de jóvenes con ASF seleccionados aleatoriamente. La muestra final está conformada por 591 jóvenes tal como muestra la tabla 1.

**Tabla 1.** Conformación de la muestra

	<b>N</b>	<b>Sobremuestreo</b>	<b>Peso</b>
Escaso Apoyo Social	<b>195</b>	<b>195</b> (33%)	1
Apoyo Social Funcional	<b>1072</b>	<b>396</b> (67%)	2.707
<b>N</b>	<b>1267</b>	<b>591</b> (100%)	

**Fuente:** Elaboración propia con base en la metodología de Mallo (2011) y Siddiqi (2006).

La “nueva” muestra para trabajar contiene los 195 jóvenes con EAS y 396 con ASF con una peculiaridad: se considera cada observación de joven con EAS como representativa de una cantidad o peso en la población total, por lo que utilizando el peso (WOE) para ponderar los casos de jóvenes con ASF se puede recuperar la tasa real de jóvenes con EAS de la población. La variable peso, se construye a partir del cociente del número total de jóvenes con ASF partido por la cantidad de jóvenes con ASF incluidos en la muestra, por lo que el valor del peso resultó ser de 2.707. De esta forma, esta variable indica a cuántos casos de la población real representa cada caso de la muestra, garantizando así la representatividad de la selección. La muestra pesada consta de 1071.97 observaciones con una tasa de EAS de 15.4% aproximadamente, lo cual representa el comportamiento real de la población.

### III.3 Variables y constructos incluidos en el estudio

La caracterización de los jóvenes se realizó a través de las variables sociodemográficas (sexo, edad y clase social), el peso y la talla, expresados en el Índice de Masa Corporal (IMC), y las respuestas a algunos instrumentos ampliamente utilizados, fiables y válidos en el contexto nacional e internacional, considerando los siguientes constructos:

- Estilos de vida
  - Patrón de alimentación
  - Actividad física
  - Adicción al tabaco
  - Consumo de bebidas alcohólicas
- Percepción del entorno
  - Tipo de familia
  - Satisfacción con la vida
  - Calidad de Vida Relacionada con la Salud (CVRS)
  - Salud mental
  - Preocupación por la figura
  - Satisfacción con la imagen corporal

La selección de los instrumentos fue realizada en la búsqueda de proyecciones a esta investigación, de tal forma que los resultados obtenidos fueran comparables con estudios previos (o futuros) y otras poblaciones no sólo del contexto nacional. Cada uno de los instrumentos aplicados se presenta en la tabla 2, donde se muestra la categoría en la que se clasificó, el constructo que busca medir y los autores que lo proponen. Posteriormente cada uno de los cuestionarios utilizados es presentado detalladamente.

Tabla 2. Instrumentos incluidos en el estudio.

<b>Constructo</b>	<b>Test utilizado</b>	<b>Autor (es)</b>
Percepción del Apoyo Social	<i>DUKE-UNC-11</i>	Broadhead, Gehlbach, De Gruy & Kaplan (1988)
Clase Social	<i>Escala de Goldthorpe</i>	Goldthorpe (1997)
Calidad de alimentación	<i>KIDMED</i>	Serra-Majem et al. (2002)
Actividad Física	<i>IPAQ - Corto</i>	IPAQ Research Committee (2002)
Adicción a la nicotina	<i>Test de Fagerström</i>	Fagerström (1978)
Consumo de bebidas alcohólicas	<i>Test de CAGE</i>	Ewing (1984)
Función familiar	<i>APGAR - Familiar</i>	Smilkstein (1978)
Satisfacción con la vida	<i>SWLS - C</i>	Diener, Emmons, Larsen & Griffin (1985)
Calidad de vida relacionada con la salud	<i>KIDSCREEN-10</i>	Ravens-Sieberer et al. (2005)
Salud mental	<i>GHQ - 12</i>	Goldberg et al. (1997)
Preocupación por la figura	<i>BSQ - 8C</i>	Cooper, Taylor, Cooper & Fairbum (1987)
Satisfacción con la imagen corporal	<i>Figuras de Stunkard</i>	Stunkard & Stellar (1990)

Fuente: Elaboración propia.

### III.3.1 Peso y talla

Para el peso corporal de los adolescentes se utilizaron básculas marca “Nuevo León” con capacidad para 160 kg. con los participantes utilizando el mínimo de ropa, cuidando una posición central y simétrica en la plataforma de la báscula.

La estatura fue medida siguiendo el protocolo de la International Society for the Advancement of Kinanthropometry (ISAK, 2010) por lo que todas fueron realizadas en un horario similar, donde se utilizó un estadímetro adosado a una pared para que los y las adolescentes se pudieran alinear verticalmente de manera apropiada.

Con estos valores se calculó el Índice de Masa Corporal ( $IMC = \text{peso}/\text{talla}^2$ ) catalogando los resultados, de acuerdo a la clasificación internacional del IMC en jóvenes menores de 18 años recomendada Cole, Bellizzi, Flegal & Dietz (2000) mostrada en la tabla 3.

**Tabla 3.** Puntos de corte internacional para el IMC por sexo

Edad	IMC: 25kg/m <sup>2</sup>		IMC: 30kg/m <sup>2</sup>	
	Hombres	Mujeres	Hombres	Mujeres
15	23.29	23.94	28.30	29.11
16	23.90	24.37	28.88	29.43
17	24.46	24.70	29.41	29.69
18	25	25	30	30

**Fuente:** Elaboración propia con base en Cole et al. (2000).

---

**IV**

**INSTRUMENTOS**

---

### IV.1.1 Cuestionario de Apoyo Social Funcional (DUKE-UNC-11)

El Cuestionario de Apoyo Social Funcional, de acuerdo con Broadhead, Gehlbach, De Gruy & Kaplan (1988), fue diseñado en 1982, modificado y validado por el mismo autor y tiene como propósito medir la percepción de un individuo sobre la cantidad y el tipo de apoyo social personal. Consta de 11 ítems que permiten conocer junto al apoyo total percibido, el Apoyo Emocional o Afectivo (demostración de cariño y empatía) y el Apoyo Confidencial (posibilidad de contar con personas con las que comunicarse) (Rodríguez et al., 2004). Es decir, el Duke-UNC-11 valora el Apoyo Social Funcional en sus dos dimensiones: el Apoyo Confidencial y el Afectivo (Álamo et al., 1999).

El instrumento original contaba con 14 ítems, agrupados en 4 subescalas: Cantidad de Apoyo, Apoyo Confidencial, Apoyo Afectivo y Apoyo Instrumental (*Quantity of Support, Confidant Support, Affective Support, and Instrumental Support*). El cuestionario fue diseñado, a partir de una revisión de la literatura para la validez de contenido y de fiabilidad y evaluado en 401 pacientes que acuden a una clínica de medicina familiar por Broadhead et al. (1988). El análisis factorial y del ítem, redujeron a 11 la cantidad de ítems otorgando un cuestionario breve y fácil de completar con dos escalas después de validar y demostrar que cinco ítems pertenecen al Apoyo Confidencial y tres al Apoyo Afectivo, el resto sugieren tres temas restantes distintos que necesitan un mayor estudio (Broadhead et al., 1988).

Cada ítem tiene cinco opciones de respuesta tipo Likert<sup>2</sup>, que va desde 1 (menos de lo que quisiera) hasta 5 (tanto como me gustaría). Puntuaciones mayores reflejan mayor percepción de Apoyo Social (Broadhead et al., 1988).

---

<sup>2</sup> La escala Likert (1932) fue propuesta como una técnica para medir actitudes, creada como la respuesta de cinco puntos opuestos de la opinión respecto a un ítem, esta puede variar según el elemento que se esté intentando medir, va de menos a más (o viceversa) solicitando se indique que tan de acuerdo o en desacuerdo, aprobando o desaprobando, se encuentra con la pregunta que se está realizando.

Los ítems relacionados con el Apoyo Afectivo son las 2, 3, 5, 9 y 11. Las preguntas que miden el Apoyo Confidencial son el 1, 4, 6, 7, 8 y 10.

Al realizar la suma de puntos de los ítems la puntuación máxima que puede obtenerse es 55 y la mínima de 11, y se considera que el sujeto estudiado tiene un Apoyo Social bajo cuando el resultado está por debajo de los 33 puntos (Martín, 2005). En cuanto al Apoyo Afectivo, cifras inferiores a 15 (siendo 25 el puntaje máximo y 5 el mínimo) indican bajos recursos afectivos; para el Apoyo Confidencial el punto de corte es de 18 puntos (ver tabla 4).

**Tabla 4.** Puntos de corte para el DUKE-UNC-11

	Máximo	Mínimo	Punto de corte
Apoyo total percibido	55	11	33
Apoyo confidencial	25	5	15
Apoyo afectivo	30	6	18

**Fuente:** Elaboración propia con base en Bellón et al. (1996a).

De acuerdo con Martín (2005) presentar un valor anormal en la escala DUCKE-UNC-11 se asocia de forma significativa con la presencia de trastornos afectivos, trastorno depresivo, presencia de hiperfrecuentación (una mayor frecuentación de los servicios sanitarios). Así también, presentar un valor anormal en la subescala de apoyo confidencial condiciona claramente al abuso de asistencia a los servicios sanitarios de atención primaria.

El cuestionario tiene la ventaja, por su brevedad y sencillez, de que se puede aplicar fácilmente en atención primaria, donde el tiempo, en ocasiones es muy limitado. La desventaja del cuestionario es que de los estudios de validación existentes (Bellón et al., 1996a; Broadhead et al., 1988; De la Revilla et al., 1994; Piña & Rivera, 2007) ninguno coincide en la asignación de los ítems a cada factor (Apoyo Social o Afectivo), aunque todos precisan que es una herramienta útil para conocer el Apoyo Social de los pacientes.

Además, Bellón et al. (1996a) encontraron una fiabilidad del cuestionario, medida con el test-retest con y sin entrevistador, aportó coeficientes de correlación para cada ítem aceptables (mayor que 0.5) y

bastante buenos para el conjunto de la escala (superior al 0.75). Piña & Rivera (2007) validan el cuestionario en personas seropositivas al VIH obteniendo una  $\alpha$  de Cronbach del instrumento en general de 0.92 (cada ítem fue analizando obteniendo como mínimo 0.90 y máximo 0.92) por lo tanto es posible concluir, de acuerdo a estos autores, que el instrumento es confiable para la población mexicana.

#### **IV.1.2 Esquema de Goldthorpe**

El Esquema de Goldthorpe clasifica la ocupación del sustentador y brinda comprensión de la movilidad social de los individuos (Salazar, 2012). La teoría de las clases está basada en las relaciones laborales para evaluar qué implicaciones tienen las posiciones de clase de los individuos para su vida económica (Goldthorpe & McKnight, 2004).

“Los individuos en diferentes posiciones de clase difieren sistemáticamente no solo en función de sus niveles de renta, sino también al menos en otros tres aspectos más (Goldthorpe y McKnight, 2006; Chan y Goldthorpe, 2007): primero, en su grado de seguridad de la renta; segundo, en su estabilidad de la renta a corto plazo y, tercero, en sus perspectivas de renta a largo plazo” (Goldthorpe, 2012, p. 47).

El modelo tiene en cuenta la situación de trabajo y la situación de mercado, combinadas con la situación de empleo (González, 1992). Junto a ello, este autor incorpora en la clasificación la situación en el trabajo, es decir, la localización de la ocupación en los sistemas de autoridad y control que rigen en los procesos de producción en la que se encuentra (Feito, 1995). “El modelo de Goldthorpe está constituido por categorías ocupacionales cuyas unidades últimas proceden de la escala de deseabilidad social de las ocupaciones construida a principios de los años setenta, a partir de una amplia muestra de varones ingleses” (González, 1992, p. 105).

Erikson & Goldthorpe (1992) dividían la clase de servicio en alta y subalterna; la clase intermedia en la de cuello blanco, pequeña burguesía y trabajador alta; y la clase trabajadora en calificada y no calificada, desarrollando un esquema que se desprendía de este para clasificar las clases sociales según su ocupación laboral (o profesión) del principal sustentador del hogar. El Esquema de Goldthorpe está compuesto por once clases mismas que pueden estudiarse a detalle en la tabla 5.

**Tabla 5.** Clasificación de las clases sociales según el Esquema de Goldthorpe

Clase elitaria	Clasificación	Significado
Clases de servicio	I	Profesionales superiores; directivos de grandes establecimientos y grandes empleadores (más de 25 empleados).
	II	Profesionales de nivel medio e inferior; técnicos superiores; directivos de pequeños establecimientos (menos de 25 empleados); supervisores de empleados no manuales.
Clases intermedias	III a	Empleados no manuales de rutina en la administración y el comercio.
	III b	Trabajadores de servicios personales y de seguridad.
	IV a	Pequeños propietarios, artesanos, etc., con empleados (menos de 25).
	IV b	Pequeños propietarios, artesanos, sin empleados.
	IV c	Agricultores, pescadores, etc.
Clases obreras	V	Supervisores de trabajadores manuales, técnicos de nivel inferior etc.
	VI	Trabajadores manuales cualificados.
	VII a	Trabajadores semicualificados y sin cualificar no agrarios.
	VII b	Trabajadores agrarios.

Fuente: Elaboración propia con base en (Goldthorpe, 1997).

El esquema desarrollado para su uso en la investigación empírica ha llegado a ser ampliamente adoptado, especialmente en la investigación de

movilidad social es conocido como esquema Goldthorpe, Erikson-Goldthorpe-Portocarero o CASMIN (Goldthorpe & McKnight, 2004). “La clasificación de Goldthorpe se ha aplicado en varios países tanto para los estudios de estratificación social y movilidad social, como para los estudios de diferencias socioeconómicas en salud (Kunst & Mackenbach, 1994a; Kunst & Mackenbach, 1994b)” (Regidor, 2001, p.19).

### **IV.1.3 Índice de la Calidad de la Dieta Mediterránea (KIDMED)**

“El término dieta mediterránea se ha utilizado extensamente para describir los hábitos dietéticos tradicionales de la gente de Creta, Italia del sur y otros países mediterráneos durante la década de los años 60” (Kontogianni et al., 2008). Se trata de probablemente uno de los modelos dietéticos más saludables conocidos hoy en día, habiendo sido corroborada por numerosos estudios nutricionales epidemiológicos y experimentales que muestran que los países mediterráneos se benefician de tasas más bajas de morbilidad por enfermedades crónicas y mayor esperanza de vida (Mariscal-Arcas et al., 2009; Prado, Fernández & Anunciabay, 2007; Serra-Majem et al., 2004; Serra-Majem, Ribas, García, Pérez-Rodrigo & Aranceta, 2003).

Los patrones dietéticos mediterráneos tradicionales se caracterizan por una elevada ingesta de alimentos vegetales como frutas y verduras, pan y cereales, legumbres y nueces (Serra-Majem et al., 2004). Aunque el KIDMED fue diseñado para evaluar la adherencia a la dieta mediterránea en general, analiza el consumo de alimentos saludables (Cházaro-Molina & Muñoz-Cano, 2015) sus variantes lo hacen mucho más amplio que sólo mediterránea, los estudios descriptivos hasta ahora han corroborado la importancia de la dieta mediterránea y su relación con buenos estilos de vida (Rodríguez, Espinoza, Gálvez, Macmillan & Solis, 2013).

El KIDMED es un cuestionario corto de 16 ítems desarrollado por Serra-Majem et al. (2002) basado en principios que sustentan los patrones alimentarios mediterráneos, así como los que lo debilitan; es el primero en

evaluar la adecuación de los patrones dietéticos mediterráneos en niños y jóvenes, confirmando que este colectivo está experimentando importantes cambios, lo que los convierte en un objetivo prioritario para las intervenciones nutricionales (Serra-Majem et al., 2004).

“El índice Kidmed está basado en instrumentos desarrollados previamente como el Índice de Adecuación Mediterránea de las Dietas Italianas desarrollado por Alberti-Fidanza & Fidanza (2004) y la Dieta de Supervivencia Global en Ancianos de Trichopoulou et al., (1995) sin embargo como tal constituye el primer índice que evalúa la adecuación de los patrones dietéticos mediterráneos entre la población de 2 a 24 años” (Serra-Majem et al., 2004, p. 934).

En estos últimos años se ha venido empleado el KIDMED como herramienta para evaluar las características de las ingestas alimentarias en relación a un patrón ideal sobre el prototipo de dieta mediterránea, siendo utilizado en diversos estudios en población infantil (Prado et al., 2007). Ha sido considerado como una herramienta rápida y simple para valorar la calidad de la dieta en la población infantil y adolescente (Marrodán et al., 2014). Oseguera (2003) citado por Salazar (2012) considera que la población de la región occidente de México se ha fusionado y toma como base la tradición alimentaria española-europea-criolla, por lo que muchos de los platillos e ingredientes se han mestizado dando por resultado preparación similar de alimentos, por lo que Salazar (2012) y García (2015) muestran la utilidad del KIDMED en jóvenes preuniversitarios de México.

Algunas preguntas del KIDMED denotan connotación negativa con respecto a la dieta mediterránea (ítem 6, 12, 14 y 16) y se les asignó un valor de -1, y aquellas con un aspecto positivo, +1; las sumas de los valores de la prueba se clasificaron en tres niveles (Serra-Majem et al., 2004) o cuatro (Prado et al., 2007) como puede observarse en la tabla 6.

**Tabla 6.** Puntos de corte para la clasificación de la dieta según el Índice de KIDMED

Autor	Puntuación	Clasificación
Serra-Majem et al. (2004)	$\geq 8$	Dieta óptima
	Entre 4 y 7	Es necesario mejorar y ajustar
	$\leq 3$	Muy baja calidad
Prado et al. (2007)	$\geq 8$	Óptima
	6 y 7	Media/alta
	4 y 5	Media/baja
	$\leq 3$	Muy baja

**Fuente:** Elaboración propia con base en Serra-Majem et al. (2004) y Prado et al. (2007).

#### IV.1.4 Cuestionario Internacional de Actividad Física (IPAQ - Corto)

Dada la importancia de la actividad física para la salud, se hizo necesario desarrollar herramientas válidas dirigidas a medir este comportamiento; para llevar a cabo este objetivo varios instrumentos han sido diseñados durante varias décadas de investigación, sin embargo, hasta mediados de 1990, la comparación de los datos recopilados de actividad física estaba obstaculizada por el uso de múltiples instrumentos (Hallal et al., 2010). La Organización Mundial de la Salud (OMS) en su informe de 2002 citado por Ainsworth et al. (2006), señaló que no existía una definición o medida de la actividad física internacionalmente aceptada y esta dificultad era el resultado de una serie de fuentes y datos directos e indirectos, así como una gama de instrumentos y metodologías para estimar los niveles de actividad física.

“Debido a que la actividad física es considerada un problema de salud mundial, en 1996 un grupo de expertos internacionales convocados por el Instituto Karolinska, la Universidad de Sydney, la OMS y los Centros para el Control y Prevención de Enfermedades, han estado trabajando en la elaboración, implementación y mejora de un cuestionario que examina diferentes dimensiones de la actividad física, para tener

información que pueda utilizarse en los sistemas de monitorización y vigilancia sanitaria de alcance poblacional” (Mantilla & Gómez-Conesa, 2007, pp. 50-51).

Con el propósito de proveer instrumentos comunes que puedan ser utilizados para obtener datos internacionalmente comparables relacionados con la actividad física relacionada con la salud es que se desarrollan los Cuestionarios Internacionales de Actividad Física (*IPAQ*, por sus siglas en inglés) (IPAQ Research Committee, 2002). Los cuestionarios se plantearon para ser utilizados por adultos de 18 a 65 años en sus ocho versiones, éstas varían según el número de preguntas (corto o largo), el periodo de repetición (“usualmente en una semana” o “los últimos 7 días”) y el método de aplicación (encuesta autoaplicada, entrevista cara a cara o por vía telefónica) (Craig et al., 2003; Mantilla & Gómez-Conesa, 2007). La versión corta de 9 ítems permite obtener información sobre el tiempo dedicado a caminar, en actividades de intensidad vigorosa y moderada, y en actividad sedentaria. La versión larga de 31 ítems sirve para indagar de forma detallada dentro de los dominios de actividad doméstica y de jardinería, ocupacional, transporte autónomo, física de ocio, y actividades sedentarias (Craig et al., 2003).

Varios instrumentos están disponibles para medir la actividad física, incluyendo cuestionarios de auto-reporte, calorimetría indirecta, observación directa, telemetría de frecuencia cardiaca y sensores de movimiento como los acelerómetros etcétera, sin embargo, todos estos métodos tienen limitaciones conocidas y para la actividad física no existe una forma de medición perfecta (Lee, Macfarlane, Lam & Stewart 2011). El IPAQ desarrollado como un instrumento para el monitoreo internacional de la actividad física y la inactividad (Craig et al., 2003) es recomendado por la OMS (2009) como parte de los indicadores básicos relacionados con las acciones que recomienda a los estados miembros para llevar a cabo la Estrategia Mundial sobre el régimen alimentario, actividad física y salud, dado que se trata de un instrumento válido y fiable.

Craig et al. (2003) demostraron que el IPAQ-corto fue generalmente mejor recibido en sitios donde se administraron las dos versiones, ya que el IPAQ-largo fue reportado como “demasiado aburrido y repetitivo” y aunque los autores originales recomiendan esta versión para los estudios de vigilancia de actividades físicas más países expresaron una preferencia cualitativa por el formulario corto. Mantilla & Gómez-Conesa (2007) señalan que “se ha sugerido que sea la versión corta la utilizada en estudios poblacionales” (p. 50).

Los cuestionarios IPAQ ha sido utilizado en Australia, Brasil, Canadá, Finlandia, Guatemala, Holanda, Japón, Portugal, Sudáfrica, Suecia, Estados Unidos y Reino Unido (Craig et al., 2003), además Colombia y Brasil (Hallal et al., 2010), Países Bajos, Vietnam, Suiza, Noruega, China, Grecia y Puerto Rico (Lee et al., 2011). El IPAQ demostró tener aceptables propiedades de medición para utilizarse en diferentes lugares en distintos idiomas; son apropiadas para hacer estudios nacionales poblacionales de prevalencia de participación en actividad física (Booth, 2000; Craig et al., 2003; Hallal et al., 2010; IPAQ Research Committee, 2002). El IPAQ originalmente fue recomendado para adultos de 18 a 65 años, sin embargo, varios estudios en Colombia y Brasil han incluido a adultos mayores (Hallal et al., 2010) y en México ha sido utilizado para medir jóvenes entre 15 y 18 años (García, 2015; Salazar, 2012). Ambas versiones (corto y largo) evalúan tres características específicas de actividad: intensidad (leve, moderada o vigorosa), frecuencia (medida en días por semana) y duración (tiempo por día), dado que la actividad física se mide a través del registro en METs<sup>3</sup>-min-semana (Mantilla & Gómez-Conesa, 2007).

---

<sup>3</sup> Un MET describe el gasto de energía de una persona en reposo (Tasa Metabólica de Reposo o Equivalente Metabólico de Tarea, *MET* por sus siglas en inglés) y se expresa generalmente en términos de absorción de oxígeno. Un MET corresponde a 3,5 ml O<sub>2</sub>/kg x min, que es el consumo mínimo de oxígeno que el organismo necesita para mantener sus constantes vitales (Howley, 2000).

El protocolo para calificar las respuestas al IPAQ utilizan la referencia de Ainsworth et al. (2000)<sup>4</sup> e indican que caminar tiene una tasa metabólica de 3.3 METs; la actividad física moderada, 4 METs; y la actividad física vigorosa, 8 METs. Teniendo en cuenta lo anterior debe multiplicarse el valor en METs de la actividad realizada por la frecuencia y la duración de la actividad para obtener el índice de actividad física y después clasificarlo en alguna de las tres categorías (IPAQ Research Committee, 2005) (ver tabla 7).

**Tabla 7.** Clasificaciones a las respuestas del IPAQ

Categoría	Criterio
Bajo	Aquellas personas que no cumplen con los criterios para las categorías 2 o 3 se consideran que tienen un nivel de actividad física "bajo".
Moderado <sup>5</sup>	<ul style="list-style-type: none"> <li>a) 3 o más días de actividad vigorosa de al menos 20 minutos al día.</li> <li>b) 5 o más días de actividad de intensidad moderada y/o caminar al menos 30 minutos por día.</li> <li>c) 5 o más días de cualquier combinación de actividades de caminata, de intensidad moderada o intensa, alcanzando un mínimo de actividad física total de al menos 600 MET-min/semana.</li> </ul>
Elevado <sup>5</sup>	<ul style="list-style-type: none"> <li>a) Actividad de intensidad vigorosa en al menos 3 días logrando al menos 1500 MET-minutos / semana.</li> <li>b) 7 o más días de cualquier combinación entre caminar y actividades de intensidad moderada o de intensidad vigorosa, logrando un mínimo de actividad física total de al menos 3000 MET-min/semana.</li> </ul>

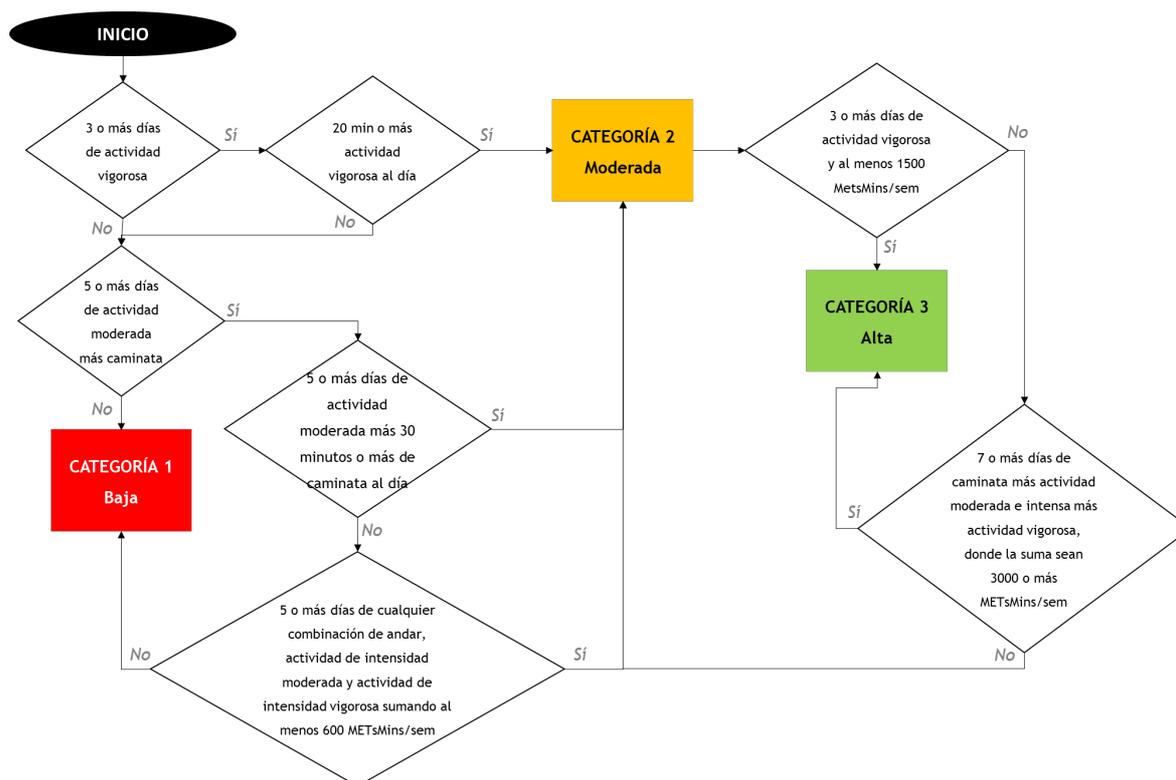
**Fuente:** Elaboración propia con base en IPAQ Research Committee (2005).

En la figura 2 se presenta el algoritmo que permite clasificar las respuestas al IPAQ-Corto siguiendo las indicaciones presentadas anteriormente.

<sup>4</sup> Las actividades en el Compendio de Actividades Físicas van desde 0.9 METs (por dormir) hasta 18 METs (por hacer una carrera a 17.5 Km/h) (Ainsworth et al., 2000).

<sup>5</sup> Si cumple al menos uno de los criterios.

Figura 2. Clasificaciones a las respuestas del IPAQ-Corto



**Fuente:** Elaboración propia con base en “Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms” (IPAQ Research Committee, 2005).

#### IV.1.5 Test de Fagerström para la Dependencia de la Nicotina (FTND)

El Test de Fagerström (*FTQ* por sus siglas en inglés) fue diseñado para evaluar por medio de ocho preguntas la dependencia a la nicotina y determinar si se necesita o no terapia de reemplazo de nicotina para tratar el síndrome de abstinencia (Fagerström, 1978) siendo uno de los cuestionarios que más atención recibió en la literatura (Becoña & Vázquez, 1998; Carrasco, Luna & Vila, 1994). El cuestionario fue creado de acuerdo a los principios de adicción y parece tener validez de cara a la dependencia al tabaco (Fagerström, 1978, p. 236), sin embargo Heatherton, Kozlowski, Frecker & Fagerström (1991) examinaron y redefinieron el FTQ reduciéndolo a seis ítems ya que encontraron dos preguntas (del cuestionario original) que no estaban relacionadas con

ninguna de las medidas bioquímicas y eran los que contribuían a las deficiencias psicométricas del instrumento, esta nueva versión fue denominada Test de Fagerström para la Dependencia a la Nicotina (*FTND* por sus siglas en inglés).

La FTND tiene niveles aceptables de consistencia interna y está estrechamente relacionada con los índices bioquímicos de fumar (Heatherton et al., 1991), ha sido traducido a distintos idiomas y se ha utilizado en distintos países como Austria, Dinamarca, Finlandia, Francia, Polonia, Estados Unidos, Suecia y Reino Unido (Fagerström et al., 1996) además de China, Holanda, Canadá, Alemania, España, Brasil, Suiza y Japón (Meneses-Gaya, Zuardi, Loureiro & Crippa, 2009). Al español hicieron la traducción del test Becoña & Vázquez (1998) quienes señalan que su utilidad para estudiar la dependencia a la nicotina en fumadores españoles, así como para identificar a los individuos que tienen mayor dependencia y por lo tanto mayor riesgo.

En México, la FTND ha sido utilizada para identificar indicadores de abstinencia (Lara-Rivas, Ramírez-Venegas, Sansores-Martínez, Espinosa & Regalado-Pineda, 2007) y para estudiar los hábitos tabáquicos de jóvenes (Delgado-Quiñone & Martín-Guzmán, 2015) y adolescentes preuniversitarios (Linage-Rivadeneira & Lucio-Gómez-Maqueo, 2012; Luque, Ruíz, Gómez & Rodríguez, 2008; Salazar, 2012). La TFND demostró tener una estructura factorial consistente, válida y confiable para la población mexicana (Linage-Rivadeneira & Lucio-Gómez-Maqueo, 2012). De Leon et al. (2003) proponen el Índice del Fumador Crónico (HSI por sus siglas en inglés) tomando los ítems 1 y 4 del FTND, sin embargo, señalan que su utilidad aún no ha sido completamente establecida y además hacer las seis preguntas puede predecir el éxito de dejar de fumar.

Dadas las diferencias entre la FTQ y la FTND la clasificación entre las respuestas otorgadas a cada instrumento también varía. La tabla 8 muestra la clasificación de dependencia según las respuestas otorgadas a cada instrumento.

**Tabla 8.** Clasificación a las respuestas del Test de Tolerancia de Fagerström

Autor	Cuestionario	Puntaje	Clasificación
Fagerström (1978)	FTQ	0 a 1 puntos	Mínima dependencia física
		11 puntos	Máxima dependencia física
Fagerström et al. (1996)	FTND	0 a 3 puntos	Dependencia leve
		Entre 4 y 6 puntos	Dependencia moderada
		7 puntos o más	Dependencia intensa

Fuente: Elaboración propia con base en Fagerström (1978) y Fagerström et al. (1996).

#### IV.1.6 Test de alcoholismo (CAGE)

Con la finalidad de detectar los principios de alcoholismo en adolescentes se utilizó el test de CAGE-4, mismo que está inserto entre 12 ítems, para evitar el rechazo por parte de los jóvenes. De acuerdo con Herrán & Ardila (2005), el cuestionario de CAGE fue originalmente desarrollado por John Ewing en 1968 para detectar bebedores rutinarios. Ewing (1984) señala que, con cuatro preguntas de entrevista clínica, las preguntas de CAGE, han demostrado ser útiles para ayudar a hacer un diagnóstico de alcoholismo. Munera, Zuluaga & Vélez (1994) indica que se trata de un cuestionario simple y corto pero muy específico para diagnosticar alto consumo de alcohol, indicativo de alcoholismo en sujetos de la población general, las cuatro preguntas clave tienen que ver con la crítica social, la culpabilidad individual, la ingesta matutina y la necesidad sentida del sujeto de consumir alcohol.

Mayfield, McLeod & Hall (1974) publican el primer estudio de validación y posteriormente ha sido utilizado en distintos campos y poblaciones y ha sido traducido a diferentes idiomas. Beresford, Low, Adduci & Goggans (1982), lo utilizaron en un estudio en un servicio de cirugía ortopédica; Masur & Monteiro (1983) en un hospital psiquiátrico brasileño; Markowitz (1984) en su estudio el abuso de alcohol como una respuesta a la impotencia percibida en la organización; King (1986) en Londres con asistentes de medicina general; Seixas, Washburn & Eisen (1988) en un sistema penitenciario; Iacoponi,

Laranjeira & Jorge (1989) en Brasil; Dabkowski, Ziótkowski, Rogiewicz & Rybakowski (1989) en Polonia; Rueff, Crnac & Darne (1989) en Francia; Niquille, Burnand & Yersin (1990) en Alemania; y otros estudios más.

Desde su introducción en 1970 el test de CAGE fue reconocido como el más eficiente y efectivo dispositivo de test de screening, no sólo es fácil de administrar y confiable al distinguir personas alcohólicas, sino que, además es una herramienta menos intimidante respecto a otros instrumentos (Herrán & Ardila, 2005). El test de CAGE ha demostrado ser eficaz al momento de establecer un diagnóstico de presunción de alcoholismo, además, el test posee una elevada sensibilidad y especificidad, así como un elevado valor predictivo positivo y presenta una serie de ventajas que lo hacen muy adecuado para ser aplicado, como su brevedad, fácil manejo y bajo coste (González et al., 1997). Ewing (1984) señala que CAGE, viene del acrónimo de las cuatro preguntas importantes en el test:

- *Have you ever felt you should Cut down on your drinking?*<sup>6</sup>
- *Have people Annoyed you by criticizing your drinking?*<sup>7</sup>
- *Have you ever felt bad or Guilty about your drinking?*<sup>8</sup>
- *Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (Eye opener)?*<sup>9</sup>

Los resultados del CAGE dependen del número de respuestas positivas a cualquiera de las cuatro preguntas que forman el acrónimo, estos se pueden clasificar de la siguiente según la tabla 9.

---

<sup>6</sup> ¿Ha intentado en alguna ocasión dejar de beber o beber menos?

<sup>7</sup> ¿Le molesta que le pregunten si bebe?

<sup>8</sup> ¿Se ha sentido alguna vez culpable por beber?

<sup>9</sup> ¿Ha tomado alguna vez un trago para “matar el gusanillo” a primera hora de la mañana?

Tabla 9. Dictamen a las respuestas del test de CAGE

Núm. de respuestas positivas	Clasificación
0 a 1	Consumo social o bebedor social
2	Consumo de riesgo
3	Consumo perjudicial
4	Dependencia alcohólica

Fuente: Elaboración propia con base en (Muracén, Martínez, Aguilar & González, 2001).

Herrán & Ardila (2005) indican que dos o tres respuestas afirmativas permite tener un alto grado de sospecha de alcoholismo; cuatro respuestas positivas son indicio de alcoholismo patológico. Dado que no se pretendía dividir a los estudiantes entre los cuatro tipos de alcoholismo, se utilizó el punto de corte en la versión original de CAGE que también utilizaron Herrán & Ardila (2005) como muestra la tabla 10.

Tabla 10. Valores del Test de CAGE

Núm. de respuestas positivas	Clasificación
0	No bebedor
≤ 3	Alto riesgo
>3	Indicio de alcoholismo patológico

Fuente: Elaboración propia con base en Ewing (1984) y Herrán & Ardila (2005).

#### IV.1.7 APGAR - Familiar

Diseñado por Smilkstein (1978) como un cuestionario breve y sencillo para explorar la función familiar poniendo a prueba cinco áreas de la función familiar, planteado para obtener el punto de vista del estado funcional de la familia de un paciente APGAR viene del acrónimo de los componentes funcionales de adaptabilidad (*Adaptability*), cooperación (*Partnership*), crecimiento (*Growth*), afectividad (*Affection*) y resolución (*Resolve*). Estos fueron elegidos empíricamente por el autor, ya que representan los temas comunes en la literatura de las ciencias sociales que se ocupa de las familias y son representados en la tabla 11 (Smilkstein, 1978).

Tabla 11. Definiciones de los componentes del APGAR Familiar

Componente	Definición	Lo que se mide
Adaptación	El uso de los recursos intra y extrafamiliares para resolver problemas cuando el equilibrio de la familia es amenazado durante una crisis.	Cómo los recursos son compartidos, o el grado de satisfacción de los miembros con el apoyo recibido cuando se necesitan recursos familiares.
Cooperación	La participación en la toma de decisiones y responsabilidades por los miembros de la familia.	Cómo las decisiones son compartidas, o la satisfacción de los miembros de la familia con la reciprocidad en la comunicación familiar y la resolución de problemas.
Crecimiento	Maduración y realización física y emocional, así como la autorrealización de los miembros de la familia que se logra a través del apoyo y la orientación mutua.	Cómo el apoyo es compartido, o la satisfacción de los miembros de la familia con la libertad disponible dentro de la familia de cambiar roles y alcanzar el crecimiento o madurez física y emocional.
Afecto	Es el cuidado o relación amorosa que existe entre los miembros de una familia.	Cómo las experiencias emocionales son compartidas, o la satisfacción de los miembros con la intimidad y la interacción emocional que existen dentro de la familia.
Resolución	El compromiso de dedicar tiempo a otros miembros de la familia para nutrirlos física o emocionalmente. Es también, usualmente, involucrar una decisión de compartir la riqueza y el espacio.	Cómo el tiempo (espacio y dinero) son compartidos, o la satisfacción de los miembros con el compromiso del tiempo que se ha hecho a los miembros de la familia.

Fuente: *The Family APGAR: A Proposal for a Family Function Test and Its Use by Physicians.* (Smilkstein, 1978).

Estos componentes se han traducido en una pauta de cinco preguntas (APGAR-familiar) que mide el funcionamiento de la familia a través de la satisfacción del entrevistado con su vida familiar, conociendo así la percepción que un integrante de la familia tiene de su funcionamiento familiar (Horowitz, Florenzano & Ringeling, 1985). Para Bellón, Delgado, Luna & Lardelli (1996b) el APGAR-familiar es una herramienta útil para conocer la situación familiar y su posible papel en el origen de los conflictos y en su solución, es decir, permite conocer si la familia puede considerarse un recurso para los individuos del grupo, o si por el contrario influirá empeorando su situación.

El test APGAR-familiar consta de cinco preguntas puntuadas con una escala de Likert que va de 0 a 2, en la que 0 equivale a “casi nunca”, 1 a “a veces” y 2 a “casi siempre” (Smilkstein, 1978). Austin y Huberty (1986) citado por Rätty, Larsson, Söderfeldt & Larsson (2005), simplifican el lenguaje y utilizan una escala de respuesta de cinco puntos que va desde “nunca” hasta “siempre”. Al final deben sumarse los puntos obtenidos para conocer la función familiar, los cuales pueden clasificarse según la tabla 12.

Tabla 12. Valores del Test APGAR-Familiar

Autor	Escala	Puntaje	Clasificación
Smilkstein (1978)	0 a 2	0 a 3 puntos	Disfunción grave
		4 a 6 puntos	Disfunción leve
		7 a 10 puntos	Normofuncional
Austin y Huberty (1986) citado por Suarez & Alcalá (2014)	0 a 4	Menor o igual a 9	Disfunción severa
		10 a 12 puntos	Disfunción moderada
		13 a 16 puntos	Disfunción leve
		17 a 20 puntos	Normal

Fuente: Elaboración propia con base en Smilkstein (1978) y Suarez & Alcalá (2014).

El APGAR-familiar puede aplicarse a diferentes miembros de la familia y en distintos momentos para palpar su variación (Horowitz et al., 1985). Bellón et al. (1996b) analizaron la estructura factorial, al poner de manifiesto que se trata de una escala unidimensional y al expresar que los 5 ítems miden aspectos del mismo concepto de la función familiar. Smilkstein, Ashworth & Montano (1982) señalan que el formato de tres opciones de respuesta (0 a 2) es para uso

clínico; y el segundo formato, con cinco opciones de respuesta (0 a 4) lo recomiendan para su uso en población general.

El APGAR original siguió un proceso de fiabilidad y validez psicométrica lo que permitió la extensión de su uso (Mendoza, 2007). Diversos estudios han demostrado la consistencia del instrumento con un alfa de Cronbach superior al 0.70, como los de Bellón et al. (1996b); Forero, Avendaño, Duarte & Campo (2006); Maddaleno, Horwitz, Jara, Florenzano & Zalazar (1987); Mendoza (2007); y Smilkstein et al. (1982), lo cual indica una consistencia dentro del rango esperado para este tipo de instrumentos. Para Forero et al. (2006) el APGAR familiar, por ser de libre uso y por el poco tiempo que se necesita para completarla, resulta una medida atractiva para la evaluación de funcionamiento familiar. Bahasas (1999) citado en Mendoza (2007), señala que fue diseñado para el uso potencial de personas en distintos estratos socioeconómicos y en diversos contextos socioculturales.

#### **IV.1.8 Escala de Satisfacción con la Vida para Niños (SWLS - C)**

Diener, Emmons, Larsen & Griffin (1985) de la Universidad de Illinois, desarrollaron la Escala de Satisfacción con la Vida (SWLS, por sus siglas en inglés) “diseñada alrededor de la idea de que uno debe pedir a los sujetos un juicio global de su vida para medir el concepto de satisfacción con la vida” (Diener et al., 1985, pp. 71-72). “La SWLS se recomienda como un complemento de las escalas que se centran en la psicopatología o el bienestar emocional, ya que evalúa el juicio evaluativo consciente de una persona de su vida mediante el uso de su propio criterio” (Pavot & Diener, 1993, p. 101).

La SWLS consta de cinco ítems con respuestas de tipo Likert de siete puntos (Diener et al., 1985). La versión en español fue validada por Atienza, Pons, Balaguer & García-Merita (2000), donde se reduce a cinco opciones de respuesta que van desde “totalmente en desacuerdo” hasta “totalmente de acuerdo”. Según Alfonso & Allison (1992) citados por Pavot & Diener (1993) la

SWLS está a nivel de lectura de sexto y décimo grado, por lo tanto, también es utilizable con la mayoría de los adultos. Gadermann (2009) y Gadermann, Schonert-Reichl & Zumbo (2010) desarrollaron la Escala de Satisfacción con la Vida adaptada para niños (SWLS-C), cuya diferencia con la versión original es que las opciones de respuesta, al igual que la versión en español, son reducidas a una escala de tipo Likert de cinco puntos, por lo que la puntuación total va de 5 (insatisfecho) a 25 (satisfecho). Tanto la versión en castellano de Atienza et al. (2000) como la adaptada para niños de Gadermann et al. (2010), presentan una estructura unifactorial como la versión original de Diener et al. (1985) y de igual forma se ha probado su alta consistencia interna y estabilidad temporal.

De acuerdo con Galiana, Gutiérrez, Sancho, Oliver & Tomás (2015) la SWLS es la más utilizada a lo largo de la literatura, Pavot & Diener (2008) señalan que se pueden encontrar versiones del cuestionario en francés, holandés, coreano, hebreo, chino mandarín, ruso, checo, español, portugués, árabe, bosnio, danés, flamenco, georgiano, alemán, friego, hindi, húngaro, islandés, italiano, japonés, jemer, noruego, persa, polaco, rumano, serbio, setswana, tailandés, turco, urdú y vietnamita. “Los datos normativos para el SWLS están disponibles para diversas poblaciones, incluyendo adultos mayores, prisioneros, individuos bajo tratamiento para el abuso de alcohol, mujeres abusadas, pacientes de psicoterapia, cuidadores de ancianos, personas con discapacidad, así como estudiantes universitarios” (Pavot & Diener, 1993, p. 103). En México las propiedades psicométricas de la SWLS fueron analizadas por Padrón, Gutiérrez & Medina (2015), sin embargo, la SWLS-C ha sido poco utilizada, algunos referentes son Salazar (2012) y García (2015) quienes abordan los resultados obtenidos de forma general.

La forma de puntuar y medir la satisfacción con la vida, también depende de la escala que se esté utilizando (ver tabla 13), en esta investigación se considera la propuesta para el SWLS-C.

Tabla 13. Puntos de corte para la SWLS-C

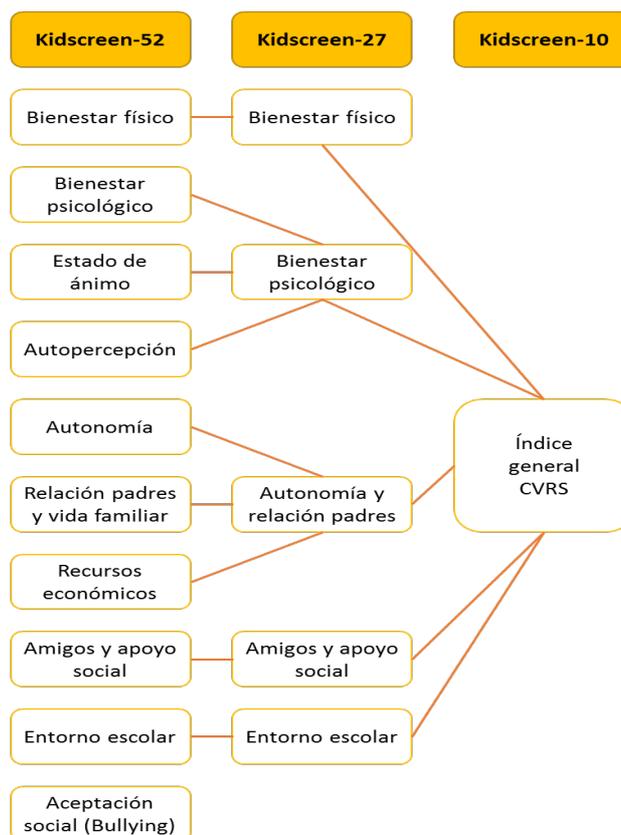
Autor	Forma de puntuar	Puntuación	Clasificación
Diener et al. (1985)	SWLS	30 a 35	Muy satisfecho
		25 a 29	Satisfecho
		20 a 24	Satisfacción media
		15 a 19	Un poco por debajo de la media de satisfacción
		10 a 14	Insatisfecho
Pavot & Diener (1993)	SWLS-C	5 a 9	Extremadamente insatisfecho
		21 a 25	Satisfecho
Atienza et al. (2000)	SWLS en español	16 a 20	Ligeramente satisfecho
		11 a 15	Satisfacción media
		6 a 10	Ligeramente insatisfecho
		0 a 5	Insatisfecho

Fuente: Elaboración propia con base en Atienza et al. (2000); Diener et al. (1985); Pavot & Diener (1993).

#### IV.1.9 Calidad de Vida Relacionada con la Salud en Niños y Adolescentes (KIDSCREEN - 10)

“Los cuestionarios KIDSCREEN son una familia de instrumentos desarrollados y normalizados para la topografía de salud relacionados con la calidad de vida (CVRS), en niños y adolescentes de entre 8 y 18 años” (KIDSCREEN Group Europe, 2006, párr. 1). De acuerdo con Ravens-Sieberer et al. (2005), fueron desarrollados tres instrumentos dirigidos a niños, adolescentes y padres; el KIDSCREEN-52, una versión larga que abarca diez dimensiones de CVRS; KIDSCREEN-27, una versión corta que abarca cinco dimensiones; y el KIDSCREEN-10 la versión más corta que mide la CVRS global, como se muestra en la figura 3.

Figura 3. Dimensiones de los cuestionarios KIDSCREEN



Fuente: KIDSCREEN Group Europe (2006).

“Los instrumentos KIDSCREEN, están conceptualmente basados en la definición de la CVRS como un constructo multidimensional que cubre los componentes físicos, emocionales, mentales, sociales y de conducta, del bienestar y de las funciones percibidas por pacientes y/o otros individuos” (KIDSCREEN Group Europe, 2006, párr. 1). Por su transculturalidad el KIDSCREEN permite establecer comparaciones entre países y disponer de una medida de CVRS especialmente útil para estudios multicéntricos internacionales (Aymerich et al., 2005). Los instrumentos pueden ser utilizados para la detección, el seguimiento y la calificación en las encuestas nacionales e internacionales de salud europea, ya que el proyecto KIDSCREEN fue financiado por la Comisión Europea, cuyo objetivo es desarrollar un instrumento estandarizado de evaluación de la calidad de vida de los niños (KIDSCREEN Group Europe, 2006).

De acuerdo con Aymerich et al. (2005), el KIDSCREEN es el primer instrumento de CVRS para población infantil y adolescente desarrollado

simultáneamente en diversos países europeos, a partir de las opiniones de la propia población diana. De esta manera, para el KIDSCREEN Group Europe (2006), los cuestionarios miden la calidad de vida desde el punto de vista del niño con respecto a la integridad física, mental y social. Además, permiten identificar poblaciones de riesgo en términos de la salud subjetiva y sugerir intervenciones tempranas apropiadas al incluir el instrumento en la investigación de servicios sanitarios e informes de salud (Salazar, 2012).

El KIDSCREEN-10 se ofrece como un instrumento fácil de administrar, de puntuar e interpretar con validez y confiabilidad demostrada (Ravens-Sieberer et al., 2010). En México se inició el proceso de adaptación en febrero de 2011 y se aprobó la versión final mexicana del instrumento por el grupo KIDSCREEN en mayo de 2011 (Hidalgo-Rasmussen, Rajmil & Espinoza, 2014).

El KIDSCREEN-10 está relacionado con la actividad física (ítems 1 y 2); estados de ánimo, emociones y sentimientos (ítems 3 y 4): ocio y tiempo libre (ítems 5 y 6); calidad de la interacción y sentimientos entre el escolar y los padres - responsables legales (ítem 7); relación del escolar con sus iguales (ítem 8) y; por último, capacidad cognitiva y la satisfacción con el rendimiento escolar (ítems 9 y 10) (Gálvez et al., 2015).

La instrucción básica del cuestionario es responder a la pregunta mediante una escala Likert de cinco puntos, que va desde nunca hasta siempre, dependiendo de cómo se han sentido los últimos siete días. Para hacer la comparación con países europeos, de acuerdo con Salazar (2012) se utiliza una escala de 0 a 100 puntos (nunca = 20; un poco = 40; moderadamente = 60; mucho = 80; y, muchísimo = 100), por lo que, una calificación alta significa mejor CVRS.

#### IV.1.10 General Health Questionnaire (GHQ - 12)

El Cuestionario General de Salud (*GHQ*, por sus siglas en inglés), diseñado por Goldberg en 1969, es una prueba autoadministrada que sirve para la identificación de las enfermedades psiquiátricas en pacientes de medicina general (Goldberg & Blackwell, 1970). “El GHQ-12 ha sido usado para valorar la salud mental en población general, así como en ciertos grupos específicos de población tales como adolescentes (López-Castedo & Fernández, 2005), adultos (Sánchez-López & Dresch, 2008), mujeres posparto (Navarro, Ascaso, Grecia-Esteve, Aguado, Torres & Martín-Santos, 2007), entre otros” (Villa, Zuluaga & Restrepo, 2013, p. 535). De acuerdo con Salazar (2012), el GHQ-12 sirve para evaluar síntomas emocionales, bienestar psicológico o trastornos mentales comunes en la población general, y en atención de primer nivel de complejidad, adolescentes y adultos.

El cuestionario tiene como objetivo, acorde con Goldberg & Blackwell (1970), ofrecer información sobre el estado actual mental, más que sobre los rasgos de la personalidad o la posibilidad de caer enfermo en el futuro. Conforme a Villa et al. (2013) el GHQ-12 ha sido usado en diversas investigaciones y ha permitido identificar y caracterizar a la población estudiada en cuanto a la salud mental.

La primera versión del GHQ contaba con 60 ítems, en la actualidad existen versiones de 36, 30, 28, 20 y 12 preguntas respectivamente. De acuerdo con Goldberg et al. (1997), si los investigadores desean utilizar un instrumento de detección como un indicador de casos, el GHQ-12 es muy robusto y funciona tan bien como el instrumento de más tiempo. Campo-Arias (2007) señala que “la versión de doce puntos es la preferida por la brevedad y más rápida administración y es de particular utilidad en poblaciones con limitadas habilidades para la lectura y la escritura, con dificultades para la concentración o que pueden ser propensas al cansancio” (p. 30). El GHQ-28 es recomendable si hay un interés en los resultados en escala previstas, además de la puntuación total (Goldberg et al., 1997).

Shevlin & Adamson (2005) indican que hay tres factores que mide el GHQ-12, ansiedad y depresión; disfunción social y pérdida de la confianza y; autoestima. Villa et al. (2013) encuentran dos dimensiones para el GHQ-12, “para el factor 1 se agruparon los ítems 1, 3, 4, 7, 8 y 12 y se nombró como ‘bienestar psicológico’; en el factor 2 se agruparon los ítems 2, 5, 6, 9, 10 y se nombró como ‘malestar psicológico’” (p. 540). Las escalas del GHQ están compuestas por síntomas de ansiedad y depresión como los clasifican en la actualidad la Organización Mundial de la Salud y la Asociación Psiquiátrica Americana (Campo-Arias, 2007).

En cualquiera de sus versiones, el GHQ, tiene cuatro opciones de respuesta y diferentes métodos de puntuación. De acuerdo con Goldberg et al. (1997), el método usual (0-0-1-1); el tipo Likert (0-1-2-3) y; el método C-GHQ en el que algunos ítems son medidos (0-1-1-1) si son ítems que indican enfermedad y aquellos que reflejan salud son puntuados como (0-0-1-1). El puntaje total es la sumatoria de los valores obtenidos en cada uno de los ítems, así se obtiene un estimado del grado de severidad de la ausencia de salud mental (García, 1999). De acuerdo con Campo-Arias (2007), para todas las formas de puntuación, más calificación indica mayor número o severidad de síntomas.

El GHQ se establece un punto de corte por encima del cual se consideran “casos”, o personas con disturbios psiquiátricos. De acuerdo con Villa et al. (2013) “en cuanto a la normalización del instrumento en regiones distintas a la inglesa, donde fue diseñado, varios investigadores han identificado rangos de valor distintos como punto de corte para la identificación de casos probables de perturbación mental” (p. 534).

Para Salazar (2012), en Latinoamérica el punto de corte concerniente a trastornos mentales comunes puede ser de 11 o 12 para la calificación tipo Likert (0-1-2-3); 7 u 8 para (0-1-1-1); 4 o 5 para (0-0-1-1) y 1 o 2 para (0-0-0-1). “Gran parte de los estudios desarrollados en nuestro contexto utilizan el GHQ-12 como un instrumento cribado, utilizando la puntuación de GHQ, no la puntuación Likert” (García, 2015, p. 116) motivo por el cual, en la presente

investigación se utilizó el método usual, es decir, a las cuatro categorías de respuesta se les adjudicó 0 (a las repuestas 1 y 2) y 1 (a las respuestas 3 y 4), se sumaron los valores obtenidos de las doce preguntas y se consideró como casos de riesgo de padecer un trastorno mental a las personas con puntuación  $\geq 5$ . La tabla 14 muestra los puntos de corte más utilizados en las adaptaciones del GHQ-12 al habla hispana.

**Tabla 14.** Puntos de corte para el GHQ-12

Autor	Forma de puntuar		Clasificación
Goldberg et al. (1997)	Método usual (0-0-1-1)		$\leq 2$ ausencia de psicopatología $> 2$ presencia de psicopatología
	C-GHQ (0-1-1-1)		$\leq 5$ ausencia de psicopatología $> 5$ presencia de psicopatología
	Likert (0-1-2-3)		$\leq 12$ ausencia de psicopatología $> 12$ presencia de psicopatología
	Fullerton, Acuña, Florenzano, Cruz & Weil (2003); Lewis & Araya (1995)	Chile	C-GHQ
5 a 6 psicopatología subumbral 7 a 12 indicativos de presencia de psicopatología.			
García (1999)	Cuba	Likert	Deterioro de la salud mental: > 10 bajos 10 a 16 normal > 16 altos
Mari & Williams (1985)	Brasil España	Método usual	0 a 4 ausencia de psicopatología
López-Castedo & Fernández (2005); Rocha, Pérez, Rodríguez-Sanz, Borrell & Obiols (2011)			$\geq 5$ indicadoras de psicopatología

**Fuente:** Elaboración propia con base en Fullerton et al. (2003); García (1999); Goldberg et al. (1997); Lewis & Araya (1995); López-Castedo & Fernández (2005); Mari & Williams (1985); Rocha et al. (2011).

#### IV.1.11 Cuestionario sobre la Forma Corporal (BSQ - 8C)

El trastorno de la imagen corporal se define como una preocupación exagerada, que produce malestar, hacia algún defecto imaginario o extremado de la apariencia física (Raich, Deus, Muñoz, Pérez & Requena, 1991). Las preocupaciones sobre la forma del cuerpo son comunes entre las mujeres jóvenes en las culturas occidentales y, en una forma extrema, constituyen una característica central de los trastornos de la alimentación como la anorexia nerviosa y bulimia nerviosa (Cooper et al., 1987). Para Rosen (1992) citado por Raich et al. (1996) la imagen corporal es un concepto que se refiere a la manera en que uno percibe, imagina, siente y actúa respecto a su propio cuerpo.

Cooper et al. (1987) encontraron una serie de encuestas comunitarias que demuestran que la insatisfacción con la forma del cuerpo es común entre las adolescentes, así como un estudio americano temprano se encontró que la mayoría de las muchachas adolescentes eran profundamente infelices con su tamaño y forma. Los métodos de evaluación sobre la imagen corporal casi inexistentes de acuerdo a Raich et al. (1996) llevan a Cooper et al. (1987) a desarrollar el Cuestionario sobre la Forma Corporal (BSQ por sus siglas en inglés).

Se trata de un instrumento de 34 ítems diseñado para estudiar la imagen corporal en trastornos alimentarios (Cooper et al., 1987) que además mide el miedo a engordar, la autodesvaloración por la apariencia física, el deseo de perder peso y la evitación de situaciones en las que la apariencia física pudiera atraer la atención de los otros (Raich et al., 1996). Fue traducido al español por Raich et al. (1991) en una muestra de mujeres adolescentes catalanas, después fue validado por Raich et al. (1996) en un estudio sobre estudiantes universitarias y pacientes diagnosticados de trastornos alimentarios. En México hizo la adaptación cultural y la validación en población mexicana fue llevada a cabo por Vázquez et al. (2011).

El BSQ es una escala ampliamente utilizada para evaluar la insatisfacción corporal (Pook, Tuschen-Caffier & Brähler, 2008). Se trata de un instrumento

que ha demostrado tener buena consistencia interna y validez concurrente, además permite discriminar entre personas con algún trastorno alimentario y los que no (Cooper et al., 1987; Evans & Dolan, 1993; Pook et al., 2008; Raich et al., 1996). Basándose en la afirmación de Burisch (1984, citado por Evans & Dolan, 1993) que demuestra que en versiones abreviadas un número de medidas muestran correlaciones de validación iguales a sus versiones largas, incluso cuando estas últimas presentan una consistencia interna alta, Evans & Dolan (1993) sugieren el uso de una versión de 16 ítems extraída del BSQ que puede ser más eficiente, además de cuatro versiones de 8 ítems que consideran suficientemente robustas para ser usadas como “formas alternativas” donde la velocidad de aplicación y economía son esenciales.

Las cuatro versiones de 8 ítems sugeridas por Evans & Dolan (1993) en suma son 32 de los 34 ítems que forman el BSQ, según los autores, el ítem 32 y el 26 fueron eliminados dadas sus cargas relativamente bajas en el primer componente y en el argumento de que representan correlaciones muy específicas con el trastorno de la imagen corporal, cuyas respuestas pueden ser muy sesgadas en las muestras de población en general. Raich et al. (1996) llegaron a una conclusión similar basándose en el análisis de los ítems del BSQ-34.

Aunque las cuatro versiones cortas del BSQ dieron muy buen ajuste en el diseño y en el test-retest de acuerdo con Evans & Dolan (1993), la versión denominada BSQ-8C por Pook et al. (2008) que contiene los ítems 4, 6, 13, 16, 19, 23, 29 y 33 del BSQ-34 es la versión más favorable en cuanto a sensibilidad al cambio y propiedades psicométricas, siendo la recomendada para la investigación (clínica y no clínica).

En México existen pocos estudios sobre el tema de la imagen corporal y en los que se ha utilizado en BSQ este ha tenido un comportamiento psicométrico similar, ya sea en sujetos femeninos, sobre aspectos de la percepción de la imagen, la insatisfacción corporal, la sintomatología de anorexia nerviosa o bulimia nerviosa en estudiantes universitarios y la influencia del modelo estético (Ochoa, Villareal, Molina & Bayle, 2008). Salazar

(2012) y García (2015) utilizan la versión corta BSQ-8C en jóvenes adolescentes en México, en población tanto masculina como femenina mostrando sólo resultados descriptivos de los resultados.

Para la clasificación de los resultados en jóvenes se utiliza la misma señalada por Cooper et al. (1987) con las adecuaciones indicadas por Evans & Dolan (1993) (ver tabla 15).

**Tabla 15.** Valores del Test BSQ-8

Puntos	Clasificación
8	No preocupado / Satisfacción corporal
9 a 22	Leve preocupación / Miedo a engordar
23 a 28	Moderada preocupación / Baja autoestima
29 a 40	Extrema preocupación Deseo de perder peso

**Fuente:** Elaboración propia con base en Evans & Dolan (1993).

López, Solé & Cortés (2008) señalan que:

“Un resultado de 8 puntos en el cuestionario, es decir estar clasificado con satisfacción corporal significa que en algún momento remoto puede sentir insatisfacción corporal pero no es reconocido por el adolescente como algo prioritario.

Ubicarse en una segunda clasificación, es decir tener una leve preocupación, significa que se tiene un pensamiento ocasional en conseguir una imagen corporal lo más parecida al modelo sociocultural establecido, aportando en algún momento miedo a engordar en los pensamientos del adolescente sin trascender a un pensamiento angustioso o negativo para él, lo cual manifiesta una preocupación moderada o normal por su imagen corporal

Tener una moderada preocupación por la figura implica que el adolescente tiene un pensamiento angustioso, negativo que le afecta, provocando insatisfacción con referencia a su imagen corporal al alejarse de la imagen de modelo corporal ideal preestablecido por el adolescente

y, que al compararlo con su imagen corporal real no corresponde. Entonces en el adolescente se acentúan los valores negativos produciéndose una baja autoestima por su apariencia física y un deseo de perder peso y de hacer dieta. Así pues, aparece una preocupación extrema por su figura y peso. Todo ello le incluye dentro de un margen de riesgo de manifestar Trastorno de la conducta Alimentario (TCA)

Tener una preocupación extrema significa que el pensamiento angustioso y negativo de su imagen corporal ocupa prácticamente todos los pensamientos del adolescente y su deseo por perder peso se convierte en su objetivo, en su máxima finalidad y para ello puede desarrollar estrategias de todo tipo para conseguirlo. Cuanto mayor sea la preocupación por el peso y las dietas, tanto mayor es la insatisfacción por la imagen corporal, peor la autoimagen general, más depresivo estará, más malhumor le envolverá y mayor malestar asintomático tendrá en general. Este pensamiento negativo le afecta provocando insatisfacción con referencia a su imagen corporal al alejarse de la imagen de modelo corporal ideal preestablecido por el adolescente y, que al compararlo con su percepción de imagen corporal que ve no se corresponde e incluso puede no corresponderse a la imagen real que tiene, conduciendo al adolescente a una situación de trastorno dismórfico le conduce a un riesgo de manifestar Trastorno en la Conducta Alimentaria (TCA)” (pp. 130-131)

#### **IV.1.12 Figuras de Stunkard**

“Las investigaciones sobre la alteración de la imagen corporal han incrementado en años recientes, en gran parte debido al amplio reconocimiento de su papel fundamental en los trastornos alimentarios y la obesidad” (Gardner, Jappe & Gardner, 2009, p. 114). Dentro de los métodos que han sido usados para medir varios componentes de la imagen corporal, se incluyen cuestionarios, entrevistas, dibujos y técnicas de distorsión de la imagen (Rowe

et al., 2005 citado por García & Garita, 2007). De acuerdo con Marrodán et al. (2008), los criterios que se vienen utilizando más habitualmente son determinados test psicométricos como los que recogen los protocolos denominados DSM-IV (Manual diagnóstico y estadístico de los trastornos mentales) y CIE-10 (Clasificación Internacional de Enfermedades, *CIE* por sus siglas en inglés, en su décima edición) y que se basan en signos y evidencias de comportamientos alterados que manifiesta el sujeto. “A pesar de su amplia difusión presentan ciertas limitaciones, ya que se orientan casi exclusivamente al diagnóstico de la anorexia y bulimia, que son sólo una parte de los TCA (Trastornos Alimentarios). Además, no son aplicables en la fase temprana de la enfermedad” (Marrodán et al., 2008, p. 16).

Las Figuras de Stunkard es un instrumento pictórico desarrollado por Stunkard & Stellar (1990), modificado y adaptado a niños y adolescentes por Collins (1991) para examinar la percepción de la figura corporal. Se trata de nueve figuras esquemáticas tanto masculinas como femeninas, ubicadas desde muy delgado (silueta 1) hasta obesidad severa (García & Garita, 2007; Trejo, 2015). Cada silueta tiene asignado un IMC correspondiente de manera que la más delgada equivaldría a un IMC de 17 y la más obesa de 33 (Salazar, 2012). Los sujetos son invitados a seleccionar la figura que representa su actual talla y luego la talla que consideran ideal; la diferencia entre los datos es llamado el índice de discrepancia y es considerado para representar la insatisfacción corporal (Acosta & Gómez, 2003; Casillas-Estrella, Montañó-Castrejón, Reyes-Velázquez, Bacardí-Gascón & Jiménez-Cruz, 2006; García, 2015; Gardner, Stark, Jackson & Friedman, 1999; Marrodán et al., 2008; Salazar, 2012).

Cuanto mayor sea la diferencia resultante mayor será la insatisfacción; una diferencia positiva indica la insatisfacción que el adolescente siente porque desea estar más delgado mientras que una negativa es la insatisfacción que se siente porque se desea estar más grueso, y una diferencia igual a cero se interpreta como satisfacción (Acosta & Gómez, 2003; García, 2015) (ver tabla 16).

Tabla 16. Categorización del índice de discrepancia

Índice de discrepancia	Clasificación
Diferencia nula	Satisfacción con la imagen corporal.
Diferencia positiva	Insatisfacción en cuanto al deseo de una figura más delgada.
Diferencia negativa	Insatisfacción en cuanto al deseo de una figura más gruesa.

Fuente: García (2015).

Los test de autopercepción de la imagen constituyen un buen instrumento de detección precoz de la distorsión e insatisfacción con la propia figura que puede llegar a desembocar en una grave alteración de la conducta alimentaria (Smolak, 2004; Cash, 2004; Mutusek et al., 2004 citados por Marrodán et al., 2008). El inicio de percepciones dispares de la figura y las expectativas con respecto a la delgadez entre las mujeres puede ser evidente a partir de los 6 y 7 años de edad (Collins, 1991); por lo que puede ser perfectamente utilizado en jóvenes preuniversitarios.

En México ha sido ampliamente utilizado mostrando ser un test confiable en adultos y jóvenes (Casillas-Estrella et al., 2006; Cortes-Martínez, Vallejo-De la Cruz, Pérez-Salgado & Ortiz-Hernández, 2009; García, 2015; Salazar, 2012; Trejo, 2015).

---

V

**MÉTODOS**

---

## V.1 Sobremuestreo

Cuando se están estudiando grupos minoritarios de una población (p.e. inmigrantes, enfermedades poco comunes, detección de fraudes, niños con actitudes sobresalientes, grupos étnicos, gestión de riesgos, etcétera) el hacer inferencias significativas sobre el comportamiento de la agrupación se vuelve una situación complicada, aumentando la dificultad del análisis al hacer comparativas con el resto de los individuos (o unidades taxonómicas). Analizar datos desequilibrados en una nube de puntos multivariante no es trivial, especialmente cuando se trabajan con grandes bases de datos.

Una solución a este problema se encuentra en los métodos de muestreo estratificado, de esta forma, la población se divide en diferentes subgrupos o estratos (mutuamente excluyentes y colectivamente exhaustivos) según la característica que se desea analizar (por ejemplo, edad, país de origen, desempeño académico, etcétera), luego los sujetos son seleccionados al azar en cada uno de los estratos para elegir la muestra. Esta selección puede ser simple o proporcional según los intereses de los investigadores. El problema es que estos métodos sólo toman en cuenta una proporción de los individuos del grupo que se desea analizar (clase minoritaria), al ser pequeña la población y considerar menos cantidad para el análisis se pierde información que puede ser crucial para el estudio. Además, se puede caer en el “efecto de enmascaramiento”<sup>10</sup> ya que en muchos casos el comportamiento de la clase mayoritaria puede ocultar el comportamiento de la minoritaria.

El problema del desequilibrio de los datos es uno de los problemas (relativamente) nuevos surgidos cuando el *machine learning* (aprendizaje automático o aprendizaje de máquina) maduró de una ciencia embrionaria a

---

<sup>10</sup> “el efecto de enmascaramiento es la tendencia que tienen observaciones extremas, a esconder o enmascarar el efecto de observaciones más extremas que sí son realmente outliers” (Gil & Vargas, 1993, p. 38).

una tecnología aplicada, ampliamente utilizada en el mundo de los negocios, industria y la investigación científica (Chawla, Japkowicz & Ko, 2004). Los métodos de *machine learning* se enfrentan al problema de construir clasificadores a partir de datos desbalanceados por clase, donde los métodos convencionales tienden a construir modelos que seleccionan instancias de clase mayoritaria; una alternativa de solución a este problema es el sobremuestreo (*oversampling*) el cual es un método de muestreo utilizado comúnmente para datos desbalanceados, lo que aumenta el número de muestras de clase minoritaria para reducir el grado de distribución desigual (Siddiqi, 2006; Zhao et al., 2016). En general existen dos tipos de métodos de sobremuestreo uno donde se trabaja sobre el algoritmo (al tener que asignar un nuevo dato a un análisis) y los niveles de aproximación de los datos; y la solución a nivel externa donde se incluyen diferentes tipos de técnicas de remuestreo a los datos (Dubey, Zhou, Wang, Thompson & Ye, 2014).

Los enfoques algorítmicos implican el diseño de nuevos algoritmos de clasificación o la modificación de los existentes para manejar el sesgo introducido debido al desequilibrio de las clases, donde muchos investigadores estudian el problema en relación con la sensibilidad al costo de una asignación errónea, así la regla de asignación va modificándose según las clasificaciones correctas o incorrectas realizadas a las distintas muestras (Akbari, Kwek & Japkowicz, 2004; Bradford, Kunz, Kohavi, Brunk & Brodley, 1998; Chen, Liaw & Breiman, 2004; Elkan, 2001; Knoll, Nakhaeizadeh & Tausend, 1994; Pazzani, Merz, Murphy & Ali, 1994).

Las soluciones a nivel de los datos (externo) incluyen técnicas de sobremuestreo o submuestreo; el primero donde se sobre-examina a la clase minoritaria para que coincida con el tamaño de la clase mayoritaria y el segundo donde se subestima la clase mayoritaria (Liu, Yu, Huang & An, 2009). Estas técnicas de remuestreo aleatorio seleccionan de forma aleatoria los puntos de datos que se van a replicar, mismos que pueden incurrir en un costo de pérdida de información al realizar el ajuste (Dubey et al., 2014).

Japkowicz (2000) propone remuestrear los individuos de las clases minoritarias cercanos al límite de la otra clase. Kubat & Matwin (1997) plantean remuestrear la clase mayoritaria de manera que los puntos de datos del límite sean eliminados. Yen & Lee (2009) sugieren el enfoque de submuestreo basados en clúster para la selección de los datos representativos como datos de entrenamiento para mejorar la precisión de la clasificación. Liu et al. (2009) desarrolla dos sistemas de aprendizaje conjunto para superar la deficiencia de la pérdida de información introducida en el método de submuestreo. Chawla, Bowyer, Hall & Kegelmeyer (2002) diseñaron un método (SMOTE) que combina el sobremuestreo de la clase minoritaria y el submuestreo de la clase mayoritaria en la búsqueda de un mejor clasificador. Padmaja, Krishna & Bapi (2008) recomiendan el algoritmo de la predicción de las minorías basadas en el filtro de las mayorías (MFMP) basado en dos pasos: en el primero las minorías se agrupan y se identifican mejorando su tasa de predicción; en el segundo paso, la mayoría de las muestras se seleccionan aleatoriamente en grupos individuales. Japkowicz (2000) emplea algoritmos de k-medias sobre cada clase por separado para sobremuestrear aleatoriamente hasta conseguir un equilibrio entre las clases. Jo & Japkowicz (2004) proponen el clúster basado en oversampling (CBOS). En esta línea Dotan-Cohen, Melkman & Kasif (2007) también utiliza algoritmos de clustering pero a través de un algoritmo aglomerativo jerárquico. Zheng & Srihari (2003) proponen una solución de nivel de características basada en seleccionar y combinar óptimamente características positivas y negativas.

Un ámbito donde el sobremuestreo es utilizado con regularidad para los modelos predictivos es en el cálculo de riesgos crediticios en banca (Crone & Finlay, 2012; Mallo, 2011). Para el ajuste de modelos de riesgo crediticio con sobremuestreo existen dos métodos principales, uno por medio de los pesos de muestreo y otro como método compensatorio (Siddiqi, 2006). La decisión sobre la mejor técnica está influenciada en gran medida por la naturaleza del clasificador y la medida de efectividad (Van Hulse, Khoshgoftaar & Napolitano, 2007). En el caso de esta investigación los objetivos de realizar sobremuestreo son maximizar grupo minoritario con el fin de capturar los distintos perfiles de comportamiento y eliminar los posibles sesgos que se producen al momento de

extraer los datos en conjunto, por lo que se trabajó por medio de los pesos de muestreo propuesto por Siddiqi (2006), el cual consiste en la selección deliberada de grupo minoritario.

El método de sobremuestreo estratificado por medio de pesos de muestreo de Siddiqi (2006) para obtener estimaciones precisas sobre las propiedades de datos desbalanceados, considera que se debe de tomar el total del grupo minoritario y hacer una extracción aleatoria de, al menos el doble del grupo mayoritario con el fin de garantizar la existencia de una muestra de modelización suficientemente grande. Al considerar cada caso en la muestra del grupo mayoritario como representativa de la población total, se utiliza una variable como el peso de la evidencia (*WOE* por sus siglas en inglés) para ponderar las observaciones y poder recuperar la tasa real en la población; esta se calcula a través del cociente del número total del grupo mayoritario de la población dividido entre la cantidad del grupo mayoritario incluidos en la muestra (Siddiqi, 2006).

El ajuste para el sobremuestreo ha producido pronósticos realistas proporcionando información sobre el desempeño de las cuentas en banca y estudios de crédito, durante los agrupamientos de atributos, proporcionando así una ventaja comercial valiosa, así como beneficios estadísticos (Mallo, 2011; Siddiqi, 2006). Dadas las similitudes que tienen los datos obtenidos con jóvenes preuniversitarios con los señalados por Mallo (2011) y Siddiqi (2006) donde el interés debe enfocarse en aquellos adolescentes con EAS sin omitir las características de los estudiantes con ASF.

## V.2 Análisis de Clúster

### V.2.1 Antecedentes

Cuando todas las variables que intervienen en el estudio son consideradas al mismo nivel, es decir, no existe una variable respuesta, la forma estándar de llevar a cabo la clasificación de los individuos en grupos homogéneos, es el Análisis de Clúster (Dorado, 1998). Se trata de una gran variedad de técnicas que tienen como objetivo primordial la búsqueda de grupos en un conjunto de individuos (Fernández, 1991). No se trata de un método único, sino un procedimiento estadístico multivariante que pretende “clasificar los elementos de un conjunto en subconjuntos homogéneos, siguiendo un determinado criterio de clasificación donde cada elemento pertenece a un único subconjunto que a menudo tiene un nombre que lo caracteriza” (Cuadras, 2014, p. 187).

El análisis de clúster es una técnica ampliamente utilizada en aplicaciones de minería de datos para descubrir patrones subyacentes en los datos (Chiu, Fang, Chen, Wang & Jeris, 2001). Para realizar este análisis se consideran los tres aspectos que Fernández (1991) considera necesarios:

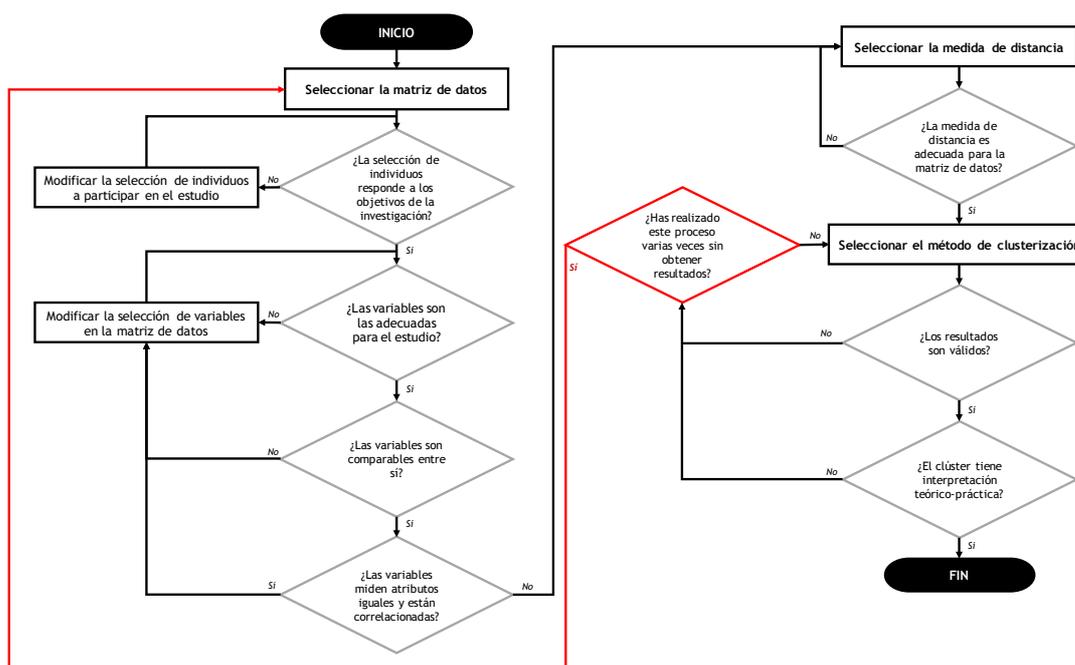
1. Selección de la matriz de datos
2. Selección de la medida de distancia
3. Selección del método de clusterización.

Añadiendo la cuarta etapa que consideran básica Aldenderfer & Blashfield (1984):

4. Validación de los resultados obtenidos

De esta forma el proceso de clusterización se lleva a cabo a través de un diagrama que permite volver al inicio siempre que se considere necesario. En la figura 4 se presenta como un proceso iterativo que puede modificarse según la matriz de los datos, las medidas de distancias y los métodos existentes.

Figura 4. Proceso para llevar a cabo un análisis de clúster



Fuente: Elaboración propia.

Así, una vez establecidos los objetivos del análisis de clúster la selección de la matriz de datos es el primer aspecto a considerar para realizar el análisis de acuerdo con Fernández (1991), o lo que es lo mismo tener en cuenta los individuos a agrupar y las variables en función. De acuerdo con Fernández (1991) y Martínez (1984), el número de variables no debe ser muy elevado ya que introducir una gran cantidad no hace sino perjudicar la interpretabilidad de los resultados. Si es muy grande, se podría realizar antes un Análisis de Componentes Principales para, de esta forma, poder resumir el conjunto de variables óptimas y reducir la dimensionalidad del problema (Santos, 2015).

Martínez (1984) señala que la unidad de medida debe ser la misma para todas las variables incluidas en el estudio. Fernández (1991) considera que las variables deben ser continuas<sup>11</sup>, comparables entre sí, tanto en su lógica como en sus escalas, todas relativas a un mismo tema para de esta forma proporcionar

<sup>11</sup> Actualmente existen técnicas que permiten trabajar con datos dicotómicos o con atributos nominales, ordinales o mixtos, lo cual se abordará posteriormente en este trabajo.

las mejores condiciones para obtener un resultado de clúster que tenga sentido. Así también, es importante que las variables no estén correlacionadas (Fernández, 1991; Martínez, 1984), si miden la misma dimensión, entonces ese aspecto recibe una oportunidad mayor a tener efecto en el proceso de agrupación.

Uno de los factores que ocasionan mayores divergencias en los resultados es la elección de la medida de similitud destinada a cuantificar la separación entre las unidades de análisis (Fernández, 1991). Es necesario establecer una medida de asociación entre los elementos de un clúster que permita medir la proximidad de éstos para realizar el método de agrupamiento (Santos, 2015). Existe una lista extensa de medidas de distancia para datos de tipo *cuantitativos o de intervalo*<sup>12</sup>, en cambio, cuando los atributos son binarios, nominales, ordinales o de tipo mixto, las opciones para medir las similitudes (o disimilitudes) se reducen.

Una vez considerada la matriz de datos y determinada la medida de distancia según las características de las variables, el siguiente paso es la selección del método de agrupación. La elección del método a emplear, depende de la naturaleza de los datos que se vayan a utilizar y del objetivo de la investigación (Santos, 2015). Los distintos procedimientos estadísticos para llevar a cabo el análisis de clúster, pueden dividirse según se haga la clasificación de los individuos (unidades taxonómicas) entre jerárquicas o no jerárquicas (Cuadras, 2014). La diferencia principal es la forma de asignar los individuos en cada clúster, en los métodos jerárquicos las asignaciones que se crean permanecen estables durante todo el proceso y no se permiten reasignaciones, mientras que en los no jerárquicos la reasignación es posible a posteriori (Santos, 2015).

---

<sup>12</sup> Debido a que los datos analizados en esta investigación no tienen esta característica no se enfocará el análisis en su estudio. Para más información referente a medidas de disimilaridad en variables cuantitativas pueden consultarse los trabajos de Gower (1985) y Cuadras (1989).

La gran proliferación de métodos impide desarrollar cada uno de los métodos de clusterización en esta tesis, un claro ejemplo de la bastedad de métodos existentes puede observarse en la figura 5. Si se desea realizar un análisis más detallado sobre cada uno de ellos puede consultarse Aldenderfer & Blashfield (1984); Cuadras (2014); Fernández (1991); Gordon (1998); Martínez (1984); Pedret (1986); y Santos (2015).

**Figura 5.** Diferentes tipos de métodos de análisis de clúster



**Fuente:** Técnicas cuantitativas al servicio del marketing: métodos descriptivos de análisis multivariable (Pedret, 1986).

Finalmente después de aplicar la técnica, debe validarse la decisión sobre el número ideal de clústers. Algunas variantes para la evaluación del número de clústeres encontrados son la Raíz Cuadrática Media (RMSSTD),  $R^2$ ,  $R^2$  semiparcial, y la distancia entre dos clúster. Fernández (1991) señala que esta labor se satisface cuando el resultado tiene una interpretabilidad teórico-práctica.

### V.2.2 TwoStep Cluster (Clúster bietápico)

El procedimiento de análisis de conglomerados en dos fases, también llamado bietápico (TwoStep), es una herramienta de exploración diseñada para descubrir las agrupaciones naturales de un conjunto de datos con la característica de que es un procedimiento automático del número óptimo de conglomerados, permite trabajar con datos categóricos como continuos y permite trabajar con archivos de datos de gran tamaño (Rubio-Hurtado & Vilà-Baños, 2017).

La mayoría de los algoritmos tradicionales de agrupamiento están limitados a manejar conjuntos de datos que contienen atributos continuos o categóricos, sin embargo, los de tipo mixto son comunes en los problemas de minería de datos (Chiu et al., 2001) donde los datos recogidos en el mundo real a menudo contienen atributos numéricos y categóricos (Ming-Yi et al., 2010). La elección de una medida de distancia inadecuada puede sesgar el tratamiento y por tanto los resultados (IBM Knowledge Center, 2001). A partir de esta problemática, derivado de una medida a partir de un modelo de mezcla Gaussiana, equivalente a la disminución en la log-verosimilitud resultante de la combinación de dos grupos, Banfield & Raftery (1993) introducen una medida en modelos para datos con variables continuas. Meila & Heckerman (1998) aplican este concepto probabilístico y derivaron una medida de distancias, pero para datos categóricos.

Inspirados en el algoritmo ROCK (algoritmo de agrupación robusta para los atributos categóricos) de Guha, Rastogi & Shim (2000) quienes desarrollan

un algoritmo robusto de agrupación jerárquica empleando enlaces y no distancias cuando se fusionan clústeres, y extendiendo la medida de distancia de Banfield & Raftery (1993) basada en modelos a situaciones que incluyen variables continuas y categóricas el método de TwoStep propuesto por Chiu et al. (2001) extiende la medida de cuyo cálculo es eficiente en cuanto a la memoria (rapidez), ya que sólo depende del par de racimos de fusión y no de todos los otros clústeres, lo cual lo hace idóneo para el Big Data.

Otro de los problemas de los métodos “tradicionales” se presenta cuando se tiene un número amplio de unidades a clasificar ya que los métodos clásicos de agrupamiento son efectivos y precisos en un pequeño conjunto de datos (Chiu et al., 2001; IBM Knowledge Center, 2001; Santos, 2015). Basados en el concepto de que los métodos tradicionales agruparían grandes conjuntos de datos de forma efectiva si estos fueran reducidos primero en conjuntos más pequeños, Zhang, Ramakrishnan & Livny (1996) proponen el método BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*). Su método propone hacer el proceso en dos etapas, donde en la primera se aplica un método de agrupación secuencial rápida al conjunto de datos grandes para comprimir las regiones densas y formar subgrupos; y en la segunda, el método de clúster será aplicado a estos sub-clústeres para encontrar un número deseado de clústeres (IBM Knowledge Center, 2001). El método BIRCH es el primer algoritmo de agrupamiento vinculado al ámbito de las bases de datos (Santos, 2015).

En otras palabras, al igual que el método BIRCH, en el algoritmo TwoStep se realiza como primer paso un pre-agrupamiento escaneando el conjunto de datos completo y almacenando las regiones densas de los registros de datos en términos de estadísticas de resumen; seguido de la aplicación de un algoritmo de agrupamiento jerárquico para unir las regiones densas (Chiu et al., 2001). Para la agrupación jerárquica basada en modelos pueden utilizarse distintas estrategias para determinar el número de clústeres. Fraley & Raftery (1998) proponen utilizar el criterio de información Bayesiano (BIC); Banfield & Raftery (1993) sugieren utilizar el peso aproximado de la evidencia (AWE).

El TwoStep Clúster maneja variables continuas y categóricas extendiendo la medida de distancia basada en modelos utilizada por Banfield & Raftery (1993) utilizando el enfoque de agrupamiento en dos etapas similar al BIRCH (Zhang et al., 1996) proporcionando la capacidad de encontrar automáticamente el número óptimo de clústeres (Chiu et al., 2001).

“El componente SPSS TwoStep Clustering es un algoritmo de análisis de clúster escalable diseñado para manejar grandes conjuntos de datos. Capaz de manejar variables continuas y categóricas, requiere solamente un paso en el procedimiento de los datos. En el primer paso del procedimiento, pre-agrupa los registros en muchos sub-clústeres pequeños. A continuación, agrupa los sub-clústeres del paso anterior en el número deseado de clústeres. Si el número deseado de clústeres es desconocido, el componente SPSS TwoStep Cluster encontrará el número adecuado automáticamente” (IBM Knowledge Center, 2001, p. 2).

El TwoStep Clúster difiere del BIRCH en que se agrega un procedimiento que permite que el algoritmo determine automáticamente el número apropiado de clústeres y una nueva estrategia de asignar la pertenencia de clúster a datos atípicos (outliers) (Chiu et al., 2001).

#### ***V.2.2.1 Paso 1 del TwoStep: pre-clustering***

En este primer paso se utiliza un enfoque de agrupamiento secuencial propuesto por Theodoridis & Koutroumbas (1999) analizando los registros uno a uno y decidiendo si el registro actual debe fusionarse con los clústeres previamente formados o iniciar un nuevo clúster basado en un criterio de distancia (IBM Knowledge Center, 2001). El objetivo es calcular una nueva matriz de datos con menos casos para el siguiente paso (Bacher, Wenzig & Vogler, 2004). Para hacer esto, se implementa el procedimiento construyendo un árbol de características de agrupamiento modificado, propuesto por Zhang et al. (1996).

“El árbol de características de agrupamiento (*CF* por sus siglas en inglés) consiste en niveles de nodos, y cada nodo contiene un número de entradas. Una entrada de hoja (en el nodo hoja) representa un subclúster que desea. Los nodos no-hoja y sus entradas guían un nuevo registro en un nodo de hoja correcto rápidamente” (IBM Knowledge Center, 2001, p. 3).

Una entrada al *CF* es un triple resumen de la información que se mantiene sobre un grupo de puntos de datos y es descrito por Zhang, Ramakrishnan & Livny (1997, pp. 147-148) como:

Dado  $N$  puntos de datos  $d$ -dimensionales en un clúster:  $\{\vec{X}_i\}$  donde  $i = 1, 2, \dots, N$ . La entrada de la característica de agrupación (*CF*) del clúster se define como un triple:  $CF = (N, \vec{LS}, SS)$ , donde  $N$  es el número de puntos de datos en el clúster,  $\vec{LS}$  es una suma lineal de los  $N$  puntos de datos, es decir,  $\sum_{i=1}^N \vec{X}_i$ , y  $SS$  es la suma cuadrada de los  $N$  puntos de datos ( $\sum_{i=1}^N \vec{X}_i^2$ ).

Cuando dos clústeres se fusionan, se asume el teorema de la aditividad, si  $CF_1 = (N_1, \vec{LS}_1, SS_1)$  y  $CF_2 = (N_2, \vec{LS}_2, SS_2)$  son las entradas *CF* de dos subclústeres separados. Entonces la entrada *CF* del subclúster que se forma mediante la fusión es:

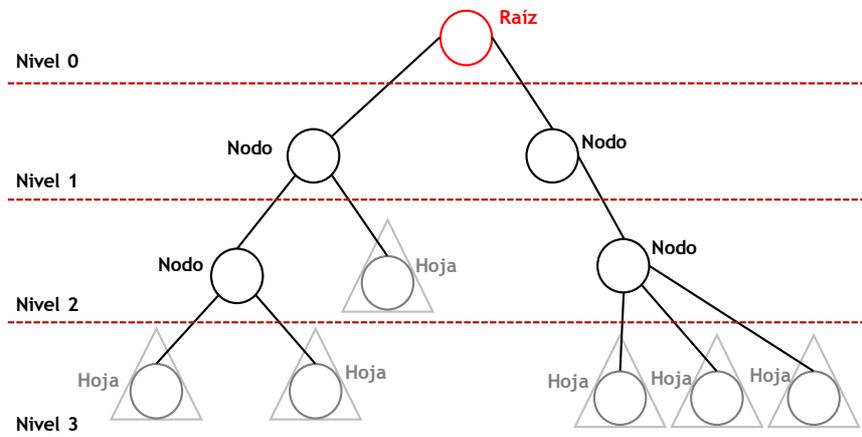
$$CF_1 + CF_2 = (N_1 + N_2, \vec{LS}_1 + \vec{LS}_2, SS_1 + SS_2)$$

Esta entrada *CF* no sólo es compacta porque almacena mucho menos que todos los puntos de datos en el subclúster, también es precisa porque es suficiente para calcular las mediciones necesarias para tomar decisiones de clúster en BIRCH, “las entradas *CF* se pueden almacenar y calcular de forma incremental y consistente a medida que se fusionan los subgrupos o se insertan nuevos puntos de datos” (Zhang et al., 1997, p. 148).

Santos (2015) presenta la figura 6 donde de forma visual muestra como el árbol *CF* consiste en una serie de nodos organizados por niveles y a su vez,

de cada nodo salen ramas que pueden contener también más nodos o terminar en una hoja, esa hoja representa el subclúster que se está buscando.

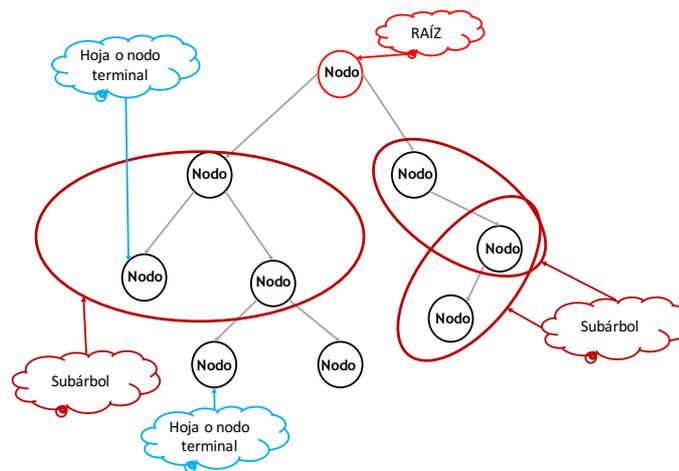
**Figura 6.** Gráfica de formación del árbol de características (CF)



Fuente: ‘Two-Step Cluster’ en SPSS y técnicas relacionadas (Santos, 2015).

A las CF con el número de registros de la entrada, la media y la varianza de cada variable continua, más los conteos para cada categoría de cada variable categórica caracterizan cada entrada. Cada registro sucesivo, a partir del nodo raíz, es recursivamente guiado por la entrada más cercana en el nodo para encontrar el nodo hijo más cercano, luego desciende a lo largo del árbol CF. Al llegar a un nodo de la hoja, encuentra la entrada de hoja más cercana, si el registro está dentro de una distancia umbral de la entrada de hoja más cercana, la entrada de hoja la absorbe y actualiza la CF. De otra manera inicia su propia entrada de hoja. Si el nodo hoja no tiene espacio para crear una nueva entrada de hoja nueva, el nodo de hoja se divide en dos. Las entradas en el nodo original de la hoja, se dividen en dos grupos usando el par más lejano como semillas, redistribuyendo las entradas restantes basándose en los criterios más cercanos. Si un árbol CF crece más allá del número máximo de niveles, el árbol CF reconstruye el árbol existente aumentando el criterio de distancia umbral. El árbol CF reconstruido es más pequeño, por lo que tiene espacio para nuevos registros de entrada. Así este proceso continúa a través de un pase de datos completo” (IBM Knowledge Center, 2001, p. 4).

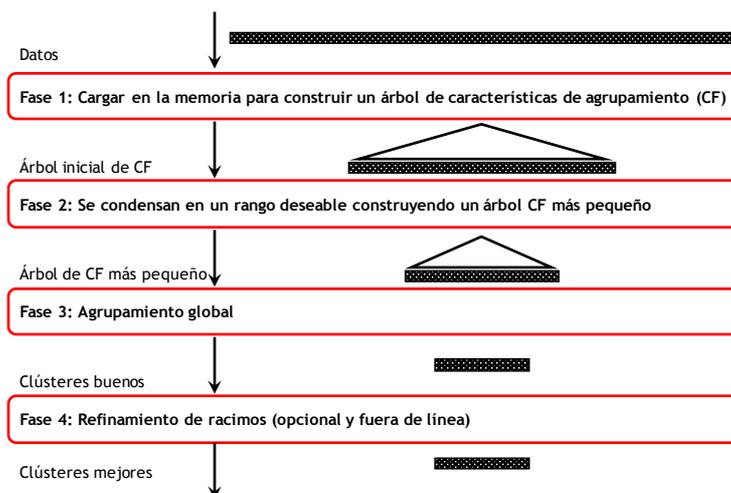
**Figura 7.** Ejemplo de descripción de las partes del árbol CF



**Fuente:** 'Two-Step Cluster' en SPSS y técnicas relacionadas (Santos, 2015).

En resumen, el algoritmo BIRCH se basa en cuatro fases. La figura 8 presenta el proceso mostrado por Zhang et al. (1996):

**Figura 8.** Descripción del BIRCH



**Fuente:** BIRCH: An Efficient Data Clustering Method for Very Large Databases (Zhang et al., 1996).

Es decir, para alcanzar el objetivo de calcular una matriz de datos con menos casos, donde los pre-clústeres calculados y sus características de agrupamiento se utilizan como nuevos casos Bacher et al. (2004). El algoritmo BIRCH presenta el inconveniente de que sólo funciona cuando las variables son continuas (Santos, 2015). Aquí una de las principales diferencias con el método BIRCH que sólo maneja variables continuas, ya que el método TwoStep incluye

conteos para cada variable categórica, así las características de agrupamiento representan colectivamente todos los registros que caen en la misma entrada (IBM Knowledge Center, 2001).

Debido a su enfoque de agrupamiento secuencial, donde se sigue a lo propuesto por Theodoridis & Koutroumbas (1999), los resultados pueden depender del orden de entrada de los casos, por lo que se recomienda utilizar un orden aleatorio para minimizar estos efectos, obteniendo incluso soluciones distintas con los casos organizados en distintos órdenes aleatorios para comprobar la estabilidad de una solución determinada (Bacher et al., 2004; Chiu et al., 2001; IBM Knowledge Center, 2001; Santos, 2015). Cuando los tamaños de archivo son demasiado grandes, pueden sustituirse varias ejecuciones por una muestra de casos estructurados con distintos órdenes aleatorios (Rubio-Hurtado & Vilà-Baños, 2017).

#### ***V.2.2.2 Paso 2 del TwoStep: clustering***

El paso dos se lleva a cabo tomando los subgrupos creados en el paso uno como entrada y agrupándolos en el número deseado de clústeres (Chiu et al., 2001). Dado que el número de subgrupos es mucho menor se pueden utilizar métodos tradicionales, SPSS utiliza el método de agrupación jerárquica aglomerativo (IBM Knowledge Center, 2001).

Se aplica una técnica jerárquica basada en un modelo distinto a las técnicas jerárquicas aglomerativas utilizando un modelo estadístico subyacente. Siguiendo a Bacher et al. (2004) el modelo asume que las variables continuas  $x_j$  ( $j = 1, 2, \dots, p$ ) están dentro del clúster  $i$  y son independientes con una distribución normal con medias  $\mu_{ij}$  y varianza  $\sigma_{ij}^2$ , y las variables categóricas  $a_j$  están dentro del clúster  $i$  con una distribución multinomial con probabilidades  $\pi_{ijl}$ , donde  $(jl)$  es el índice de la  $l$ -ésima categoría ( $l = 1, 2, \dots, m_j$ ) de la variable  $a_j$  ( $j = 1, 2, \dots, q$ ).

### V.2.2.3 Medidas de distancia

Dos medidas de distancia están disponibles para trabajar la disimilaridad entre los individuos, la distancia euclídea, para datos continuos, y la de log-verosimilitud que permite manejar atributos de tipo mixto.

Para realizar el cálculo de la distancia euclídea para datos continuos, las variables suelen estandarizarse previamente buscando que la escala de medición no distorsione los resultados, después se calcula el espacio euclídeo entre dos puntos a través de:

$$d(x, y) = \sqrt{\sum_{j=1}^J (x_j - y_j)^2}$$

donde  $d(x, y)$  es el valor de la distancia entre las unidades de análisis  $x$  e  $y$ ; y  $x_j$  e  $y_j$  son los valores que toman las variables desde 1 hasta  $j$ .

La distancia de log-verosimilitud entre dos clústers  $i$  y  $s$  se define por medio de:

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

donde:

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

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

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

$\xi_v$  se puede interpretar como una medida de dispersión (varianza) dentro del grupo  $v$  ( $v = i, s, \langle i, s \rangle$ ).  $\xi_v$  consta de dos partes, la primera  $-n_v \left( \sum_{j=1}^p \frac{1}{2} \log(\hat{\sigma}_{vj}^2 + \hat{\sigma}_j^2) \right)$  mide la dispersión de las variables continuas  $x_j$  dentro del clúster  $v$ . Si sólo  $\hat{\sigma}_{vj}^2$  fuera utilizado,  $d(i, s)$  sería exactamente la disminución en la función log-verosimilitud después de fusionar los clústeres  $i$  y  $s$ . El término  $\hat{\sigma}_j^2$  se añade para evitar la situación de degeneración para cuando  $\hat{\sigma}_{vj}^2 = 0$ . La entropía<sup>13</sup>  $-n_v \left( \sum_{j=1}^q \sum_{l=1}^{m_j} \hat{\pi}_{vjl} \log(\hat{\pi}_{vjl}) \right)$  se utiliza en la segunda parte como medida dispersión para las variables categóricas.

Similar a la agrupación jerárquica aglomerativa, los clústers con la distancia  $d(i, s)$  menor serán fusionados en cada paso. La función log-verosimilitud para el paso con  $k$  clústeres es calculada por medio de:

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

La función  $l_k$  no es exactamente la función log-verosimilitud pero puede ser interpretada como dispersión dentro de los clústeres. Si sólo se utilizan variables categóricas, entonces  $l_k$  se refiere a la entropía dentro de los  $k$  clústeres.

#### V.2.2.4 Predictor en el agrupamiento

Dada la importancia que tiene las variables en la formación de un clúster nuevo, tanto dentro como entre, se distinguen dos situaciones en la que la importancia del predictor está definida, siguiendo a Santos (2015) por:

$$VI_i = \frac{-\log_{10}(p - \text{valor}_i)}{\max_{j \in \Omega} (-\log_{10}(p - \text{valor}_j))}$$

donde  $\Omega$  denota la matriz de datos, el p-valor se calcula de la siguiente forma:

---

<sup>13</sup> Medida del desorden de un sistema (RAE, 2001c).

### V.2.2.4.1 Entre clústeres

- a) Para **variables categóricas**, basado en  $\chi^2$ , donde p-valor = Prob ( $\chi_d^2 > \chi^2$ )

$$\chi^2 = \sum_{i=1}^{I'} \sum_{j=1}^{J'} \frac{(N_{ij} - \hat{N}_{ij})^2}{\hat{N}_{ij}}$$

$$\hat{N}_{ij} = \frac{N_{i.} \cdot N_{.j}}{N}$$

donde los grados de libertad son  $(I' - 1, J' - 1)$ . Si  $N = 0$ , la importancia será indefinida o desconocida. Si  $N_{i.} = 0$  entonces  $I'$  se calcula como  $(I' = I - 1)$ . Si  $N_{.j} = 0$ , entonces  $J'$  se calcula como  $(J' = J - 1)$ . Si  $J' \leq 1$  o  $I' \leq 1$  la importancia será indefinida o desconocida.

- b) Para **variables continuas**, el p-valor se basa en una F de Snedecor (1938), donde p-valor = Prob $\{F(J - 1, N - 1) > F\}$

$$F = \frac{\sum_{j=1}^J (\bar{x}_j - \bar{x})^2 / (J - 1)}{\sum_{j=1}^J (N_j - 1) s_j^2 / (N - 1)}$$

donde los grados de libertad son  $(J' - 1, N - J')$ . Si  $N = 0$ , la importancia será indefinida o desconocida. Si  $N_j = 0$  entonces  $J'$  se calcula como  $(J' = J - 1)$ . Si  $J' \leq 1$  o  $N \leq J'$  la importancia será indefinida o desconocida. Si el denominador es cero, entonces el p-valor se tomará como 1.

### V.2.2.4.2 Dentro del clúster

- a) Para **variables categóricas** el p-valor está basado de forma que la proporción de individuos en el clúster  $j$  es la misma que la proporción global, mediante  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^{I'} \frac{(N_{ij} - N_j p_i)^2}{N_j p_i}$$

donde los grados de libertad son  $(I' - 1)$ . Si  $N_j = 0$ , la importancia será indefinida o desconocida. Si  $p_i = 0$ , entonces  $I'$  se calcula como  $(I' = I - 1)$ . Si  $I' \leq 1$ , la importancia será indefinida o desconocida.

- b) Para **variables continuas** el p-valor está basado en que la medida del clúster  $j$  es la misma que la media global, mediante una T-Student:

$$t = \frac{(\bar{x}_j - \bar{x})}{s_j / \sqrt{N_j}}$$

donde los grados de libertad son  $(N_j - 1)$ . Si  $N_j \leq 1$  o si  $s_j = 0$ , la importancia será indefinida o desconocida. Si el denominador es igual a cero, entonces el p-valor se tomará como 1 y se calculará como  $p - \text{valor} = 1 - \text{Prob}\{|T(d)| \leq |t|\}$ .

### V.2.2.5 Número de clústeres

El número de clústeres puede ser determinado automáticamente, donde se utiliza un estimador de dos fases a través del Criterio de Información de Akaike (1974) (AIC)<sup>14</sup>:

$$AIC_k = 2l_k + 2r_k$$

---

<sup>14</sup> Combina la teoría de la máxima verosimilitud, información teórica y la entropía de información (Motulsky & Christopoulos, 2004). De acuerdo con Santos (2015) el parámetro de penalización no distingue si el tamaño muestral es grande o pequeño por lo que no es consistente del número adecuado de factores comunes.

donde  $r_k$  es el número de parámetros independientes o el Criterio Bayesiano de Información (BIC) es calculado en la primera fase por medio de:

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

$BIC_k$  y  $AIC_k$  son una buena estimación del número máximo de clústeres (Chiu et al., 2001). El número máximo de clústeres se establece igual al número de clústeres donde la relación  $\frac{BIC_k}{BIC_1} \leq c_1$  por primera vez<sup>15</sup>.

La segunda fase utiliza el cambio de relación  $R(k)$  en la distancia para  $k$  clústeres, definido como:

$$R(k) = d_{k-1}/d_k$$

donde  $d_{k-1}$  es la distancia si  $k$  clústeres se agrupan en  $k - 1$  clústeres. El número de racimos se obtiene para la solución donde se produce un gran salto en el cambio de relación:

$$R(k_1)/R(k_2)$$

para los dos valores mayores de  $R(k)$  ( $k = 1, 2, \dots, k_{max}$ ;  $k_{max}$  obtenidos del primer paso). Si el cambio de la relación es mayor que el valor del umbral<sup>16</sup>  $c_2$  el número de agrupaciones se establece igual a  $k_1$ , de lo contrario el número de agrupaciones se establece igual a la solución con  $\max(k_1, k_2)$ .

### V.2.2.6 Calidad de los conglomerados

La “calidad de los conglomerados” se miden a través de la silueta de cohesión, de esta forma en los resultados que ofrece el SPSS (IBM, 2016) se presenta una

---

<sup>15</sup> Actualmente SPSS maneja  $c_1 = 0.04$  (Bacher et al., 2004).

<sup>16</sup> Actualmente SPSS maneja  $c_2 = 1.15$  (Bacher et al., 2004).

ventana resumiendo el resultado en “pobres”, “suficientes” y “buenos” según el coeficiente de silueta de cohesión, donde un valor de “1”, significaría que todos los casos se encuentran en sus centros de los conglomerados; y un valor de “-1” significa que todos los casos se encuentran en los centros de otros conglomerados a los que no pertenecen.

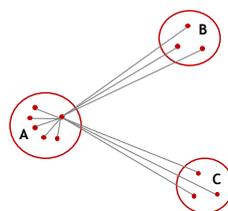
La medida de silueta de cohesión se obtiene, sobre todos los casos, calculando  $(B - A) / \max(A, B)$ , donde  $A$  es la distancia del caso a su centro del conglomerado y  $B$  es la distancia del caso al centro del conglomerado más cercano al que no pertenece (Rubio-Hurtado & Vilà-Baños, 2017).

Santos (2015) lo explica a través de la figura 9 y señala que para construir la medida de silueta se necesita la partición obtenida mediante el algoritmo y la colección de proximidades entre los individuos y objetos, en donde para cada objeto  $i$  se introduce un valor llamado  $s(i)$  determinado por

$$s(i) = \frac{b_1 - a_i}{\max\{a(i), b(i)\}}$$

Si se llama  $i$  al objeto que ha sido asignado al clúster “A”, se puede calcular la media de las disimilaridades entre este objeto y el resto de objetos dentro del clúster, de forma que se tendrá  $a(i)$  como medida de las disimilaridades de  $i$  con el resto de objetos incluidos en el clúster A;  $d(i, C)$  como la medida de las disimilaridades de  $i$  con todos los objetos incluidos en el clúster C, sabiendo que ambos clústeres A y B son distintos entre sí; y  $b(i) = \min d(i, C)$ .

**Figura 9.** Ejemplo gráfico del cálculo de las similaridades entre clústeres



**Fuente:** ‘Two-Step Cluster’ en SPSS y técnicas relacionadas (Santos, 2015).

Por lo tanto, el valor  $s(i)$  se obtiene como combinación entre los valores estimados anteriormente, de manera que:

$$s(i) = \begin{cases} 1 - (a_i/b_i) & \text{si } a_i < b_i \\ 0 & \text{si } a_i = b_i \\ (a_i/b_i) - 1 & \text{si } a_i > b_i \end{cases}$$

La interpretación de los resultados de la silueta de cohesión está basada en los trabajos de Kaufman & Rousseeuw (1990), donde un conglomerado en la zona de “buenos” significa que los datos reflejan de forma razonable o fuerte la estructura de los conglomerados; un resultado “suficiente” implica que los datos revelan de forma justa esta estructura de conglomerados, y un resultado en la región de “pobres” refleja que los datos no aportan evidencias significativas de la estructura de conglomerados.

Para los datos con variables mixtas, los resultados experimentales de Chiu et al. (2001) confirman que el algoritmo no sólo genera clústeres de mejor calidad que los algoritmos k-means tradicionales, sino que también presenta buenas propiedades de escalabilidad siendo capaz de identificar el número subyacente de clústeres en los datos correctamente. Gracias a sus ventajas sobre otras metodologías entre las que se encuentra el hecho de que procesa datos tanto cualitativos como cuantitativos, el análisis de clúster bietápico ha sido utilizado en distintos campos como: Salud Mental (Francisco, 2012); Economía de la Industria (Arrighetti & Ninni, 2014); Diagnóstico de Obesidad en el Ámbito Escolar (Sigmund, Sigmundová, Šnoblová & Gecková, 2014); Análisis Urbanístico (Alho & De Abreu, 2015); Análisis de los Factores de Suicidio en Adolescentes (Chin & Choi, 2015); entre otros.

### V.3 Descomposición de matrices

El uso de grandes matrices con las que se trabaja en algunos campos en la actualidad (medicina, biología, astronomía, química, economía, psicología, informática, por mencionar algunos) proporciona una estructura natural para codificar información sobre  $n$  objetos, cada uno de los cuales está descrito por  $p$  características. En algunos casos, un paso importante para la interpretación es construir una representación comprimida de esta matriz para que sea más fácil de analizar e interpretar a través de un campo específico del conocimiento. La representación más común se obtiene al realizar la Descomposición en Valores Singulares (SVD) (Eckart & Young, 1936) y el procedimiento más utilizado es el Análisis de Componentes Principales (PCA) (Hotelling, 1933; Pearson, 1901).

Recordemos que el SVD es una técnica de reducción de la dimensión que describe la información de un conjunto de variables observadas mediante un conjunto de variables más pequeño (componentes principales), por lo tanto, lo que busca es disminuir la dimensión y extraer características de un conjunto de datos de entrada, mediante una transformación que los proyecte en nuevas direcciones (componentes principales) absorbiendo la mayor cantidad posible.

Si se tiene una matriz  $X_{ij}$ . Dado  $X$  existen las matrices ortogonales  $U = [u^1, u^2 \dots u^j]$  y  $V = [v^1, v^2 \dots v^j]$ , donde  $u^j$  y  $v^j$  son tales que  $U^T X V = \Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$ , donde  $\Sigma \in \mathbb{R}^{i \times j}$ ,  $p = \min\{i, j\}$ ,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ , y  $\text{diag}(\cdot)$  representa una matriz diagonal con los elementos especificados en la diagonal. De esta forma  $X = U \Sigma V^T$  donde  $U$ ,  $V$  y  $\Sigma$  constituyen la SVD de  $X$ , los  $\sigma_i$  son valores singulares de  $X$  y los vectores  $u^i$  y  $v^i$  son el  $i$ -ésimo valor izquierdo y el  $i$ -ésimo vector singular derecho, respectivamente.

De esta forma, las componentes principales son unas variables compuestas incorreladas tales que unas pocas explican la mayor parte de la variabilidad, este método estima a partir de los factores comunes y prescinde de los factores únicos calculando los valores y vectores propios directamente de la matriz de correlaciones de las variables originales sin utilizar las

comunalidades en la diagonal (Cuadras, 2014). En las aplicaciones cabe esperar que las primeras componentes expliquen un elevado porcentaje de la variabilidad total.

El SVD es ampliamente utilizado en el análisis de datos, a menudo a través de métodos como PCA, en gran parte porque los subespacios generados por vectores proporcionan la mejor aproximación de rango  $k$  a la matriz de datos  $\mathbf{X}$  sin embargo, los vectores  $u^i$  y  $v^i$  por si mismos pueden carecer de significado en términos del campo en el cual se extraen y proyectan los datos (Mahoney & Drineas, 2009). Después de todo, los vectores singulares son abstracciones matemáticas que se pueden calcular para cualquier matriz de datos, no se trata de objetos o variables observables sino de conceptos; en algunos casos pueden interpretarse en términos de las direcciones de los ejes del elipsoide del que se dibujan los datos, sin embargo, esta interpretación no puede surgir de la matemática sola, sino que requiere un conocimiento del campo del cual se extraen los datos. Esto es problemático cuando se está interesado en extraer información de la salida de los algoritmos de análisis de datos y cuando se está interesado en los vectores experimentales reales no en bases creadas artificialmente.

Buscando la construcción de aproximaciones matriciales de rango bajo que dependen de elementos de datos reales que al igual que la SVD tengan optimalidad y propiedades algorítmicas, cuya interpretación sea estadística natural asociada a su construcción y que funcione bien en la práctica, Mahoney & Drineas (2009) proponen la CUR como matriz de descomposiciones. Es decir, descomposiciones de matriz de bajo rango que se expresan explícitamente en términos de un pequeño número de **columnas reales y/o filas reales de la matriz de datos original**.

Siguiendo a Mahoney & Drineas (2009), la matriz  $\mathbf{X}_{ij}$  se descompone como el producto de tres matrices  $\mathbf{C}$ ,  $\mathbf{U}$  y  $\mathbf{R}$ , donde  $\mathbf{C}$  es un pequeño número de las columnas reales de  $\mathbf{X}$ ;  $\mathbf{R}$  es un pequeño número de las filas reales de  $\mathbf{X}$ ; y  $\mathbf{U}$  es una matriz construida cuidadosamente que garantiza que el producto de la  $\mathbf{CUR} \approx \mathbf{X}$ . La medida en que  $\mathbf{X} \approx \mathbf{CUR}$  y en relación con la medida en que la

CUR se pueda usar en lugar de  $\mathbf{X}$ , dependerá sensiblemente de la elección de  $\mathbf{C}$  y  $\mathbf{R}$ , así como de la construcción de  $\mathbf{U}$ . Al tratarse de elementos de los datos reales y utilizarse en lugar de las columnas o filas propias, la interpretación será en términos y características del campo desde el cual se dibujen los datos.

Basados en estudios previos realizados por Berry, Pulatova & Stewart (2005); Drineas, Kannan & Mahoney (2006); Frieze, Kannan & Vempala (2004); Goreinov, Tyrtysnikov & Zamarashkin (1997); Goreinov & Tyrtysnikov (2001); Mahoney, Maggioni & Drineas (2006); y Stewart (1999), el método de mejora de la descomposición de matrices a través del apalancamiento (*leverage*) propuesto por Mahoney & Drineas (2009) consiste en construir  $\mathbf{C}$  (y de forma similar  $\mathbf{R}$ ), calculando un “puntaje de importancia” para cada columna de  $\mathbf{X}$ , probando aleatoriamente un pequeño número de columnas  $\mathbf{X}$ , utilizando esa puntuación como distribución de probabilidad de muestreo de importancia. Para seleccionar los puntos de muestreo se puede expresar la  $j$ -ésima columna de  $\mathbf{X}$  (denotada por  $X_j$ ) como:

$$X^j = \sum_{\xi=1}^r (\sigma_{\xi} u_{\xi}^{\xi}) v_j^{\xi}$$

donde  $r = \text{rango}(\mathbf{X})$  y  $v_j^{\xi}$  es la  $j$ -ésima coordenada del  $\xi$ -vector singular derecho, es decir, la  $j$ -ésima columna de  $\mathbf{X}$  es una combinación lineal de todos los vectores singulares izquierdos y valores singulares, y los elementos de la  $j$ -ésima fila de  $\mathbf{V}$  son los coeficientes. De esta forma se pueden aproximar  $X^j$  como una combinación lineal de los  $k$  vectores singulares de izquierda y los valores singulares correspondientes como:

$$A^j \approx \sum_{\xi=1}^k (\sigma_{\xi} u_{\xi}^{\xi}) v_j^{\xi}$$

Puesto que se buscan las columnas de  $\mathbf{X}$  que se correlacionen simultáneamente con la extensión de todos los  $k$  vectores singulares superiores, entonces se calculan las puntuaciones de los leverages normalizadas:

$$\pi_j = \frac{1}{k} \sum_{\xi=1}^k (v_j^\xi)^2$$

para todos los  $j = 1, \dots, n$ . Con esta normalización, es sencillo demostrar que  $\pi_j \geq 0$  y que  $\sum_{j=1}^n \pi_j = 1$ , y por lo tanto estas puntuaciones forman una distribución de probabilidad sobre las  $n$  columnas. Estas cantidades, hasta el escalamiento, son iguales a los elementos diagonales de la llamada “matriz de sombreros”, es decir, la matriz de proyección sobre el tramo de los  $k$  vectores singulares derechos de  $\mathbf{X}$ .

Con este procedimiento, la matriz  $\mathbf{C}$  contiene  $c'$  columnas, donde  $c' \leq c$  en la esperanza de que  $c'$  esté estrechamente concentrado con su expectativa. Esta puntuación depende de la matriz  $\mathbf{X}$ , y tiene una interpretación natural como la captura de la “influencia estadística” o “influencia” de una determinada columna en el mejor ajuste de bajo rango de la matriz de datos. Al elegir las columnas que ejercen una influencia desproporcionadamente grande sobre el mejor ajuste de rango bajo, aseguran que  $\mathbf{CUR}$  sea tan bueno como  $\mathbf{X}_k$  en la captura de la parte dominante del espectro de  $\mathbf{X}$ .

De esta forma Mahoney & Drineas (2009) proponen la selección aleatoria de filas y columnas de la matriz según la distribución de probabilidad determinada por sus leverages. Bodor, Csabai, Mahoney & Solymosi (2012) desarrollan el paquete rCUR en entorno R (R Development Core Team, 2013) y añaden además del método original (random) las opciones de la selección de tantas filas y columnas como se le solicite (*exact.num.random*); la elección de filas y columnas con los mayores puntajes de leverages devueltos de forma determinística (*top.scores*); la selección en una iteración basada en un factor que combine no sólo el puntaje del leverage sino también la ortogonalidad del siguiente vector con el subespacio ya seleccionado (*ortho.top.scores*) y; la opción de escoger filas y columnas con un alto nivel de leverage sobre un parámetro otorgado previamente (*highest.ranks*).

“Además de escoger columnas con alta influencia estadística se pueden aprovechar las ideas del análisis de regresión diagnóstica para aplicar las

descomposiciones de matriz **CUR**, como una herramienta para el análisis exploratorio de datos” (Mahoney & Drineas, 2009, p. 699). Los “puntos de importancia” o “leverage score” han sido ampliamente utilizados para la identificación de valores atípicos en el análisis de regresión de diagnóstico (Bodor et al., 2012). “Es importante señalar que este enfoque CUR es muy diferente a la perspectiva estadística clásica, donde los puntos de leverage se han utilizado en el análisis de regresión diagnóstica para identificar los valores atípicos y los errores” (Bodor et al., 2012, p. 14).

Dado que se trata de una metodología nueva, la descomposición de matrices a través de las CUR han tenido pocas aplicaciones, entre las que se encuentran su análisis en oncología sobre análisis en tumores (Bodor et al., 2012); la cuantificación del papel cambiante de las publicaciones (Orosz, Farkas & Pollner, 2016); y la identificación de iones y posiciones importantes en espectrometría de masa de datos de imágenes de composiciones químicas de muestras biológicas complejas (Yang, Rübél, Prabhat, Mahoney & Bowen, 2015).

## V.4 Análisis de Segmentación

### V.4.1 Antecedentes

El análisis de segmentación fue concebido y debe ser utilizado principalmente con un objetivo exploratorio (Escobar, 1998) con la finalidad última de, a través de sus resultados, desarrollar y crear estrategias de actuación para cada segmento objetivo (Kotler, 1988).

“El análisis de segmentación es una técnica estadística que pretende, a partir de la información que nos suministran ciertas variables (independientes o predictoras, respecto de una variable dependiente) clasificar un conjunto de individuos y observaciones, en grupos capaces de describir de la mejor manera posible la variable dependiente” (Dorado, 1998, p. 11).

Sin duda, los métodos de segmentación, como técnica de dependencia entre variables (Escobar, 1998) iniciada por Smith (1956) dentro del Marketing, conforman en la actualidad propuestas importantes en la agrupación de individuos, de acuerdo con determinadas características (perfil socio-demográfico, comportamiento de consumo, opinión, etcétera) (Avila, 1997; Castro, 2005; Dorado, 1998; Ramos, 2015).

En su uso se distinguen por un lado una variable cuya distribución se desea explicar y, por el otro, un conjunto de variables, nominales y ordinales, con estatus de independientes que reciben el nombre de pronosticadoras (Escobar, 1998). Las variables en análisis pueden ser de distintos tipos, algunas pueden ser medidas objetivamente (edad, peso, ingresos, nivel de estudios, etcétera), mientras que otras pueden ser inferidas o tratarse de mediciones subjetivas (actitudes, preferencias, percepciones, etcétera) (Frank, Massy & Wind, 1972). Construyendo de esta forma perfiles más precisos para los individuos (o unidades analizadas) que componen un colectivo bajo estudio, agrupados para conocer mejor los subgrupos poblaciones y así obtener mejores pronósticos sobre el comportamiento de grandes grupos de datos (Castro, 2005).

Kotler (1988) describe seis fases en un estudio de segmentación:

1. Identificación de las variables para realizar la segmentación.
2. Desarrollo de los perfiles de cada segmento obtenido.
3. Evaluación del atractivo de cada segmento.
4. Selección del segmento o segmentos objetivos.
5. Identificación de posibles motivos para posicionarse en los segmentos seleccionados.
6. Seleccionar, desarrollar y crear estrategias de actuación para cada segmento objetivo.

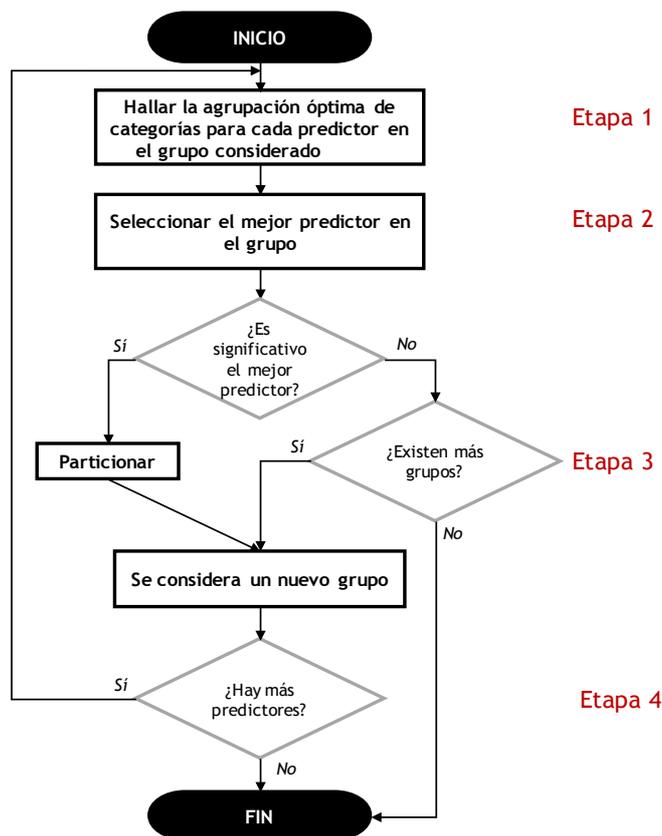
Tradicionalmente el análisis de segmentación se había reducido al estudio de variables dependientes cuantitativas, utilizando el algoritmo AID (Automatic Interaction Detection) presentado por Morgan & Sonquist (1963). Sin embargo, en la actualidad han proliferado infinidad de propuestas acerca de metodologías para la segmentación, muchas de ellas denominadas “sistemas de caja negra” es decir, métodos poco documentados o basados en criterios poco científicos (Castro, 2005).

El procedimiento para el análisis AID es de tipo iterativo, donde para obtener los grupos finales se desarrolla un proceso por fases o etapas, en cada una de las cuales se selecciona el mejor predictor y con este se particiona el grupo inicial en varios subgrupos definidos para las categorías del predictor elegido, procediendo de la misma manera en cada subgrupo hasta que no sea posible subdividir ninguno de los grupos que se van obteniendo (Morgan & Sonquist, 1963).

La propuesta más conocida es el CHAID (Chi-square Automatic Interaction Detection) de Kass (1980), el cual utiliza variables categóricas que trabajan con tablas marginales y el test  $\chi^2$  en distintas etapas del proceso, segmentando las categorías según el mejor predictor. El procedimiento CHAID es considerado como un algoritmo general de segmentación (Dorado, 1998). Al igual que otras prácticas de segmentación, las operaciones elementales que realiza son la agrupación de las categorías de las variables pronosticadoras, seguida de la comparación de los efectos entre distintas variables y finalizando

el proceso de segmentación (Escobar, 1998). El proceso como se presenta en la figura 10 puede dividirse en cuatro etapas, la primera de agrupación de categorías, seguida de la selección del mejor predictor, continuando con la segmentación, repitiendo el proceso hasta llegar a un grupo terminal.

Figura 10. Diagrama de flujo correspondiente al algoritmo de Kass (1980).



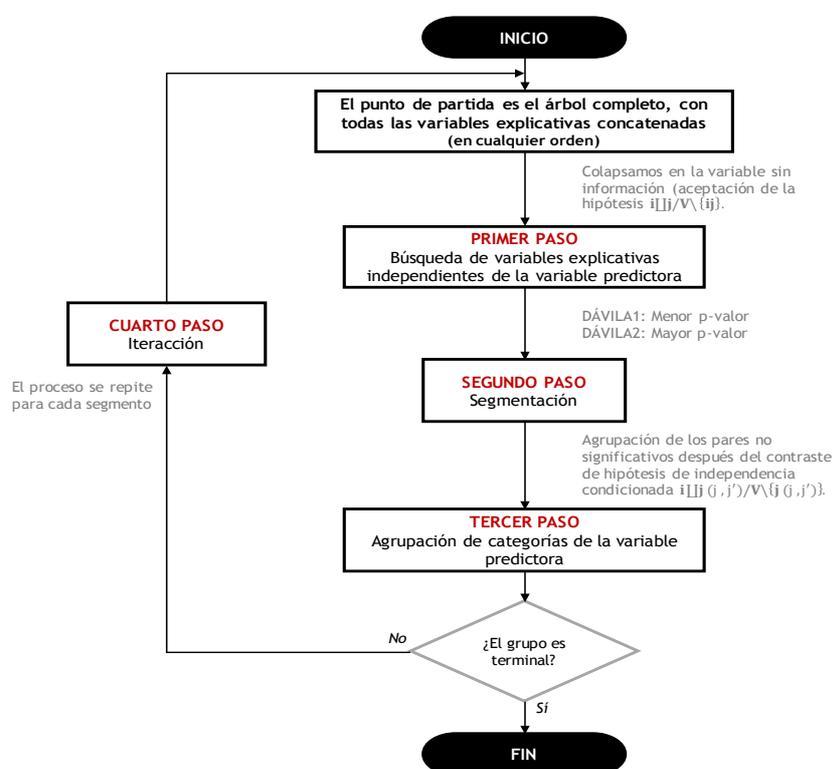
Fuente: Una alternativa al análisis de segmentación basada en el análisis de hipótesis de independencia condicionada (Ávila, 1997).

La utilidad del análisis de segmentación por el método CHAID es múltiple y aunque ha sido diseñada con propósitos descriptivos con cierta cautela puede ser útil para un previo análisis causal de las variables (Escobar, 1998). El algoritmo CHAID ha tenido numerosas aplicaciones desde su introducción en distintos campos de la ciencia (Castro-López, Vicente-Galindo, Patino-Alonso & Galindo-Villardón, 2016). Sin embargo, entre las limitaciones del CHAID se encuentra que el test  $\chi^2$  no captura el carácter asimétrico, sólo considera

información nominal, colapsa sin garantías<sup>17</sup> y considera una sola respuesta por lo que tiene alto riesgo tipo 1; además no recoge información ordinal y no considera respuestas multivariantes (Castro, 2005).

Buscando una solución a los problemas detectados en el algoritmo CHAID, Ávila (1997) propone dos algoritmos de hipótesis de independencia condicionada con la forma  $i \perp\!\!\!\perp j / V \setminus \{ij\}$ , en la que es posible colapsar en  $j$  a los que denominó BSA-I (mejor conocido como DÁVILA 1) y BSA-II (también llamado DÁVILA 2). Los algoritmos que propone Ávila (1997) se pueden resumir en el esquema de la figura 11.

Figura 11. Esquema de los algoritmos BSA-I y BSA-II



Fuente: Elaboración propia con base en Ávila (1997).

<sup>17</sup> Ávila (1997) demostró como el CHAID al trabajar utilizando contrastes sobre distribuciones marginales se puede estar en presencia de variables de confusión las cuales pueden ser independiente marginalmente y no serlo en presencia de otras variables. A este fenómeno se le conoce como paradoja de Simpson (1951).

Basados en los criterios de entropía Dorado (1998) propone utilizar el índice de Shannon (1948) quienes definen entropía o incertidumbre media enfrentada como la ganancia media de información. Utilizando este índice Dorado (1998) propone dos algoritmos, uno ascendente (ADORADO) y otro descendente (DDORADO).

El ADORADO es un algoritmo divisivo que se lleva a cabo en dos fases, en la primera valora el desorden, es decir el grado de heterogeneidad existente en el conjunto de partida a partir de calcular la entropía, a través de los cuales pretende encontrar una segmentación de la población considerada que proporcione los segmentos más homogéneos posibles, en relación a la respuesta; repitiendo el proceso hasta que no haya ninguna variable nueva que produzca un descenso de la entropía significativo (Dorado, 1998).

Recogiendo las ventajas de los algoritmos DÁVILA2 y ADORADO, Dorado (1998) propone DDORADO, el cual parte del árbol completo con todas las variables concatenadas, sin que el orden en que se haga tenga el mayor interés, después se buscan las variables independientes de la respuesta tratando de saber si es posible eliminar las que no aporten información al comportamiento de la variable respuesta. Estas variables serán aquellas que no proporcionen un incremento significativo de la entropía, donde todas las que verifiquen esa condición se eliminan del estudio y el árbol pierde automáticamente las ramas relacionadas con esa(s) variable(s).

Basado en un estudio riguroso de los métodos previos, Castro (2005) propone el algoritmo TAID, el cual otorgan ventajas a los algoritmos planeados anteriormente, principalmente la posibilidad de considerar aspectos multivariantes y ordinales de la variable respuesta.

### V.4.2 Algoritmo TAID

El objetivo del algoritmo TAID, planteado por Castro (2005) es proponer un algoritmo de segmentación que considere la colapsabilidad de categorías que poco aportan información relevante o bien informan de manera similar; que tengan en cuenta el carácter asimétrico de las variables; que consideren el aspecto ordinal de las categorías de la variable respuesta; y que consideren aspectos multivariantes en la variable respuesta.

“Nuestra propuesta se centra en considerar una variable latente, como variable respuesta y varias variables categóricas como predictoras y construir un árbol de clasificación exploratorio que enfatice los predictores más significativos en cada nivel de segmentación” (Castro, 2005, p. 237). De esta forma comienza el algoritmo con la construcción de una variable latente capaz de recoger el carácter multivariante del conjunto de variables con las que se está trabajando.

La búsqueda de las variables latentes es diferente según la naturaleza de los predictores de la respuesta; Castro (2005) considera la solución a partir de los Modelos de Clases Latentes (MCL) a través del modelo mixtura para agrupamiento, el cual considera que cada clase latente representa una clase subyacente a la cual pertenecen los casos o individuos agrupados, considerando siempre el supuesto de independencia local y que las clases latentes son internamente homogéneas. Para los propósitos que se buscan lograr en esta investigación no se considera iniciar el análisis por medio de árboles de decisión con una respuesta de carácter multivariante, sin embargo, si el lector desea ampliar la información al respecto y aplicar esta metodología se recomienda además de los trabajos de Castro-López et al. (2016); Castro (2005) y Ramos (2015) considerar aquellos donde se desarrollan a detalle como Sepúlveda (2004) y Vermunt & Magidson (2005).

### V.4.2.1 Búsqueda del mejor predictor

De acuerdo con Castro (2005), la elección del mejor predictor se extiende de los algoritmos CHAID (Kass, 1980) pero en lugar del test  $\chi^2$  utiliza el coeficiente de predictividad  $\tau$  (Goodman & Kruskal, 1954), definido para medir el incremento relativo de la probabilidad de predecir correctamente la variable fila, conociendo el nivel de la variable columna. Este mismo coeficiente fue utilizado por Light & Margolin (1971) para analizar la heterogeneidad (variabilidad) de los datos categóricos.

El índice  $\tau$  está definido por:

$$\tau_{i/j} = \frac{\left( \sum_{i=1}^I \sum_{j=1}^J \frac{f_{ij}^2}{f_{i.} f_{.j}} - \sum_{i=1}^I (f_{i.}/f_{..})^2 \right)}{1 - \sum_{i=1}^I (f_{i.}/f_{..})^2} = \frac{\Phi^2}{1 - \sum_{i=1}^I (f_{i.}/f_{..})^2}$$

en donde el denominador de esta expresión corresponde a la medida de heterogeneidad total de las categorías de la variable respuesta en el sentido del coeficiente de heterogeneidad de Gini (1912); y el numerador a la heterogeneidad explicada (parte de la total), es la heterogeneidad provocada por el poder predictivo de las categorías del predictor (Castro, 2005). El resultado del índice  $\tau$  varía entre 0 (sin poder predictivo) y 1 (predicción perfecta).

De acuerdo con Castro (2005) se procederá a realizar la primera segmentación siempre y cuando alguno de los índices de predictividad sean significativos a través del índice CATANOVA, introducido por Light & Margolin (1971) y utilizado en árboles ternarios por Siciliano & Mola (1998), el cual se calcula a través de:

$$C = (I - 1)(J - 1) \tau_{i/j}$$

donde  $I$  y  $J$  representan, respectivamente, el número de categorías de la respuesta y el predictor. Dado que sigue una distribución  $\chi^2$  con  $(I - 1)(J - 1)$  grados de libertad, si el correspondiente p-valor es menor de 0.05 se puede afirmar que la capacidad predictiva de la variable explicativa es significativa.

“De entre todos los predictores significativos, elegiremos como mejor predictor, para realizar la primera segmentación, el que presente un mayor coeficiente de predictividad  $\tau$ ” (Castro, 2005, p. 40).

#### ***V.4.2.2 Criterio para segmentar***

En el estudio de una tabla de contingencia no siempre se está frente a la situación en que la variable fila y la variable columna tengan un rol simétrico, es decir, es posible que una de las variables esté condicionando el comportamiento de la otra. Cuando esto sucede, se debe considerar la alternativa de analizar la tabla de contingencia con un método que contemple datos asimétricos (Castro, 2005).

De acuerdo con las distribuciones condicionales, el numerador del índice de predictividad  $\tau$  puede ser descompuesto a través de los ejes principales tal como ocurre con el coeficiente de asociación  $\Phi^2$  en el Análisis Factorial de Correspondencias Simétrico (Lauro & D’Ambra, 1984). En el caso de análisis, siguiendo a Castro (2005) es posible visualizar la influencia de las  $j$  categorías de la variable columna sobre las categorías de la variable respuesta; es decir sobre las distribuciones condicionales empíricas  $p_{ij}/p_{.j}$  con  $i = 1, \dots, I$ , relativas a la hipótesis de ausencia de independencia dada por las frecuencias marginales  $p_{i.}$ , utilizando el Análisis de Correspondencias No Simétrico (ACNS) propuesto por Lauro & D’Ambra (1984).

“Una vez puesto de manifiesto que la inercia recogida en el índice de predictividad  $\tau$  puede ser representada en el plano factorial del Análisis de correspondencias asimétrico se pueden utilizar las coordenadas de las diferentes categorías, en la representación factorial para asociar las categorías del predictor en tres grupos siguiendo las ideas de (Siciliano & Mola, 1998), es decir agruparemos todas las categorías del predictor en fuertes y débiles” (Castro, 2005, p. 241).

Siciliano & Mola (1998) definen las categorías fuertes y débiles de la siguiente forma:

$$\begin{aligned} \varphi_{j1} \geq 1 & \quad \text{Categorías fuertes por la derecha.} \\ |\varphi_{j1}| < 1 & \quad \text{Categorías débiles.} \\ \varphi_{j1} \leq -1 & \quad \text{Categorías fuertes por la izquierda.} \end{aligned}$$

donde  $\varphi_{j1}$  representa la correspondiente coordenada sobre el primer eje factorial del ACNS conformado con la variable respuesta y el correspondiente predictor.

Obtener segmentos homogéneos formados a partir de la predictividad de la categoría es una de las mayores ventajas que se tienen al considerar esta propuesta, además permite emplear variables manifiestas de diferentes escalas e introducir variables exógenas al modelo (Castro, 2005).

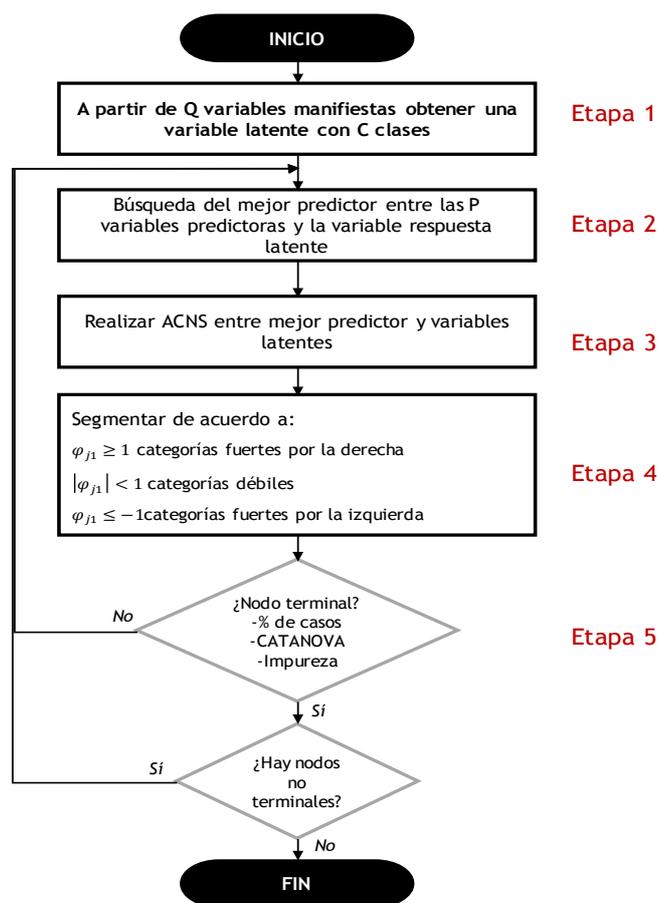
#### ***V.4.2.3 Etapas del algoritmo***

El algoritmo propuesto por Castro (2005) consta de cinco etapas como puede verse en la figura 12, donde considera el análisis de clases latentes para la primera; el coeficiente de predictividad (Goodman & Kruskal, 1954) y el índice de Catanova (Light & Margolin, 1971) para la segunda; y toma en cuenta el análisis de correspondencias no simétrico de Siciliano & Mola (1998) para segmentar los casos continuando así hasta llegar a los grupos terminales. Tomando como filtros de parada un  $p$  – valor  $> 0.05$  para la significación de predictividad, o tamaños de muestra en el segmento menores del 10% del total<sup>18</sup>.

---

<sup>18</sup> Este criterio es arbitrario y puede ser sustituido por otro o por otra regla como el número de sujetos en el segmento antes de proceder a segmentar o número de sujetos en los segmentos (Castro, 2005).

Figura 12. Algoritmo TAID



Fuente: Contribuciones a la detección de variables relevantes en tablas de contingencia multivariantes (Castro, 2005).

Dadas las ventajas propuestas y el tipo de datos con el que se está trabajando, se analizará a los jóvenes preuniversitarios colimenses siguiendo esta metodología.

Previo al paso a los resultados obtenidos, en la tabla 17 se presenta un resumen de las distintas técnicas aplicadas en esta investigación. Los objetivos que busca cada una, el método más conocido, sus limitaciones, el método aplicado y sus ventajas.

**Tabla 17.** Resumen de los métodos aplicados en esta investigación

	Sobremuestreo	Reducción de dimensión	Clúster	Árbol de Segmentación
<b>Objetivo</b>	Aumentar el número de muestras de la clase minoritaria para reducir el grado de distribución desigual (Zhao et al., 2016).	Busca la construcción de aproximaciones matriciales de rango bajo que describa la información de un conjunto de variables observables mediante un conjunto de variables más pequeño (Mahoney & Drineas, 2009).	La búsqueda de grupos en un conjunto de individuos (Fernández, 1991). Clasificar los elementos de un conjunto en subconjuntos homogéneos, siguiendo un determinado criterio de clasificación donde cada elemento pertenece a un único subconjunto (Cuadras, 2014).	Es una técnica que pretende, a partir de la información que suministran ciertas variables clasificar un conjunto de individuos y observaciones, en grupos capaces de describir de la mejor manera posible la variable dependiente (Dorado, 1998).
<b>Método similar más conocido</b>	Métodos de muestreo estratificado.	Descomposición en Valores Singulares (Eckart & Young, 1936) y el procedimiento más utilizado es el Análisis de Componentes Principales (Hotelling, 1933; Pearson, 1901).	Depende del tipo de datos con el que se esté trabajando. Están los métodos jerárquicos y no jerárquicos.	CHAID (Kass, 1980)
<b>Limitaciones</b>	Sólo toman en cuenta una proporción de los individuos del grupo que se desea analizar (clase minoritaria), al ser pequeña la población y considerar menos cantidad para el análisis se pierde información que puede ser crucial para el estudio.	Los vectores singulares son abstracciones matemáticas que en algunos casos pueden interpretarse en términos de las direcciones de los ejes del elipsoide. Sin embargo, es problemático cuando se está interesado en extraer información de los vectores experimentarles reales.	<ul style="list-style-type: none"> <li>• La mayoría están limitados a manejar conjuntos de datos que contienen atributos continuos o categóricos.</li> <li>• La elección de una medida de distancia inadecuada puede sesgar el tratamiento y por tanto los resultados.</li> <li>• Son efectivos y precisos en un pequeño conjunto de datos.</li> </ul>	<ul style="list-style-type: none"> <li>• El test <math>\chi^2</math> no captura el carácter asimétrico de las variables.</li> <li>• Sólo considera información nominal.</li> <li>• Colapsa sin garantías.</li> <li>• Considera una sola respuesta por lo que tiene alto riesgo tipo 1.</li> <li>• No recoge información ordinal.</li> <li>• No considera respuestas multivariantes.</li> </ul>
<b>Método utilizado</b>	Sobremuestreo estratificado (Siddiqi, 2006)	CUR	TwoStep Cluster (Chiu et al., 2001)	TAID (Castro, 2005)
<b>Ventajas del método utilizado sobre métodos similares</b>	Maximizar al grupo minoritario con el fin de capturar los distintos perfiles de comportamiento y eliminar los posibles sesgos que se producen al momento de extraer los datos en un conjunto (Siddiqi, 2006).	Número pequeño de número de las columnas reales y un número pequeño de las filas reales construidas de tal manera que $CUR \approx X$	<ul style="list-style-type: none"> <li>• Se puede trabajar con variables categóricas o continuas.</li> <li>• Proporciona la capacidad de encontrar automáticamente el número óptimo de clústeres.</li> <li>• Permite trabajar con grandes conjuntos de datos.</li> </ul>	<ul style="list-style-type: none"> <li>• Considera la colapsabilidad de categorías que poco aporta información relevante o bien informan de manera similar</li> <li>• Tiene en cuenta el carácter asimétrico de las variables.</li> <li>• Considera el aspecto ordinal de las categorías de la variable respuesta.</li> <li>• Considera aspectos multivariantes en la variable respuesta.</li> </ul>

---

**VI**

**RESULTADOS**

---

## VI.1 Descripción general de la población

Cuando las variables analizadas en una investigación son de tipo cuantitativo (en este caso constructos), la primera presentación y el análisis de los resultados generalmente se lleva a cabo a través de una comparación de medidas de tendencia central y dispersión de los resultados. Durante mucho tiempo, el análisis de datos categóricos se ha limitado al análisis de tablas de contingencia bidimensionales, que si bien son una referencia para iniciar la exploración de los datos, limitan a un mayor conocimiento acerca de la población de estudio (Castro, 2005).

Siguiendo esta línea, para comenzar el análisis del contexto de los estudiantes preuniversitarios se realizaron tablas de frecuencias (con valores absolutos y relativos) con los resultados a los instrumentos aplicados, donde se separan a los jóvenes según los puntajes que obtuvieron en el DUKE-UNC-11. Así, aquellos adolescentes que alcanzaron un valor igual o superior al punto de corte ( $\geq 33$ ) son considerados con Apoyo Social Funcional (ASF) y quienes obtuvieron un puntaje menor, son clasificados con Escaso Apoyo Social (EAS) en los cuales se enfoca esta investigación.

El análisis de tablas de contingencia puede plantearse a partir de ciertas medidas de asociación que indiquen si la relación entre las variables de interés es o no estadísticamente significativa, por lo que se anexan al análisis los p-valores obtenidos en el test Chi-Cuadrado<sup>19</sup> ( $\chi^2$ ) (Pearson, 1900), presentado como uno de los más utilizados para realizar los análisis comparativos entre dos variables cualitativas.

---

<sup>19</sup> El test  $\chi^2$  contrasta dos hipótesis estadísticas, una nula o hipótesis de independencia de las variables  $H_0: i \perp j$  y una alternativa o hipótesis de asociación de las variables ( $H_1: i \not\perp j$ ), a través de los resultados observados y los resultados teóricos calculados bajo el supuesto de que las variables fueran independientes. De tal forma que, si los resultados observados difieren significativamente de los teóricos, es posible rechazar la independencia de los mismos y considerar que las variables se encuentran asociadas.

“Es comúnmente aceptado por la comunidad científica que, en general, una presentación gráfica proporciona mayor información acerca de las características y patrones de los datos, que un texto o una presentación tabular de los mismos” (Avila, 1993, p. 19). Así, al analizar las respuestas otorgadas a los instrumentos aplicados según las categorías consideradas (sociodemográficas, estilos de vida y percepción del entorno) se presentan, junto a las tablas, los gráficos de barras como un método básico usual para mostrar los resultados, las cuales funcionan como un apoyo visual a los investigadores.

De los 1267 jóvenes analizados 195 (15.4%) son considerados adolescentes con EAS. De estos jóvenes 101 (51.8%) son del sexo masculino y 94 (48.2%) del sexo femenino; con una media de edad 16.55 de ( $\pm 0.8$ ); donde la mayoría pertenecen a la clase obrera (67.1%) y clase intermedia (20.9%) (ver tabla 18).

**Tabla 18.** Resultados de las variables sociodemográficas

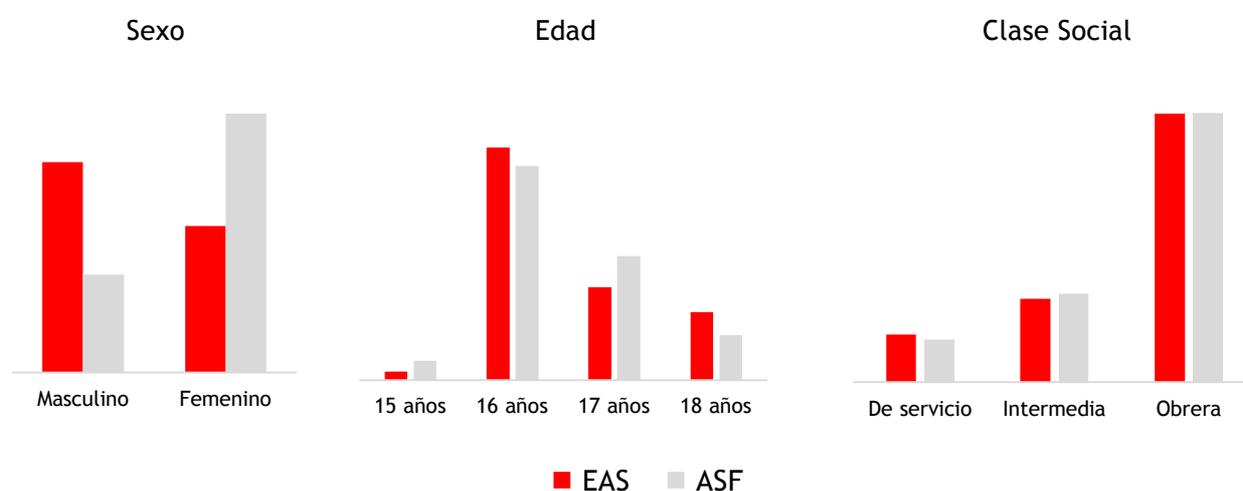
	EAS		ASF	
	n	(%)	n	(%)
N	195	(15.4)	1072	(84.6)
<b>Sexo</b>				
Masculino	101	(51.8)	488	(45.5)
Femenino	94	(48.2)	584	(54.5)
p-valor	0.106			
<b>Edad</b>				
15 años	4	(2.1)	51	(4.8)
16 años	113	(57.9)	571	(53.3)
17 años	45	(23.1)	331	(30.9)
18 años	33	(16.9)	119	(11.1)
p-valor	0.009			
<b>Clase Social</b>				
De servicio	19	(12.0)	94	(10.6)
Intermedia	33	(20.9)	196	(22.1)
Obrera	106	(67.1)	597	(67.3)
p-valor	0.842			

Fuente: Elaboración propia.

Considerando los resultados del estadístico  $\chi^2$  aceptando un nivel de significancia ( $\alpha$ ) del 5% (confianza del 95%) es posible afirmar, dados sus p-valores no significativos<sup>20</sup> que la clase social y el sexo de los jóvenes es independiente con el AS; sin embargo, con un p-valor de 0.009 se puede considerar que este depende de la edad.

En la figura 13 se presentan el conjunto de gráficas de barras resultantes de las cifras mostradas en la tabla 18 como una alternativa visual a la presentación de los datos.

**Figura 13.** Conjunto de gráficas de las variables sociodemográficas



Los resultados de los constructos que reflejan los estilos de vida se muestran en la tabla 19, donde se puede apreciar cómo el 37% de los jóvenes con EAS fueron detectados con sobrepeso u obesidad; 42.6% tienen un patrón deficiente de alimentación; 8.7% presentan niveles bajos de actividad física; 13.8% registraron adicción al tabaco (leve, moderada o intensa) y el 11.3% tienen un consumo perjudicial o dependencia a las bebidas alcohólicas.

<sup>20</sup> Si se acepta la hipótesis nula al 5 y al 1% ( $p\text{-valor} \geq 0.05$ ), se dice que el contraste es “no significativo”. En cambio, si se rechaza la hipótesis nula ( $p\text{-valor} < 0.05$ ) se dice que el contraste es “significativo”, donde si este se da tanto al 5% como al 1% se trata de un valor “altamente significativo”.

Tabla 19. Resultados de los cuestionarios sobre estilos de vida

	EAS		ASF	
	n	(%)	n	(%)
<b>Índice de Masa Corporal</b>				
Normopeso	123	(63.1)	684	(63.8)
Sobrepeso	36	(18.5)	233	(21.7)
Obesidad	36	(18.5)	155	(14.5)
p-valor	0.273			
<b>Patrón de alimentación</b>				
Dieta óptima	112	(57.4)	775	(72.3)
Debe mejorar	76	(39.0)	275	(25.7)
De muy baja calidad	7	(3.6)	22	(2.1)
p-valor	0.000			
<b>Actividad física</b>				
Baja	17	(8.7)	71	(6.6)
Moderada	89	(45.6)	392	(36.6)
Vigorosa	89	(45.6)	609	(56.8)
p-valor	0.015			
<b>Adicción al tabaco</b>				
No fuma	155	(79.5)	889	(82.9)
Ex fumador	13	(6.7)	83	(7.7)
Leve	23	(11.8)	88	(8.2)
Moderado	3	(1.5)	9	(8.2)
Intenso	1	(0.5)	3	(0.3)
p-valor	0.400			
<b>Consumo de alcohol</b>				
Bebedor social	155	(79.5)	847	(79.0)
Consumo de riesgo	18	(9.2)	149	(13.9)
Consumo perjudicial	18	(9.2)	63	(5.9)
Dependencia	4	(2.1)	13	(1.2)
p-valor	0.090			

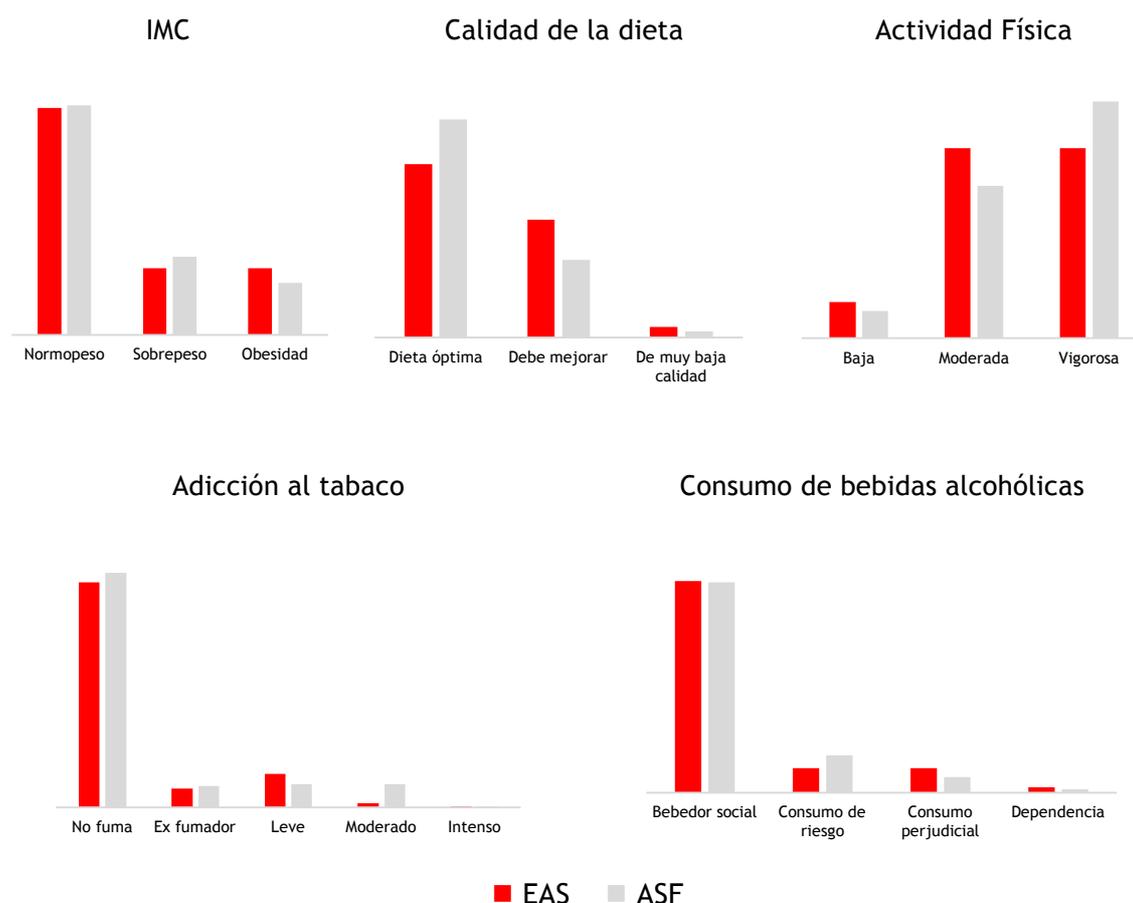
Fuente: Elaboración propia.

Si se observan los resultados del estadístico  $\chi^2$ , aceptando un nivel de significancia ( $\alpha$ ) del 5% es posible afirmar, dados sus p-valores no significativos, que el IMC de los jóvenes y su adicción al tabaco y alcohol es independiente de

su AS; sin embargo, sí se detectó relación entre el AS y el patrón de alimentación y actividad física, dados los p-valores altamente significativos.

Al igual que las variables anteriores, la tabla 19 viene acompañada del conjunto de gráficas de barras presentes en la figura 14 para una presentación alternativa de los resultados.

**Figura 14.** Conjunto de gráficas de los constructos sobre estilos de vida



Si siguiendo con la estructura con la que se presentan los resultados anteriores, en la tabla 20 se muestran los referentes a los cuestionarios que miden la percepción del entorno que tienen los jóvenes. De esta forma se puede resaltar que 75.3% de los estudiantes con EAS consideran tener un ambiente familiar con disfunción leve o grave; 16.9% se sienten insatisfechos con su vida; 44.1% tienen una baja CVRS; se detecta que un 70.8% tiene algún tipo de psicopatología; 9.7% tienen una marcada preocupación por su figura y finalmente, el 3.1% de estos adolescentes se sienten moderadamente o severamente insatisfechos con su imagen corporal.

Tabla 20. Resultado de los cuestionarios que miden la percepción del entorno

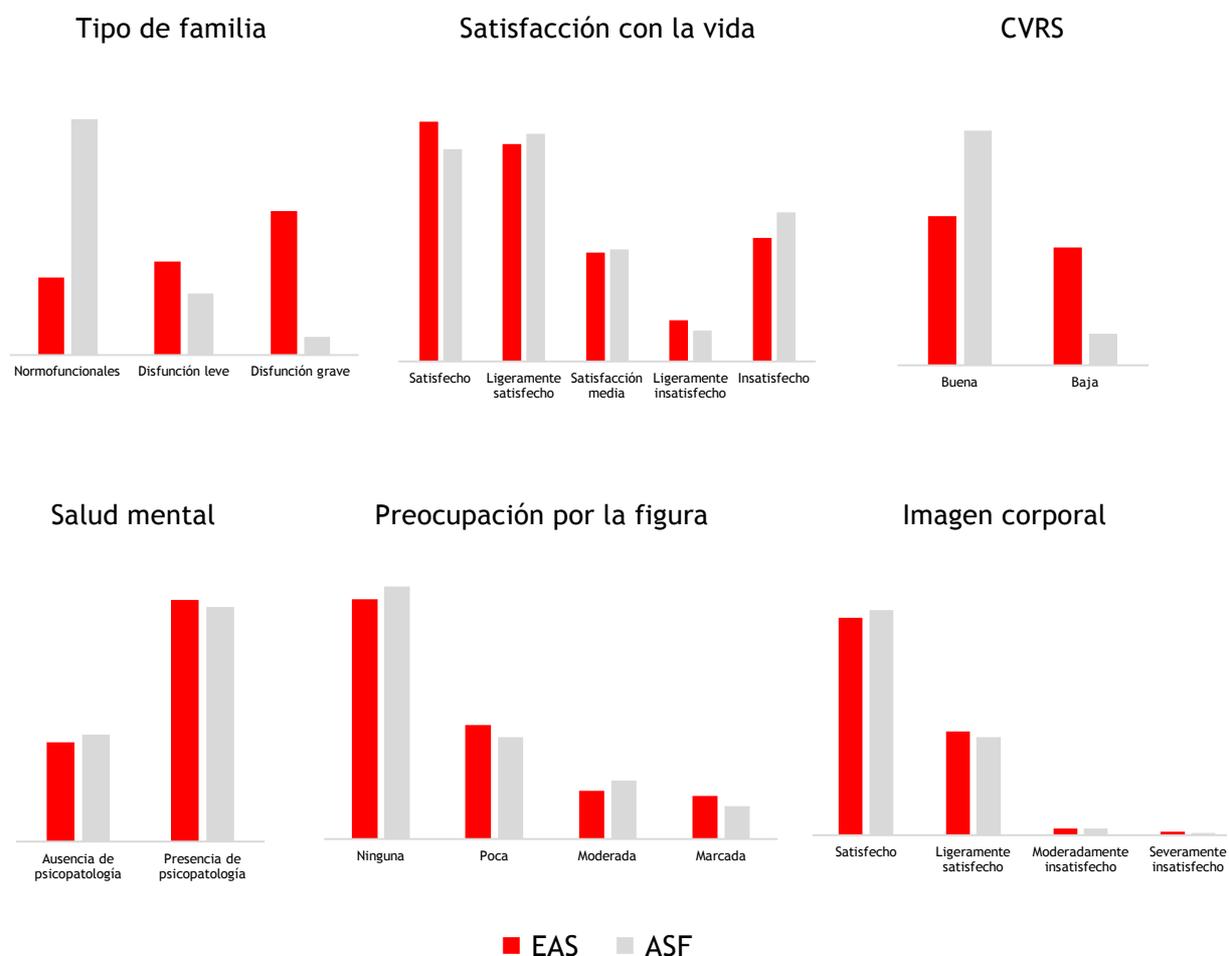
	EAS		ASF	
	n	(%)	n	(%)
<b>Tipo de familia</b>				
Normofuncionales	48	(24.6)	802	(74.8)
Disfunción leve	58	(29.7)	209	(19.5)
Disfunción grave	89	(45.6)	61	(5.7)
p-valor		0.000		
<b>Satisfacción con la vida</b>				
Satisfecho	64	(32.8)	311	(29.0)
Ligeramente satisfecho	58	(29.7)	333	(31.1)
Satisfacción media	29	(14.9)	164	(15.3)
Ligeramente insatisfecho	11	(5.6)	45	(4.2)
Insatisfecho	33	(16.9)	219	(20.4)
p-valor		0.606		
<b>Calidad de vida</b>				
Buena	109	(55.9)	946	(88.2)
Baja	86	(44.1)	126	(11.8)
p-valor		0.000		
<b>Salud mental</b>				
Ausencia de psicopatología	57	(29.2)	335	(31.3)
Presencia de psicopatología	138	(70.8)	737	(68.8)
p-valor		0.575		
<b>Preocupación por la figura</b>				
Ninguna	105	(53.8)	608	(56.7)
Poca	50	(25.6)	245	(22.9)
Moderada	21	(10.8)	140	(13.1)
Marcada	19	(9.7)	79	(7.4)
p-valor		0.444		
<b>Imagen corporal</b>				
Satisfecho	128	(65.6)	728	(67.9)
Ligeramente satisfecho	61	(31.3)	317	(29.6)
Moderadamente insatisfecho	4	(2.1)	21	(2.0)
Severamente insatisfecho	2	(1.0)	6	(0.6)
p-valor		0.835		

Fuente: Elaboración propia.

Si se analizan los resultados del test  $\chi^2$ , aceptando un nivel de significancia ( $\alpha$ ) del 5% es posible afirmar, dados los p-valores no significativos,

que el ASF es independiente de la satisfacción con la vida, salud mental, preocupación por la figura e insatisfacción con la imagen corporal; no obstante, se puede considerar relacionado con el tipo de función familiar y la CVRS con p-valores altamente significativos.

**Figura 15.** Conjunto de gráficas de los constructos sobre la percepción del entorno



En la figura 15 se presentan el conjunto de gráficas de barras resultantes de las cifras mostradas en la tabla 20 como una alternativa visual a la presentación de los datos.

### VI.1.1 Exploración de las relaciones entre variables sociodemográficas, hábitos de vida y percepción del entorno

Cuando las variables analizadas son medidas en escala ordinal, suelen utilizarse las asociaciones de Spearman (1904) ( $\rho$ ) como una de las técnicas estadísticas utilizadas con más frecuencia por los investigadores para encontrar relaciones (o interdependencia) entre dos variables cualitativas.

El  $\rho$  de Spearman es una medida de asociación entre dos variables donde para calcularla los datos son ordenados y reemplazado por su respectivo orden. Viene representada por  $\rho = 1 - \frac{6 \sum D^2}{N(N^2-1)}$ , donde D es la diferencia entre los correspondientes estadísticos de orden de x y N es el número de parejas.

Las asociaciones de Spearman de las variables presentes en este estudio se presentan en la tabla 21 y fueron calculadas con el SPSS v.22 (IBM, 2016). De esta forma se presentan las relaciones entre variables y constructos medidas a través de los cuestionarios aplicados.

**Tabla 21.** Asociaciones de Spearman

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
V1	1														
V2	0.05	1													
V3	-0.02	0.07	1												
V4	0.01	0.04	0.01	1											
V5	-0.02	-0.03	-0.06	0.06	1										
V6	0.12	0.06	-0.01	0.03	-0.06	1									
V7	0.08	0.17	0.02	0.04	0	0.07	1								
V8	-0.04	0.07	0.06	0.02	0.01	-0.01	0.01	1							
V9	0.00	0.09	0.02	0.00	0.02	0.07	0.06	0.12	1						
V10	<b>0.43</b>	0.13	0.01	-0.03	-0.01	0.12	0.06	-0.1	-0.03	1					
V11	0.03	-0.01	0.00	-0.01	-0.01	0.00	-0.03	-0.05	-0.09	-0.01	1				
V12	<b>0.31</b>	0.18	0.01	0.03	-0.02	0.13	0.09	-0.03	0.01	<b>0.30</b>	0.00	1			
V13	0.02	-0.01	0.05	0.04	-0.05	0.00	0.03	-0.01	0.07	0.06	-0.06	0.04	1		
V14	-0.02	-0.10	-0.03	0.05	0.06	-0.06	-0.07	0.01	0.02	-0.06	0.05	-0.08	0.00	1	
V15	-0.02	-0.04	0.00	-0.02	-0.06	-0.03	-0.01	0.03	0.00	-0.01	0.06	-0.02	-0.02	<b>0.35</b>	1

**Fuente:** Elaboración propia a través de SPSS v.22 (IBM, 2016).

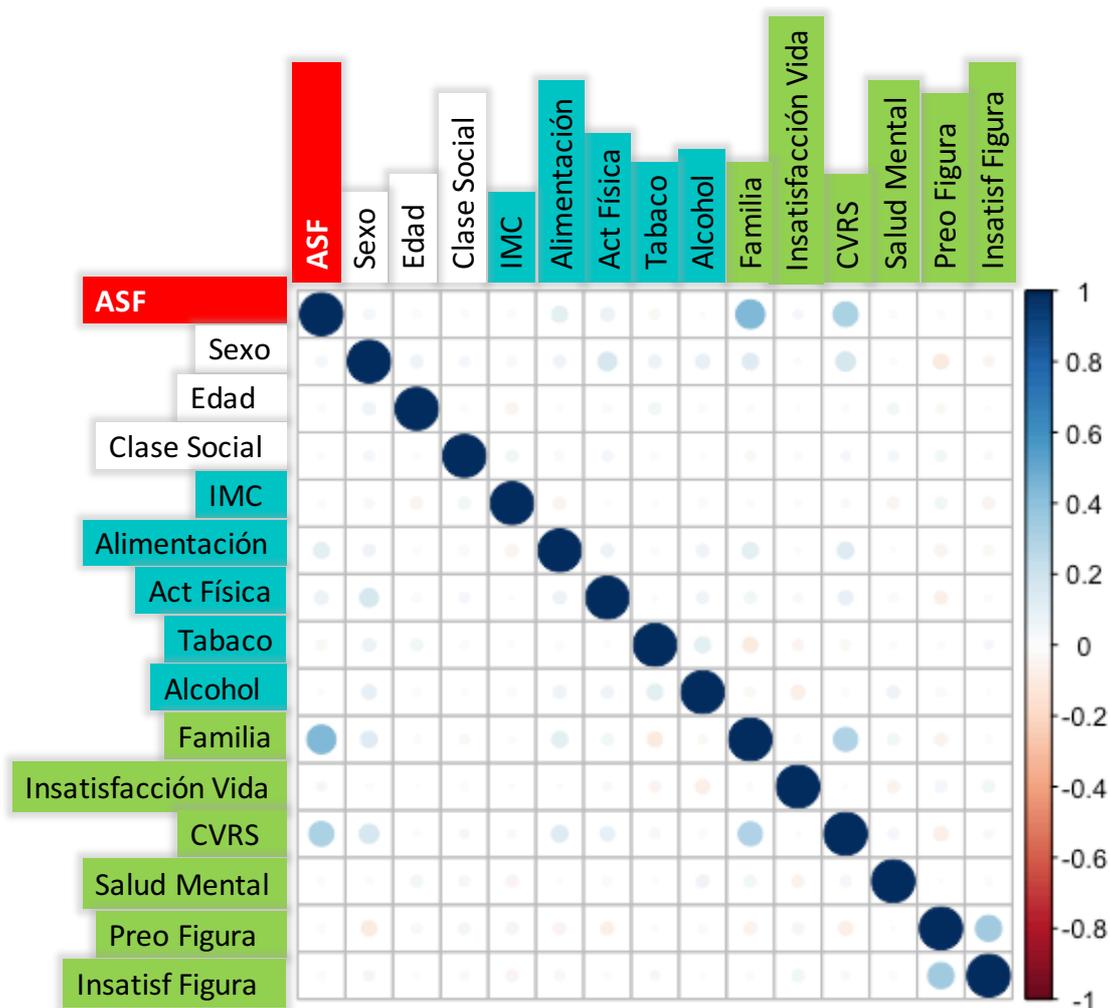
Donde:

V1: Apoyo Social	V9: Consumo de alcohol
V2: Sexo	V10: Función familiar
V3: Edad	V11: Insatisfacción con la vida
V4: Clase Social	V12: CVRS
V5: Índice de Masa Corporal	V13: Salud mental
V6: Calidad de la alimentación	V14: Preocupación con la figura
V7: Actividad física	V15: Insatisfacción con la imagen corporal
V8: Adicción al tabaco	

Prestando atención en el AS (V1) se puede encontrar una asociación media con la función familiar (V10) ( $\rho=0.43$ ) y la CVRS (V12) ( $\rho=0.31$ ). Considerando los otros constructos, el puntaje mayor se encuentra entre la preocupación por la figura (V14) y la insatisfacción con la imagen corporal (V15) ( $\rho=0.35$ ), asociación lógica ya que los dos cuestionarios miden constructos similares; y entre la CVRS (V12) y la función familiar (V10) ( $\rho=0.30$ ).

En la figura 16 se presenta la matriz de asociaciones de Spearman realizado con R (R Development Core Team, 2013), a través de la cual, los resultados de la tabla 21 se presentan de forma gráfica. En esta matriz una asociación perfecta positiva (+1) se encuentra en color azul marino (●) y una asociación perfecta negativa (-1) es representada en rojo intenso (●) como se muestra al lado derecho de la figura.

Figura 16. Matriz de asociaciones de Spearman

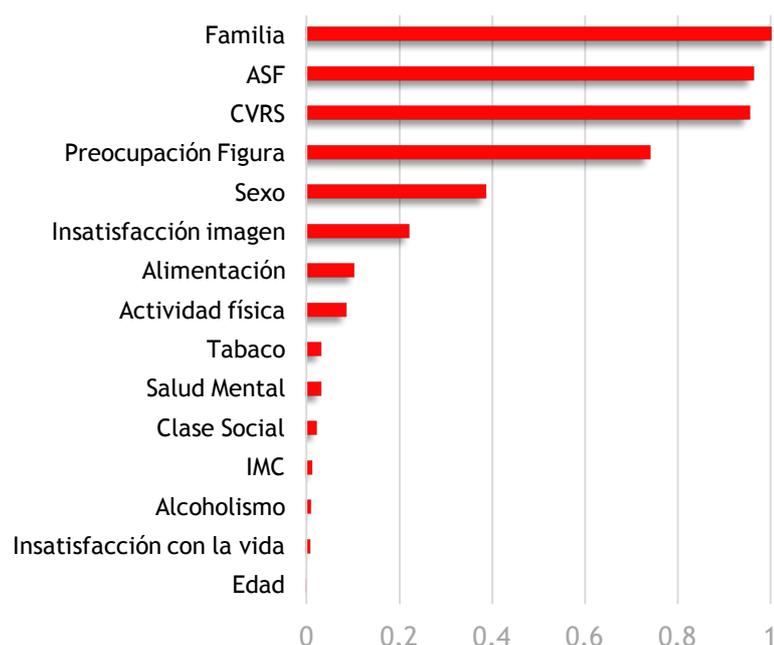


Fuente: Elaboración propia a través de R Project for Statistical Computing (R Development Core Team, 2013).

## VI.2 Inspección de las agrupaciones naturales de los jóvenes según sus variables socioeconómicas, hábitos de vida y la percepción de su entorno

En la búsqueda de particularidades que identifiquen a los jóvenes pertenecientes al grupo de EAS se procedió a realizar un análisis de clúster bietápico utilizando SPSS v.22 (IBM, 2016), donde al analizar toda la población participante en el estudio, el predictor más importante para la clasificación fue la familia<sup>21</sup>, seguido de la clasificación entre ASF y EAS, CVRS, preocupación por la figura, sexo e insatisfacción con la imagen corporal como se muestra en la gráfica 1.

Gráfica 1. Importancia del predictor para la clasificación en el clúster

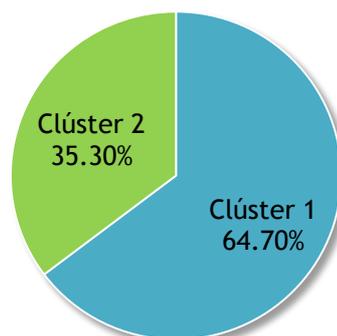


Fuente: Elaboración propia con base en los resultados de la clasificación por medio del SPSS (IBM, 2016).

<sup>21</sup> Es importante señalar que siguiendo la recomendación de Chiu et al. (2001) y del IBM Knowledge Center (2001), de probar el análisis con la selección aleatoria de las variables en repetidas ocasiones, en todas las veces el resultado salió de la forma que se presenta.

Con estos criterios se formaron dos agrupaciones en donde se logra clasificar el 82.5% de la población. El primer clúster lo conforman 676 (64.7%) individuos y el otro 369 (35.3%) (ver gráfica 2), con una silueta de cohesión y separación de 0.1 por lo que se puede considerar una agrupación poco óptima, donde no hay evidencias claras de los agrupamientos.

Gráfica 2. Tamaños de clústeres

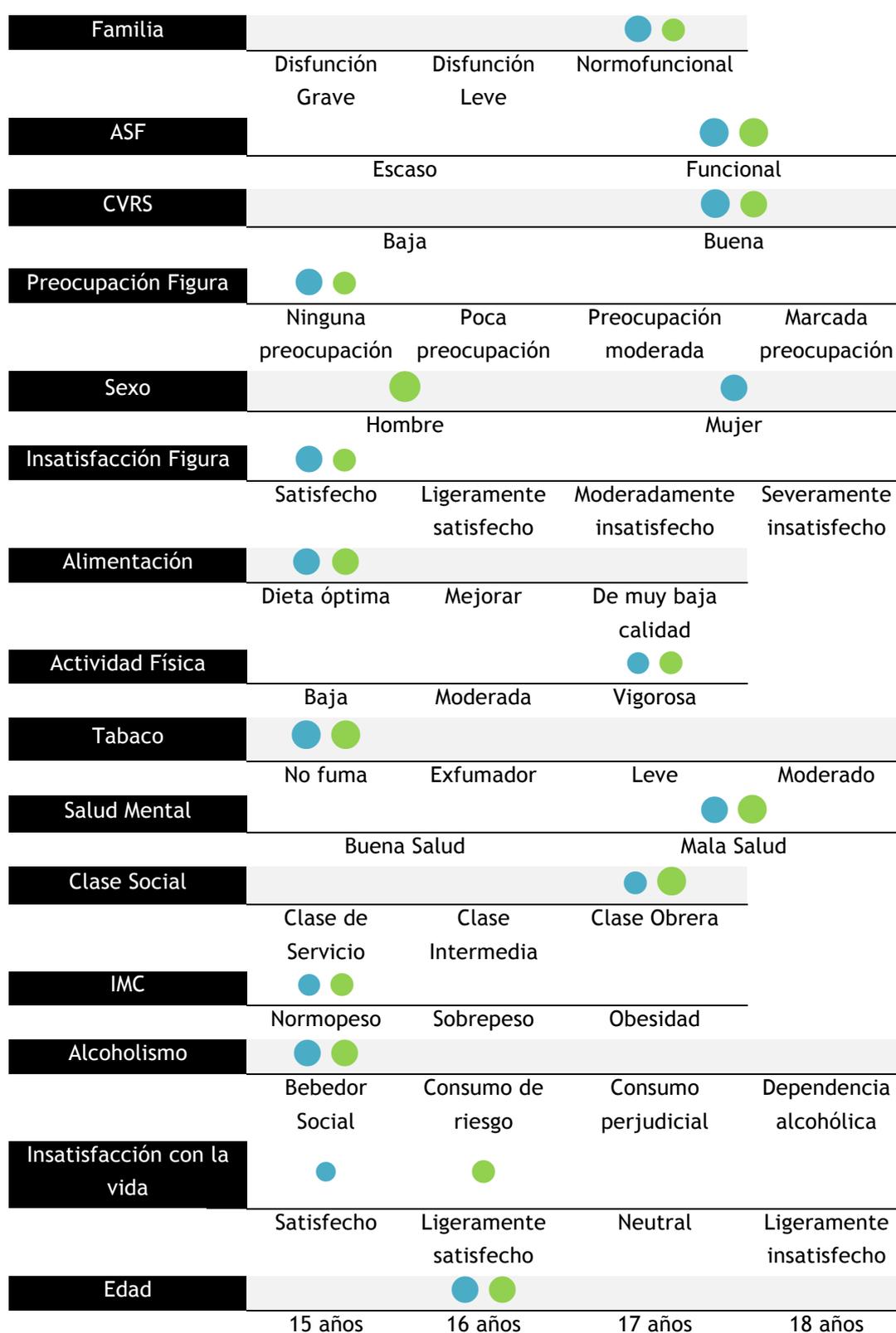


**Fuente:** Elaboración propia con base en los resultados de la clasificación por medio del SPSS (IBM, 2016).

Analizando de forma general las agrupaciones se puede observar en la figura 17, como la mayor parte de los jóvenes de los dos clústeres tiene 16 años, pertenecen a una clase social obrera, su IMC está en grados normales, tiene una dieta óptima, actividad física vigorosa, no fuman, su consumo de bebidas alcohólicas es de tipo social (considerado sin riesgo), pertenecen a familias normofuncionales, tienen una buena CVRS, mala salud mental, poca preocupación por su figura y están satisfechos con su imagen corporal.

Las únicas variables que distinguen a los dos grupos son el sexo y la satisfacción con la vida; en el primer clúster la mayor parte son mujeres y están satisfechas con la vida; y en el segundo la mayoría son hombres y están ligeramente satisfechos, como se puede observar en la figura 17.

Figura 17. Comparación de clústeres



Fuente: Elaboración propia con base en los resultados de la clasificación por medio del SPSS (IBM, 2016).

La clasificación de los preuniversitarios en clústeres a través del método bietápico permite identificar la importancia de la funcionalidad en la familia como constructo que caracteriza el entorno de los jóvenes, sin embargo, no otorga más información para acceder a las diferencias globales entre los adolescentes con EAS y aquellos con ASF. A partir de esta clasificación no se pueden proponer rutas de acción para políticas públicas ni intervenciones específicas, dado que no se cumplen los requisitos de homogeneidad en el segmento y heterogeneidad entre segmentos.

Cabe recordar que el objetivo final del análisis de clúster es segmentar la población de forma que permita tener grupos bien definidos entre sí donde cada elemento pertenezca a un único subconjunto para tomar decisiones basadas en estas agrupaciones. Siguiendo el esquema de la Figura 4, después de haber analizado los datos obtenidos sobre jóvenes preuniversitarios colimenses, a través del método TwoStep lo que sigue es la validación de los resultados obtenidos. Con una silueta de cohesión y separación de 0.1 y con la dificultad práctica de la obtención de conclusiones a través de los resultados se recomienda volver a alguno de los pasos anteriores para mejorar los resultados obtenidos.

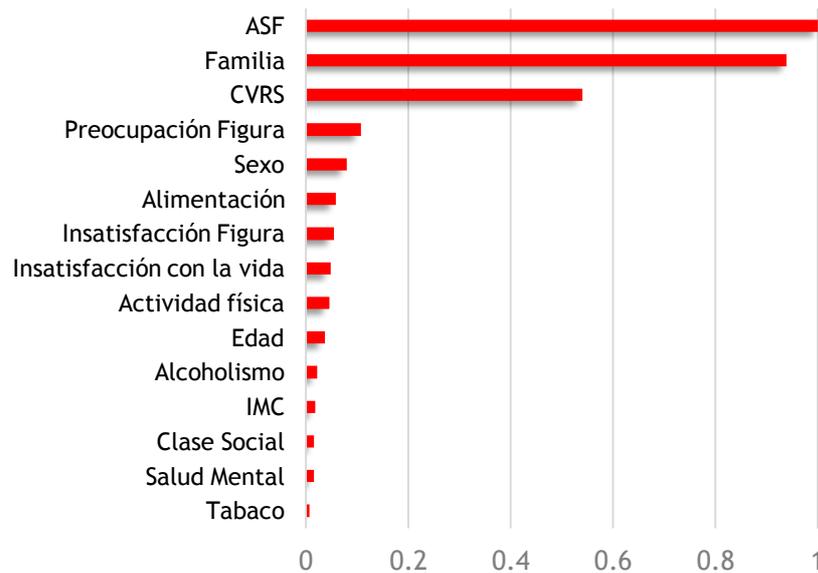
Como pudo observarse en las tablas de distribución de frecuencias y en las gráficas de barras presentadas anteriormente, el comportamiento de los dos grupos resulta ser muy similar en la mayor parte de los constructos y variables medidas, es decir las características específicas de los adolescentes con EAS quedan ocultas al intentar resaltar sus propiedades comparándolos con los jóvenes con ASF, esto probablemente debido a las pocas observaciones en el primer grupo en relación con el segundo, es decir, se está frente al problema de datos desbalanceados y enmascarados. Dadas las características de los individuos analizados no se puede esperar que la separación en grupos sea la mejor, para solucionar este problema, se debe volver al primer paso del proceso para realizar el análisis de clúster, es decir, modificar la matriz de datos de arranque.

### VI.3 Búsqueda de las agrupaciones naturales de los jóvenes a través de la población recalibrada

El análisis de clúster por medio del TwoStep (IBM, 2016) se repite ahora con la población calibrada y recalculada a través del sobremuestreo estratificado (Siddiqi, 2006), de esta forma se busca resaltar la importancia de seleccionar adecuadamente la matriz inicial de datos.

Realizando los análisis a los 591 jóvenes incluidos en la muestra (195 con EAS y 396 con ASF) logran encontrarse, a través del TwoStep dos clúster donde ahora sí el predictor más importante para la clasificación fue el ASF<sup>22</sup>, seguido de la familia y la CVRS como puede observarse en la gráfica 3.

Gráfica 3. Importancia del predictor en los clústeres basados en la muestra

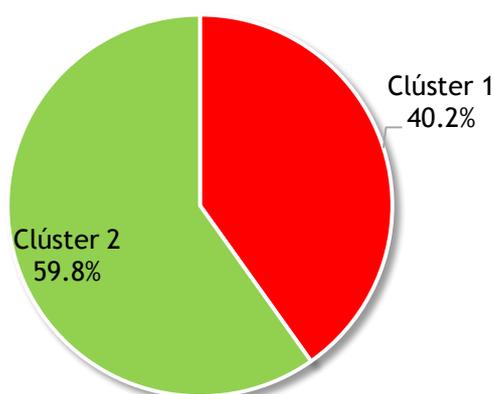


Fuente: Elaboración propia con base en los resultados de la clasificación por medio del SPSS (IBM, 2016).

<sup>22</sup> Al igual que la vez anterior, el proceso fue realizado en varias ocasiones incluyendo las variables de forma aleatoria obteniendo los mismos resultados todas las veces.

Con estos criterios se formaron dos agrupaciones donde se logra clasificar el 81.7% de la muestra. El primer clúster lo conforman 194 (40.2%) individuos y el otro 289 (59.8%), como puede verse en la gráfica 4, con una silueta de cohesión y separación de 0.2 por lo que se puede considerar una agrupación mejor que la anterior pero aun así poco óptima sin evidencias claras de los agrupamientos.

Gráfica 4. Tamaño de los clústeres basados en la muestra

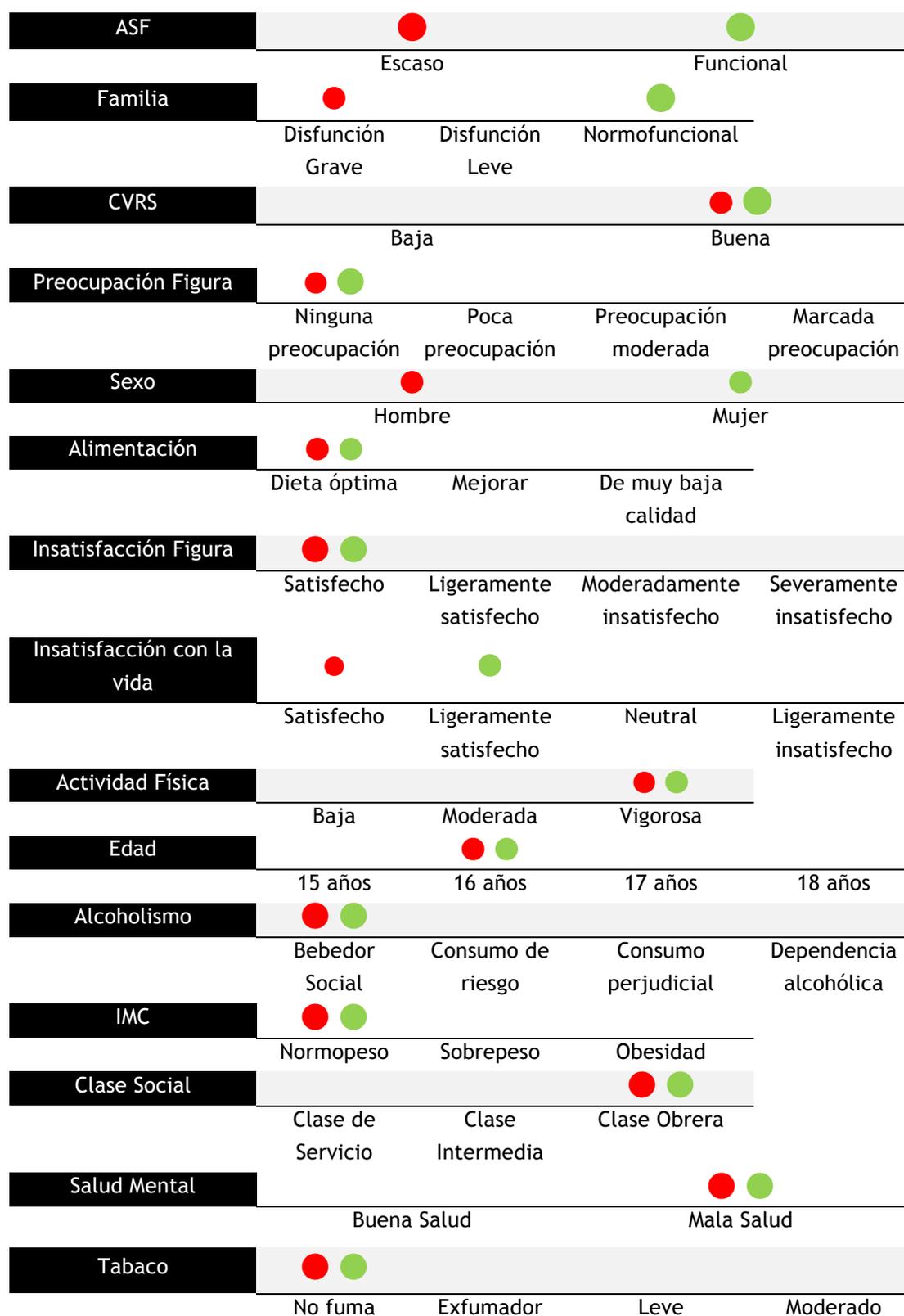


**Fuente:** Elaboración propia con base en los resultados de la clasificación por medio del SPSS (IBM, 2016).

Analizando las agrupaciones desde esta muestra, se pudieron encontrar mayores diferencias en la comparación de los clústeres generados a partir de la población total. Los constructos que diferencian primordialmente a los jóvenes con EAS es la funcionalidad de la familia, ya que la mayor parte de estos jóvenes perciben tener un ambiente familiar con disfunción grave, siendo la mayoría hombres y con una satisfacción con la vida superior al grupo con ASF.

Las variables y constructos relacionados con la edad, clase social, IMC, alimentación, actividad física, consumo de tabaco y alcohol, CVRS, salud mental, preocupación con la figura y satisfacción con la imagen corporal continúan sin mostrar más información para analizar los resultados como se puede ver en la figura 18.

Figura 18. Comparación de clústeres basados en la muestra



Fuente: Elaboración propia con base en los resultados de la clasificación por medio del SPSS (IBM, 2016).

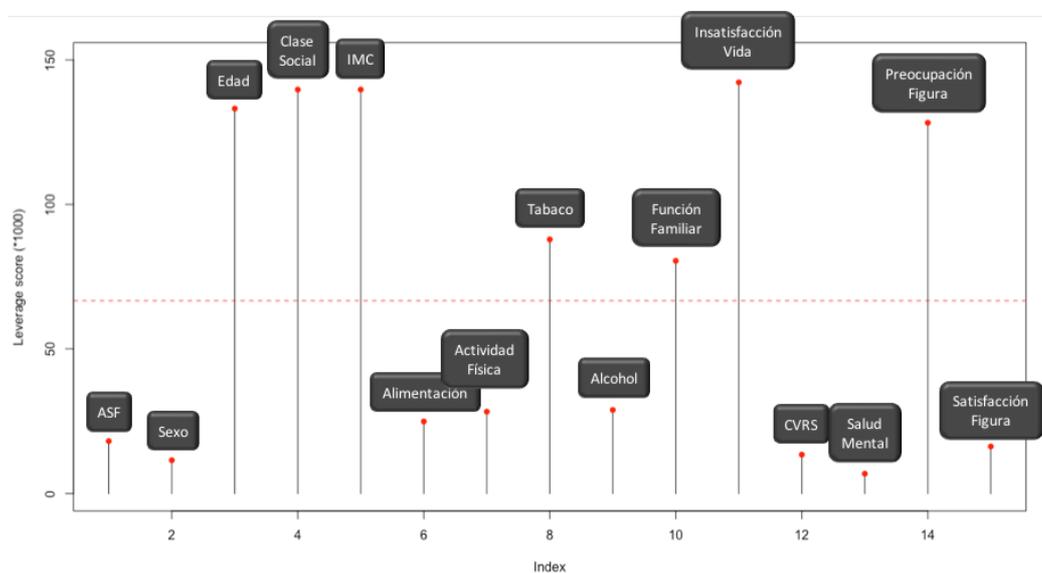
Una vez realizado el sobremuestreo estratificado de la población, logró mejorarse la clasificación de los preuniversitarios en clústeres a través del método bietápico, sin embargo, dadas las características de la población analizada esta ordenación sigue mostrando problemas importantes, ya que la silueta de cohesión sigue sin ser la óptima, por lo que se continúa sin poder presentar información a partir de la cuales se puedan proponer rutas de acción o intervenciones específicas.

## VI.4 Exploración de las variables y constructos que tienen alta influencia en la percepción del Apoyo Social en los jóvenes

Utilizando las ideas desarrolladas en el trabajo de Mahoney & Drineas (2009) y de Bodor et al. (2012) y se procedió al cálculo de los leverages de los constructos y variables a través del paquete estadístico rCUR (Bodor et al., 2012), en la búsqueda de las variables y constructos que tienen alta influencia en la percepción del apoyo social en los jóvenes analizados.

En la figura 19 se presentan los leverages optimizados para las 3 variables y 12 constructos medidos en los 591 jóvenes de la muestra, donde claramente se pueden observar como la edad, clase social, IMC, tabaco, función familiar, insatisfacción con la vida y preocupación por la figura son los aspectos que tienen alta influencia estadística en la ordenación de los adolescentes.

**Figura 19.** Leverage para la selección de variables

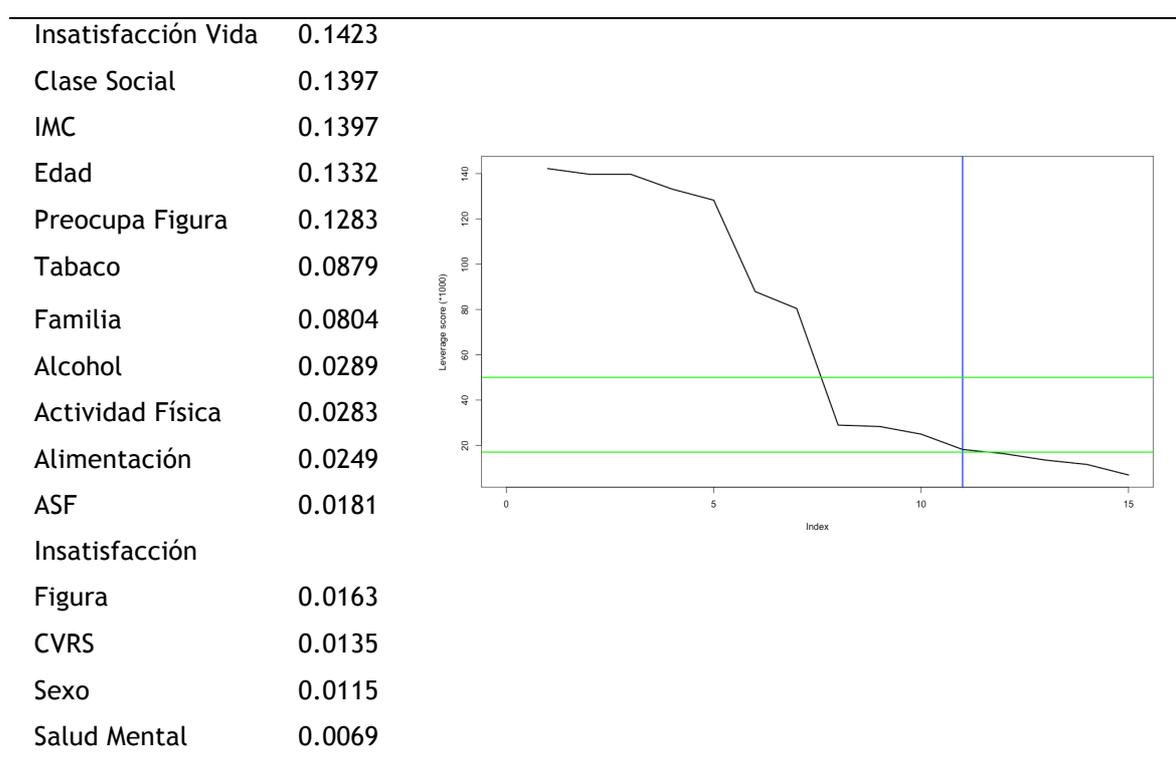


Fuente: Elaboración propia por medio del rCUR (Bodor et al., 2012).

Para la selección de las variables que tienen alta influencia estadística, los leverages pueden ordenarse de forma descendente y similar a la regla del codo utilizada para el Análisis Factorial, seleccionar aquellos hasta donde la

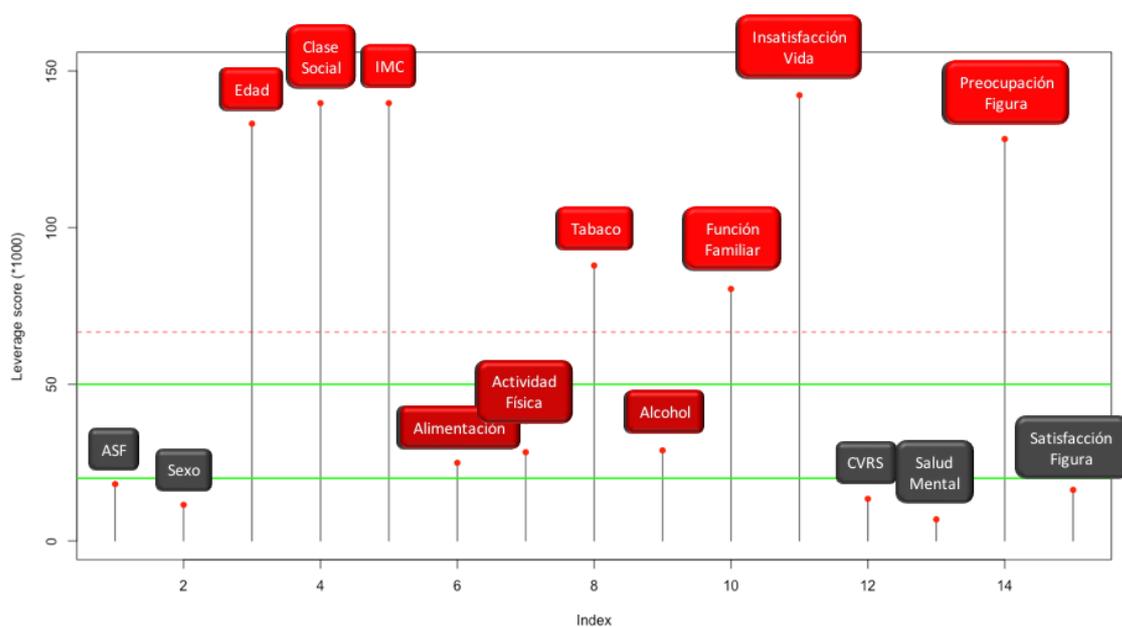
aportación de la variable deje de ser significativa respecto a la anterior. De esta forma, la gráfica 5 muestra los resultados obtenidos y como puede observarse en la tabla que la acompaña, insatisfacción con la vida, clase social, IMC, edad, preocupación por su figura, consumo de tabaco y funcionamiento familiar son las variables y constructos que más influencia estadística tienen en el estudio. No obstante, el cambio significativo se produce al incluir el consumo de alcohol, actividad física y el patrón de alimentación al análisis.

**Gráfica 5.** Puntos de leverages en orden descendente



**Fuente:** Elaboración propia con los resultados obtenidos por medio del rCUR (Bodor et al., 2012).

Finalmente, en la gráfica 6 se puede ver como efectivamente estas son las características que sobresalen analizando los puntos de leverages de las variables y constructos involucrados en el estudio. Cabe resaltar que, con estos resultados, los cinco aspectos que se incluyen al inicio en la medición de los hábitos de vida (IMC, alimentación, actividad física, tabaco y alcohol) deben mantenerse.

**Gráfica 6.** Selección de las variables y constructos de acuerdo a los leverages

**Fuente:** Elaboración propia por medio del rCUR (Bodor et al., 2012).

Los constructos y variables recuperadas a través del método CUR como aquellas que tienen alta influencia estadística en los jóvenes analizados, permiten tener un panorama sobre las relaciones que tienen en la influencia de la generación de ASF entre los adolescentes. De tal forma que las rutas de acción globales, para mejorar el AS en adolescentes, deben considerar la insatisfacción con la vida, la clase social, el IMC, la edad y la preocupación por la figura como las variables y constructos más importantes.

## **VI.5 La percepción de Apoyo Social como una respuesta a la edad de los jóvenes, su sexo, clase social, entorno y hábitos de vida. Un análisis a través de los métodos de segmentación.**

En la búsqueda de construir un perfil más preciso de los individuos que componen el colectivo de estudio, agruparlos para conocer mejor este subgrupo poblacional y tener herramientas que permitan proponer intervenciones eficaces, se realizó una clasificación del grupo EAS, a través de un árbol de segmentación basado en las técnicas TAID propuestas por Castro (2005) a través de R Project for Statistical Computing (R Development Core Team, 2013).

El algoritmo TAID propone iniciar el análisis realizando un análisis de clases latentes con las variables respuesta manifiestas para obtener una única variable respuesta y determinar sus clases latentes, sin embargo, el objetivo principal de esta investigación es conocer las características de los adolescentes según el AS percibido. De tal forma que el análisis se inicia a partir de su clasificación en los dos grupos con los que se ha venido trabajando (EAS y ASF).

El segundo paso para la creación del árbol de segmentación a través del TAID es encontrar el mejor predictor, para lo cual se realiza la tabla cruzada con los constructos iniciales (EAS y ASF) y cada una de las variables (o constructos) predictivas(os), por ejemplo, la tabla 22 muestra la relación entre AS y edad. Es importante aclarar que se consideraron las variables y constructos con los leverages más altos encontrados anteriormente<sup>23</sup> (edad, clase social, IMC, alimentación, actividad física, tabaco, alcohol, familia, satisfacción con la vida y preocupación por la figura) y la población obtenida a través del sobremuestreo estratificado<sup>24</sup>.

---

<sup>23</sup> Ver apartado VI.4

<sup>24</sup> Ver apartado III.2

**Tabla 22.** Tabla de contingencia entre el ASF como constructo respuesta y la edad de los adolescentes

	EAS	ASF
15 años	4	18
16 años	113	206
17 años	45	130
18 años	33	42

Fuente: Elaboración propia.

A partir de estos resultados se calcula el coeficiente de predictividad:

$$\tau_{i/j} = \frac{\left( \sum_{i=1}^I \sum_{j=1}^J \frac{f_{ij}^2}{f_{.j} f_{i.}} - \sum_{i=1}^I (f_{i.}/f_{..})^2 \right)}{1 - \sum_{i=1}^I (f_{i.}/f_{..})^2} = 0.019185$$

Enseguida se calcula el índice de Catanova:

$$C = (I - 1)(J - 1)\tau_{i/j} = 33.95884$$

el cual sigue una distribución  $\chi^2$  con 3 grados de libertad, encontrando un p-valor altamente significativo ( $2.021222 \cdot 10^{-7}$ ). Este proceso continúa con cada una de las otras variables y constructos participantes en el estudio. Así se determina el mejor predictor a través de los resultados mostrados en la tabla 23.

**Tabla 23.** Coeficientes de predictividad e índices de Catanova de todos los predictores

	$\tau$	Catanova	g.l.	p-valor
Edad	0.019185	33.95884	3	2.021222e-07
Clase Social	0.001931	3.417289	3	0.331649
IMC	0.003818	4.505307	2	0.1051199
Alimentación	0.011922	14.06762	2	0.0008815651
Actividad física	0.013185	15.55791	2	0.0004184493
Tabaco	0.0051745	12.21181	4	0.01584374
Alcohol	0.0054432	9.634376	3	0.02194397
<b>Familia</b>	<b>0.2714589</b>	<b>320.3215</b>	<b>2</b>	<b>2.773775e-70</b>
Satisfacción con la vida	0.0058054	13.70084	4	0.008313688
Preocupación por la figura	0.0077668	13.74715	3	0.003270324

Fuente: Elaboración propia.

Como es posible observar el mejor predictor para caracterizar a los adolescentes con EAS es la funcionalidad de su familiar, por lo que se considera segmentar a partir de este constructo.

La etapa tres del algoritmo TAID corresponde al análisis no simétrico de correspondencias entre el mejor predictor y el constructo respuesta para, a partir de los resultados agrupar las categorías. Las coordenadas estándar de cada categoría se presentan en la tabla 24.

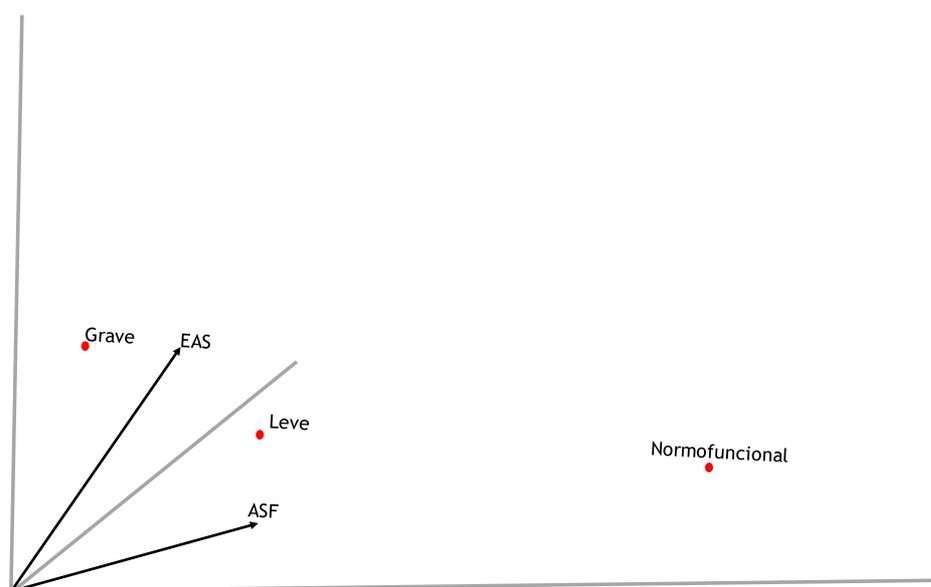
**Tabla 24.** Puntuaciones de cada categoría del predictor seleccionado.

	Eje 1	Eje 2
Disfunción Grave	-0.1493	0.8875
Disfunción Leve	-0.3324	0.3702
Normofuncional	-0.9312	-0.2745

Fuente: Elaboración propia.

Gráficamente se puede representar a través del Biplot como se muestra en la figura 20.

**Figura 20.** ACNS comparando el AS percibido y la funcionalidad de la familia



Fuente: Elaboración propia.

A través del análisis visual se puede considerar agrupar las categorías que refieren a una disfuncionalidad familiar y las cuales se separan de la que

representa a las familias normofuncionales. Clasificando las opciones según recomiendan Siciliano & Mola (1998), (ver tabla 25) se puede observar como la normofuncionalidad es la categoría más fuerte.

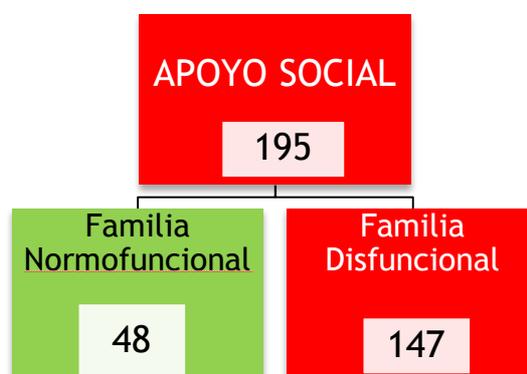
**Tabla 25.** Clasificación de las categorías del predictor

	$\varphi_{j1}$	Tipo de categoría
Disfunción Grave	-0.1493	Débil
Disfunción Leve	-0.3324	Débil
Normofuncional	-0.9312	Fuerte por la izquierda

Fuente: Elaboración propia.

A partir de este resultado se segmenta y se clasifica a los jóvenes como puede verse en la figura 21. De esta forma 48 jóvenes del grupo EAS tenían una familia normofuncional, mientras 147 familias presentaban algún tipo de disfunción.

**Figura 21.** Primer nivel del árbol de segmentación

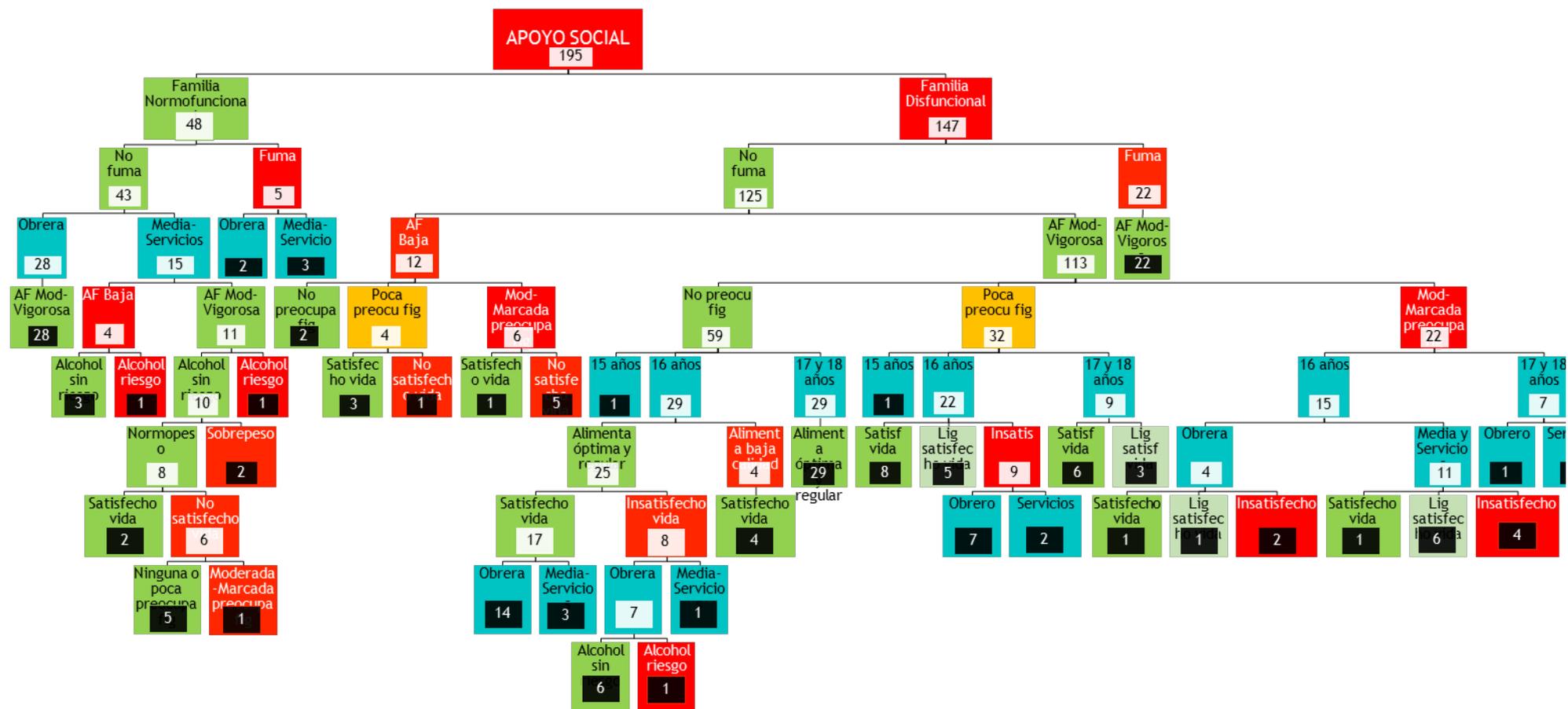


Fuente: Elaboración propia.

A partir de este resultado el proceso se repite seleccionando el mejor predictor hasta que no se encuentre alguno significativo.

Si sólo se considera este criterio de parada, sin tomar en cuenta que los nodos tengan al menos el 10% de la población en análisis se obtiene un árbol ternario como el que se presenta en la figura 22, con 38 perfiles terminales.

Figura 22. Árbol ternario de segmentación considerando hasta que ningún predictor tenga un resultado significativo



Fuente: Elaboración propia.

La sexta fase de un estudio de segmentación, de acuerdo con Kotler (1988) es seleccionar, desarrollar y crear estrategias de actuación para cada segmento objetivo. El análisis de un árbol con tantas clasificaciones no resulta óptimo para este fin, por lo que debe ser considerado añadir un criterio de parada que considere que para que un nodo sea segmentable este debe tener más del 10% de la muestra analizada.

De esta forma se obtiene el árbol que se muestra en la figura 23, con 17 nodos terminales. Puede observarse como después del nivel 4 de segmentación, la edad es el mejor predictor para continuar la división, considerando que se trata de individuos de un mismo nivel educativo, podrían considerarse sólo los 8 grupos terminales hasta este punto. De esta forma se debe tener en cuenta, para el desarrollo de estrategias dirigidas a los jóvenes con EAS, la funcionalidad de la familia, el consumo de tabaco, la clase social, la actividad física y la preocupación por la figura.

A partir de los ocho perfiles resultantes, descritos en la tabla 26, se pueden considerar tres grandes grupos de individuos. Los primeros (perfil 1 y 2) cuyos hábitos de vida y percepción del entorno están entre los rangos normales, sin embargo, consideran tener un EAS.

Un segundo grupo conformado por los perfiles 5 y 6, quienes también tienen hábitos de vida saludables, sin embargo, además de percibir un EAS también consideran tener una familia con algún tipo de disfuncionalidad (leve o grave).

En el resto de los perfiles presentan al menos algún hábito nocivo. De esta forma el perfil 3 y 8 se caracteriza por ser su adicción al tabaco. El perfil 4 por una actividad física por debajo de las recomendaciones de la OMS. Y el perfil 7, con una preocupación por su figura de moderada a marcada que puede traer otro tipo de nocivos para la salud.

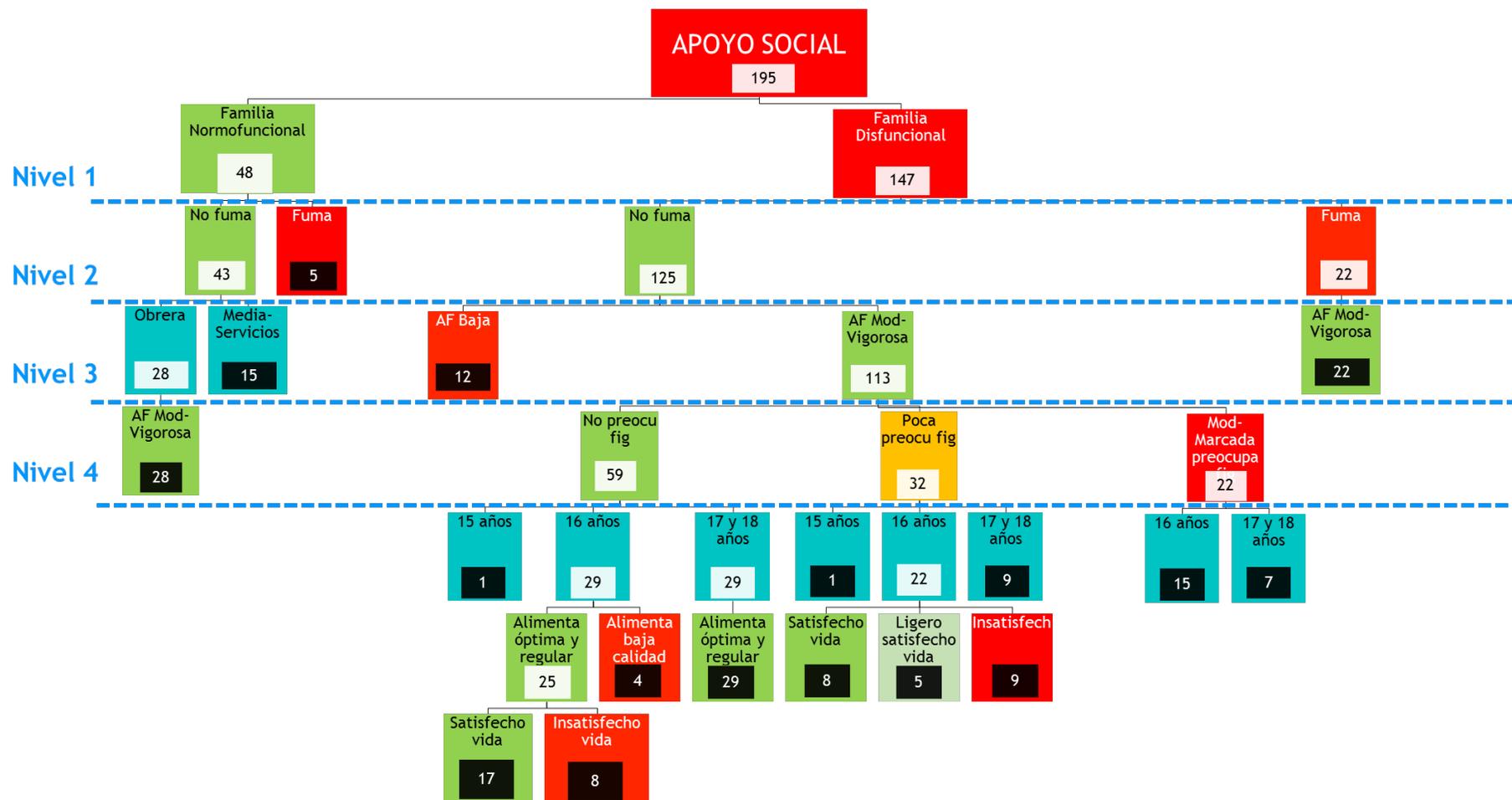
Esta caracterización puede ayudar a la creación estrategias de intervención que permitan ayudar a los jóvenes en mejorar su percepción de apoyo social de una forma más eficiente y eficaz.

Tabla 26. Perfiles de los nodos terminales del árbol final

Nodo	Descripción
1:	Familia normofuncional; no fuman; clase social obrera; actividad física vigorosa.
2:	Familia normofuncional; no fuman; clase social media o de servicios.
3:	Familia normofuncional; fuman.
4:	Familia con algún grado de disfunción; no fuman; actividad física baja.
5:	Familia con algún grado de disfunción; no fuman; actividad física moderada a vigorosa; sin preocupación por su figura.
6:	Familia con algún grado de disfunción; no fuman; actividad física moderada a vigorosa; con poca preocupación por su figura.
7:	Familia con algún grado de disfunción; no fuman; actividad física moderada a vigorosa; con preocupación por su figura de moderada a marcada.
8:	Familia con algún grado de disfunción; fuman.

Fuente: Elaboración propia.

Figura 23. Árbol ternario de segmentación final



Fuente: Elaboración propia.

---

# CONCLUSIONES

---

1. El análisis de datos a través del sobremuestreo estratificado permite enfrentar el problema de datos desbalanceados, produciendo pronósticos realistas maximizando la información de la clase minoritaria para resaltar sus características sin dejar de lado la comparación con la clase mayoritaria.
2. Se pone de manifiesto como las técnicas provenientes del Big Data como el TwoStep Cluster, la descomposición de matrices con el método CUR y los árboles de clasificación son herramientas útiles que permite detectar las variables y constructos que tienen alta influencia en Apoyo Social observado por los jóvenes preuniversitarios.
3. A través de la descomposición de matrices con el método CUR es posible señalar que la insatisfacción con la vida, la clase social y el IMC son aquellas variables que mayor influencia tienen en la percepción de Apoyo Social en los jóvenes preuniversitarios. Por lo que las intervenciones a gran escala deben considerar estas tres características al momento de planificar una intervención general.
4. Utilizando el algoritmo TAID se logró organizar y agrupar de manera eficiente a los jóvenes preuniversitarios, donde, a través de sus resultados se determinaron ocho perfiles mediante los cuales se pueden crear estrategias de intervención que permitan ayudar a los jóvenes en mejorar su percepción por el apoyo social.
5. El uso de metodología proveniente del Big Data en Ciencias Sociales permite, además de trabajar con grandes cantidades de datos en una menor cantidad de tiempo, crear estrategias de intervención más eficaces para mejorar el entorno de adolescentes y jóvenes.

---

# BIBLIOGRAFÍA

---

- Acosta G., M. V., & Gómez P., G. (2003). Insatisfacción corporal y seguimiento de dieta: una comparación transcultural entre adolescentes de España y México. *Revista Internacional de Psicología Clínica Y de La Salud*, 3(1), 9-21.
- Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., O'Brien, W. L., Bassett, D. R., Schmitz, K. H., Emplainscourt, P. O., Jacobs, D. R., & Leon, A. S. (2000). Compendium of Physical Activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, 32(9), S498-S516. DOI:10.1097/00005768-200009001-00009
- Ainsworth, B. E., Macera, C. A., Jones, D. A., Reis, J. P., Addy, C. L., Bowles, H. R., & Kohl, H. W. (2006). Comparison of the 2001 BRFSS and the IPAQ Physical Activity Questionnaires. *Medicine & Science in Sports & Exercise*, 38(9), 1584-1592. DOI:10.1249/01.mss.0000229457.73333.9a
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. DOI:10.1109/TAC.1974.1100705
- Akbani, R., Kwek, S., & Japkowicz, N. (2004). Applying Support Vector Machines to Imbalanced Datasets. En *Proceedings of the European conference on Machine Learning* (Vol. 3201, pp. 39-50). Catania, Italy: Springer-Verlag.
- Álamo M., M., Artiles R., M., Santiago N., P., Bernal B, I., Aguiar B., J., & Gómez A., A. (1999). Apoyo social funcional percibido por los pacientes en programa de atención domiciliaria en los medios rural y urbano. *Atención Primaria*, 24(1), 26-31.
- Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Beverly Hills: Sage Publications.
- Alho, A. R., & De Abreu e Silva, J. (2015). Utilizing urban form characteristics in urban logistics analysis: a case study in Lisbon, Portugal. *Journal of Transport Geography*, 42, 57-71. DOI:10.1016/j.jtrangeo.2014.11.002

- Arrighetti, A., & Ninni, A. (2014). *La trasformazione “silenziosa” : cambiamento strutturale e strategie d’impresa nell’industria italiana*. Dipartimento di Economia Università di Parma.
- Asher, S. R., Hymel, S., & Renshaw, P. D. (1984). Loneliness in Children. *Child Development*, 55(4), 1456-1464. DOI:10.2307/1130015
- Atienza, F. L., Pons, D., Balaguer, I., & García-Merita, M. (2000). Propiedades psicométricas de la Escala de Satisfacción con la Vida en Adolescentes. *Psicothema*, 12(2), 314-319.
- Ávila Z., C. A. (1993). *Métodos gráficos multivariantes y su aplicación en las ciencias de la vida* (Trabajo de grado). Universidad de Salamanca, Salamanca, España.
- Ávila Z., C. A. (1997). *Una alternativa al análisis de segmentación basada en el análisis de hipótesis de independencia condicionada* (Tesis doctoral). Universidad de Salamanca, Salamanca, España.
- Aymerich, M., Berra, S., Guillamón, I., Herdman, M., Alonso, J., Ravens-Sieberer, U., & Rajmil, L. (2005). Desarrollo de la versión en español del KIDSCREEN, un cuestionario de calidad de vida para la población infantil y adolescente. *Gaceta Sanitaria*, 19(2), 93-102. DOI:10.1157/13074363
- Bacher, J., Wenzig, K., & Vogler, M. (2004). SPSS TwoStep Cluster - a first evaluation. *Rechts-Und Diskussionspapiere / Universität Erlangen-Nürnberg, Sozialwissenschaftliches Institut, Lehrstuhl Für Soziologi*, 2, 23.
- Banfield, J. D., & Raftery, A. E. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, 49(3), 803-821. DOI:10.2307/2532201
- Barra A, E., Cerna C, R., Kramm M, D., & Véliz V, V. (2006). Problemas de Salud, Estrés, Afrontamiento, Depresión y Apoyo Social en Adolescentes. *Terapia Psicológica*, 24(1), 55-61.

- Barrón L. R., A. (1996). *Apoyo Social. Aspectos teóricos y aplicaciones*. (S. XXI, Ed.) (1a ed.). Madrid.
- Becoña, E., & Vázquez, F. L. (1998). The Fagerström test for nicotine dependence in a Spanish sample. *Psychological Reports*, 83(3), 1455-1458. DOI:10.2466/pr0.1998.83.3f.1455
- Bellón S., J., Delgado S., A., Luna C., J. de D., & Lardelli C., P. (1996a). Validez y fiabilidad del cuestionario de apoyo social funcional Duke-UNC-11. *Atención Primaria*, 18(4), 153-163.
- Bellón S., J., Delgado S., A., Luna del Castillo, J. de D., & Lardelli C., P. (1996b). Validez y fiabilidad del cuestionario de función familiar Apgar-familiar. *Atención Primaria*, 18(6), 289-296.
- Beresford, T., Low, D., Adduci, R., & Goggans, F. (1982). Alcoholism assessment on an orthopaedic surgery service. *The Journal of Bone and Joint Surgery. American Volume*, 64(5), 730-733.
- Berry, M. W., Pulatova, S. A., & Stewart, G. W. (2005). Algorithm 844: Computing Sparse Reduced-Rank Approximations to Sparse Matrices. *ACM Transactions on Mathematical Software*, 31(2), 252-269.
- Bodor, A., Csabai, I., Mahoney, M. W., & Solymosi, N. (2012). rCUR: an R package for CUR matrix decomposition. *BMC Bioinformatics*, 13(1), 103. DOI:10.1186/1471-2105-13-103
- Booth, M. (2000). Assessment of Physical Activity: An International Perspective. *Research Quarterly for Exercise and Sport*, 71(sup2), 114-120. DOI:10.1080/02701367.2000.11082794
- Born, D., Minujín, A., & Lombartdía, M. L. (2015). *Una exploración sobre la situación de las y los adolescentes y jóvenes en América Latina y el Caribe a partir de evidencia cuantitativa reciente*. Panamá.

- Bradford, J. P., Kunz, C., Kohavi, R., Brunk, C., & Brodley, C. E. (1998). Pruning Decision Trees with Misclassification Costs. En *European Conference on Machine Learning* (pp. 131-136).
- Broadhead, W. E., Gehlbach, S. H., De Gruy, F. V., & Kaplan, B. H. (1988). The Duke-UNC Functional Social Support Questionnaire. Measurement of social support in family medicine patients. *Medical Care*, 26(7), 709-723. DOI:10.1097/00005650-198807000-00006
- Broadhead, W. E., Kaplan, B. H., James, S. A., Wagner, E. H., Schoenbach, V. J., Grimson, R., Heyden, S., Tibblin, G., & Gehlbach, S. H. (1983). The epidemiologic evidence for a relationship between social support and health. *American Journal of Epidemiology*, 117(5), 521-537.
- Campo-Arias, A. (2007). Cuestionario general de salud-12: análisis de factores en población general de Bucaramanga, Colombia. *Iatreia*, 20(1), 29-36.
- Carrasco G., T. J., Luna A., M., & Vila C., J. (1994). Validez del "Fagerström Tolerance Questionnaire" como medida de dependencia física de la nicotina: una revisión. *Revista Española de Drogodependencias*, (1), 3-14.
- Casillas-Estrella, M., Montaña-Castrejón, N., Reyes-Velázquez, V., Bacardí-Gascón, M., & Jiménez-Cruz, A. (2006). A mayor IMC mayor grado de insatisfacción de la imagen corporal. *Rev Biomed*, 17, 243-249.
- Castro-López, C., Vicente-Galindo, P., Patino-Alonso, C., & Galindo-Villardón, P. (2016). Contribuciones a la Minería de Datos. En *XXVI Simposio Internacional de Estadística* (pp. 1-4). Sincelejo.
- Castro L., C. R. (2005). *Contribuciones a la detección de variables relevantes en tablas de contingencia multivariantes* (Tesis doctoral). Universidad de Salamanca, Salamanca, España.
- Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

- Chawla, N. V, Japkowicz, N., & Ko, A. (2004). Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM Sigkdd Explorations Newsletter*, 6(1), 1-6.
- Cházaro-Molina, E. A., & Muñoz-Cano, J. M. (2015). Comunicación educativa a pacientes con diabetes mellitus 2 y adherencia al tratamiento nutricional. *Revista Española de Comunicación En Salud*, 6(2), 126-137.
- Chen, C., Liaw, A., & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data. *University of California, Berkeley*, 110.
- Chin, Y. R., & Choi, K. (2015). Suicide Attempts and Associated Factors in Male and Female Korean Adolescents A Population-Based Cross-Sectional Survey. *Community Mental Health Journal*, 51(7), 862-866. DOI:10.1007/s10597-015-9856-6
- Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. En *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01* (pp. 263-268). New York: ACM Press. DOI:10.1145/502512.502549
- Cobb, S. (1976). Social support as a moderator of life stress. *Psychosomatic Medicine*, 38(5), 300-314. DOI:10.1097/00006842-197609000-00003
- Cole, T. J., Bellizzi, M. C., Flegal, K. M., & Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ (Clinical Research Ed.)*, 320(7244), 1240-3. DOI:10.1136/bmj.320.7244.1240
- Collins, M. E. (1991). Body figure perceptions and preferences among preadolescent children. *International Journal of Eating Disorders*, 10(2), 199-208. DOI:10.1002/1098-108X(199103)10:2<199::AID-EAT2260100209>3.0.CO;2-D

- Comisión Interamericana para el Control del Abuso de Drogas (CICAD), Secretaría de Seguridad Multidimensional (SSM), & Organización de Estados Americanos (OEA). (2015). *Informe sobre uso de drogas en las Américas*. Washington, D. C.
- Consejo Nacional de Población (CONAPO). (2017). Proyecciones de la población de México - Entidades Federativas 2010/2050. Recuperado de <http://bit.ly/2skr83L>
- Cooper, P. J., Taylor, M. J., Cooper, Z., & Fairbum, C. G. (1987). The development and validation of the body shape questionnaire. *International Journal of Eating Disorders*, 6(4), 485-494. DOI:10.1002/1098-108X(198707)6:4<485::AID-EAT2260060405>3.0.CO;2-O
- Cortes-Martínez, G., Vallejo-De la Cruz, N. L., Pérez-Salgado, D., & Ortiz-Hernández, L. (2009). Utilidad de siluetas corporales en la evaluación del estado nutricional en escolares y adolescentes de la Ciudad de México. *Boletín Médico Del Hospital Infantil de México*, 66(6), 511-521.
- Cortés R., E., Castro-Pozo, M. U., & Salazar C., C. M. (2015). *Juventudes contemporáneas. Visibilidad en el espacio urbano*. (1st ed.). Ciudad de México: Instituto de Altos Estudios sobre Deporte, Cultura y Sociedad, A. C.
- Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise*, 35(8), 1381-1395. DOI:10.1249/01.MSS.0000078924.61453.FB
- Crone, S. F., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*, 28(1), 224-238. DOI:10.1016/j.ijforecast.2011.07.006
- Cuadras, C. (1989). Distancias Estadísticas. *Estadística Española*, 30(119), 295-378.

- Cuadras, C. M. (2014). *Nuevos métodos de análisis multivariante*. Barcelona: CMC Editions.
- Dabkowski, M., Ziótkowski, M., Rogiewicz, M., & Rybakowski, J. (1989). Severity of alcohol dependence and social functioning of male patients with alcoholism. II. Functioning in parental, employee's and autonomous roles. *Psychiatria Polska*, 23(4), 294-299.
- De la Revilla, L., Marcos, B., Castro, J. A., Aybar, R., Marín, I., & Mingorance, I. (1994). Percepción de salud, apoyo social y función familiar en VIH positivo. *Atención Primaria*, 13(4), 161-164.
- De Leon, J., Diaz, F. J., Becoña, E., Gurpegui, M., Jurado, D., & Gonzalez-Pinto, A. (2003). Exploring brief measures of nicotine dependence for epidemiological surveys. *Addictive Behaviors*, 28(8), 1481-1486.
- Delgado-Quiñone, E. G., & Martín-Guzmán, V. M. (2015). Relación entre percepción de funcionalidad familiar y dependencia a la nicotina en jóvenes fumadores de 15 a 24 años. *Revista Médica MD*, 6.7(4), 253-257.
- Diario Oficial de la Federación (2016). Decreto de reforma. Recuperado en junio 03 de 2017, Recuperado de <http://bit.ly/1WSokXa>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1), 71-75. DOI:10.1207/s15327752jpa4901\_13
- Dorado D., A. (1998). *Métodos de búsqueda de variables relevantes en análisis de segmentación: aportaciones desde una perspectiva multivariante* (Tesis doctoral). Universidad de Salamanca, Salamanca, España.
- Dotan-Cohen, D., Melkman, A. A., & Kasif, S. (2007). Hierarchical tree snipping: clustering guided by prior knowledge, 23(24), 3335-3342. DOI:10.1093/bioinformatics/btm526

- Drineas, P., Kannan, R., & Mahoney, M. W. (2006). Fast Monte Carlo Algorithms For Matrices III: Computing A Compressed Approximate Matrix Decomposition. *Society for Industrial and Applied Mathematics*, 36(1), 184-206. DOI:10.1137/S0097539704442702
- Dubey, R., Zhou, J., Wang, Y., Thompson, P. M., & Ye, J. (2014). Analysis of sampling techniques for imbalanced data: An n = 648 ADNI study. *NeuroImage*, 87, 220-241. DOI:10.1016/j.neuroimage.2013.10.005
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211-218. DOI:10.1007/BF02288367
- Elkan, C. (2001). The Foundations of Cost-Sensitive Learning. En *17th international joint conference on Artificial intelligence* (pp. 973-978). Seattle.
- Erikson, R., & Goldthorpe, J. J. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- Escobar, M. (1998). Las aplicaciones del análisis de segmentación: El procedimiento Chaid. *Empiria: Revista de Metodología de Ciencias*.
- Estrada P., C., Chan G., E. C., Herrero O., J. B., Rodríguez D., F. J., Herrero D., F. J., & Bringas M., C. (2009). *Apoyo social y colectivos vulnerables: una herramienta para la intervención contra la violencia de género*. UNIOVI. Oviedo: Universidad de Oviedo : AECID.
- Evans, C., & Dolan, B. (1993). Body Shape Questionnaire: derivation of shortened "alternate forms." *The International Journal of Eating Disorders*, 13(3), 315-321. DOI:10.1002/1098-108X(199304)13:3<315::AID-EAT2260130310>3.0.CO;2-3
- Ewing, J. A. (1984). Detecting alcoholism. The CAGE questionnaire. *JAMA*, 252(14), 1905-1907. DOI:10.1001/jama.1984.03350140051025

- Fagerström, K. O. (1978). Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors*, 3(3-4), 235-241. DOI:10.1016/0306-4603(78)90024-2
- Fagerström, K. O., Kunze, M., Schoberberger, R., Breslau, N., Hughes, J. R., Hurt, R. D., Puska, P., Ramstrom, L., & Zatonski, W. (1996). Nicotine dependence versus smoking prevalence: comparisons among countries and categories of smokers. *Tobacco Control*, 5(1), 52-56.
- Feito A., R. (1995). *ESTRUCTURA SOCIAL CONTEMPORANEA: Las clases sociales en los países industrializados*. Madrid: Siglo XXI.
- Felton, B. J., & Shinn, M. (1992). Social integration and social support: Moving “social support” beyond the individual level. *Journal of Community Psychology*, 20(2), 103-115. DOI:10.1002/1520-6629(199204)20:2<103::AID-JCOP2290200202>3.0.CO;2-4
- Fernández S., O. (1991). El Análisis de Cluster: aplicación, interpretación y validación. *Papers: Revista de Sociología*, 37, 65-76. DOI:10.5565/rev/papers/v37n0.1596
- Forero A., L. M., Avendaño D., M. C., Duarte C., Z. J., & Campo A., A. (2006). Consistencia interna y análisis de factores de la escala APGAR para evaluar el funcionamiento familiar en estudiantes de básica secundaria. *Revista Colombiana de Psiquiatría*, 35(1), 23-29.
- Fraley, C., & Raftery, A. E. (1998). How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*, 41(8), 578-588. DOI:10.1093/comjnl/41.8.578
- Francisco, D. (2012). Poor mental health symptoms among Romanian employees. A Two-Step Cluster analysis. *Procedia - Social and Behavioral Sciences*, 33, 293-297. DOI:10.1016/j.sbspro.2012.01.130
- Frank, R. E., Massy, W. F., & Wind, Y. (1972). *Market segmentation*. New Jersey: Prentice-Hall.

- Frieze, A., Kannan, R., & Vempala, S. (2004). Fast monte-carlo algorithms for finding low-rank approximations. *Journal of the ACM*, 51(6), 1025-1041. DOI:10.1145/1039488.1039494
- Fullerton, C., Acuña, J., Florenzano, R., Cruz, C., & Weil, K. (2003). Psicopatología en pacientes hospitalizados en un hospital general. *Revista Chilena de Neuro-Psiquiatría*, 41(2), 103-109. DOI:10.4067/S0717-92272003000200003
- Gadermann, A. M. (2009). *The Satisfaction with Life Scale adapted for Children : investigating the structural, external, and substantive aspects of construct validity* (Tesis doctoral). Universidad de Columbia, Vancouver, Canadá. DOI:10.14288/1.0054573
- Gadermann, A. M., Schonert-Reichl, K. A., & Zumbo, B. D. (2010). Investigating Validity Evidence of the Satisfaction with Life Scale Adapted for Children. *Social Indicators Res*, 96(2), 229-247. DOI:10.1007/s11205-009-9474-1
- Galiana, L., Gutiérrez, M., Sancho, P., Oliver, A., & Tomás, J. M. (2015). Propiedades psicométricas de la versión española de la Escala Temporal de Satisfacción con la Vida (Temporal Satisfaction With Life Scale, TSLS): Un estudio en mayores que acuden a programas universitarios. *European Journal of Investigation in Health, Psychology and Education*, 5(3), 335-344. DOI:10.1989/ejihpe.v5i3.136
- Gálvez C., A., Rosa G., A., García-Cantó, E., Rodríguez G., P. L., Pérez-Soto, J. J., Tarraga M., L., & Tarraga L., P. (2015). Estado nutricional y calidad de vida relacionada con la salud en escolares del sureste español. *Nutrición Hospitalaria*, 31(2), 737-743.
- García F., L., & Garita A., E. (2007). Relación entre la satisfacción con la imagen corporal, autoconcepto físico, índice de masa corporal y factores socioculturales en mujeres adolescentes costarricenses. *Revista de Ciencias Del Ejercicio Y La Salud*, 5(1), 9-18. DOI:10.15517/pensarmov.v5i1.356

- García S., L. A. (2015). *Entornos socioculturales, alimentación y estilos de vida en relación con la obesidad infantil y adolescente. Zacatecas-México* (Tesis doctoral). Universidad de Extremadura, Badajoz, España.
- García V., C. R. (1999). Manual para la utilización del cuestionario de salud general de Goldberg: Adaptación cubana. *Revista Cubana de Medicina General Integral*, 15(1), 88-97.
- Gardner, R. M., Jappe, L. M., & Gardner, L. (2009). Development and validation of a new figural drawing scale for body-image assessment: the BIAS-BD. *Journal of Clinical Psychology*, 65(1), 113-122. DOI:10.1002/jclp.20526
- Gardner, R. M., Stark, K., Jackson, N. A., & Friedman, B. N. (1999). Development and validation of two new scales for assessment of bodyimage. *Perceptual and Motor Skills*, 89(3), 981-993. DOI:10.2466/pms.1999.89.3.981
- Gil G, D. S., & Vargas N, J. A. (1993). Detección de un outlier superior en muestras exponenciales basado en predicción de la mayor observación. *Revista Colombiana de Estadística*, 14(28), 37-44.
- Gini, C. (1912). *Variabilità e mutabilità: contributo allo studio delle relazioni statistiche. Studi Economico-giuridici. Facolta di Giurisprudenza*. Bologna: Università di Cagliari.
- Goldberg, D. P., & Blackwell, B. (1970). Psychiatric illness in general practice. A detailed study using a new method of case identification. *British Medical Journal*, 2(5707), 439-443. DOI:10.1136/bmj.2.5707.439
- Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, 27(1), 191-197. DOI:10.1017/S0033291796004242

- Goldthorpe, J. H. (1997). The “Goldthorpe” class schema: Some observations on conceptual and operational issues in relation to the ESRC review of government social classifications. En Rose, D., & O’Reilly, K. (Eds.), *Constructing classes: Towards a new social classification for the UK* (pp. 40-48). London: Office for National Statistics.
- Goldthorpe, J. H. (2012). De vuelta a la clase y el estatus: por qué debe reivindicarse una perspectiva sociológica de la desigualdad social. *Reis*, 137, 43-58. DOI:10.54777/cis/reis.137.43
- Goldthorpe, J. H., & McKnight, A. (2004). The Economic Basis of Social Class. En Morgan, S., Grusky, D. B., & Fields, G. S. (Eds.), *Mobility and Inequality: Frontiers of Research from Sociology and Economics* (p. 36). Stanford: Stanford University Press.
- González, J. J. (1992). La construcción empírica de las clases. *Política Y Sociedad*, 11, 99-122.
- González G., Y., López S., P., Saavedra R., M. C., González A., D., García L., L. J., & Cuesta C., B. (1997). Detección precoz de jóvenes con trastornos relacionados con el alcohol en atención primaria. *Atención Primaria*, 20(3), 133-136.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of Association for Cross Classifications. *Journal of the American Statistical Association*, 49(268), 732-764. DOI:10.1080/01621459.1954.10501231
- Gordon, A. D. (1998). Cluster Validation. En Hayashi, C., Yajima, K., Bock, H. H., Ohsumi, N., Tanaka, Y., & Baba, Y. (Eds.), *Data Science, Classification, and Related Methods* (pp. 22-39). Tokyo: Springer. DOI:10.1007/978-4-431-65950-1\_2
- Goreinov, S. A., Tyrtshnikov, E. E., & Zamarashkin, N. L. (1997). A theory of pseudoskeleton approximations. *Linear Algebra and Its Applications*, 261(1-3), 1-21. DOI:10.1016/S0024-3795(96)00301-1

- Goreinov, S., & Tyrtysnikov, E. (2001). The maximal-volume concept in approximation by low-rank matrices. *Contemporary*.
- Gower, J. C. (1985). Similarity, Dissimilarity, and Distance Measure. En *Encyclopedia of Biostatistics*. Chichester: John Wiley & Sons, Ltd. DOI:10.1002/0470011815.b2a10084
- Gracia F., E. (1997). *El apoyo social en la intervención comunitaria* (1a ed.). Paidós.
- Guha, S., Rastogi, R., & Shim, K. (2000). ROCK: A robust clustering algorithm for categorical attributes. *Information Systems*, 25(5), 345-366. DOI:10.1016/S0306-4379(00)00022-3
- Hallal, P. C., Gómez, L. F., Parra, D. C., Lobelo, F., Mosquera, J., Florindo, A. A., Reis, R. S., Pratt, M., & Sarmiento, O. L. (2010). Lecciones aprendidas después de 10 Años del uso de IPAQ en Brasil y Colombia. *Journal of Physical Activity and Health*, 7(Suppl 2), 259-264.
- Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K. O. (1991). The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction*, 86(9), 1119-1127. DOI:10.1111/j.1360-0443.1991.tb01879.x
- Herrán, O. F., & Ardila, M. F. (2005). Validez interna y reproducibilidad de la prueba CAGE en Bucaramanga, Colombia. *Biomédica*, 25(2), 231-241.
- Hidalgo-Rasmussen, C. A., Rajmil, L., & Espinoza, R. M. (2014). Adaptación transcultural del cuestionario KIDSCREEN para medir calidad de vida relacionada con la salud en población mexicana de 8 a 18 años. *Ciência & Saúde Coletiva*, 19(7), 2215-2224. DOI:10.1590/1413-81232014197.09682013
- Hidalgo-Rasmussen, C. A., Ramírez-López, G., & Hidalgo-San Martín, A. (2013). Actividad física, conductas sedentarias y calidad de vida en adolescentes universitarios de Ciudad Guzmán, Jalisco, México. *Cien Saude Colet*, 18(7),

1943-1952. DOI:10.1590/S1413-81232013000700009

Horowitz C., N., Florenzano U., R., & Ringeling P., I. (1985). Familia y salud familiar. Un enfoque para la atención primaria. *Bol of Sanit Panam*, 98(2), 144-156.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417-441. DOI:10.1037/h0071325

Howley, E. (2000). You asked for it Question Authority. *ACSM's Health & Fitness Journal*, 4(6), 6-26.

Huitrón-Bravo, G., Denova-Gutiérrez, E., Halley-Castillo, E., Santander-Rigollet, S., Bórquez-Puga, M., Zapata-Pérez, L., Huinca, B., Zubarew, T. G., & Villarroel-del-Pino, L. (2011). Conductas de riesgo en una muestra de adolescentes chilenos y mexicanos: un estudio comparativo. *Papeles de Población*, 17(70), 33-47.

Iacoponi, E., Laranjeira, R. R., & Jorge, M. R. (1989). At risk drinking in primary care: report from a survey in Sao Paulo, Brazil. *British Journal of Addiction*, 84(6), 653-658.

IBM. (2016). Software Package for Social Sciences.

IBM Knowledge Center. (2001). TwoStep Cluster Node. Recuperado de <https://ibm.co/2sU9DWI>

Instituto Mexicano de la Juventud (IMJUVE), & Universidad Nacional Autónoma de México (UNAM). (2012). *Encuesta Nacional de Valores en Juventud 2012*. Recuperado de <http://bit.ly/2umVTl7>

Instituto Mexicano de la Juventud (IMJUVE). (2010). *Encuesta Nacional De Juventud 2010 Resultados Generales*. Recuperado de <http://bit.ly/1i1Mu7B>

Instituto Nacional de Estadística y Geografía (INEGI). (2017a). Mortalidad general - Consulta interactiva de datos. Recuperado de <http://bit.ly/2tp6aQG>

Instituto Nacional de Estadística y Geografía (INEGI). (2017b). *Principales causas de mortalidad por residencia habitual, sexo y grupo de edad del fallecido*. Recuperado de <http://bit.ly/1bCDGTH>

Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRF), Instituto Nacional de Salud Pública (INSP), Secretaría de Salud (SS), Medina-Mora, M. E., Villatoro-Velázquez, J. A., Fleiz-Bautista, C., Téllez-Rojo, M. M., Mendoza-Alvarado, L. R., Romero-Martínez, M., Gutiérrez-Reyes, J. P., Castro-Tinoco, M., Hernández-Ávila, M., Tena-Tamayo, C., Alvear-Sevilla, C., & Guisa-Cruz, V. (2011). *Encuesta Nacional de Adicciones 2011: alcohol*. Ciudad de México: Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Instituto Nacional de Salud Pública; Secretaría de Salud. Recuperado de <http://bit.ly/1dHe9qe>

Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRF), Instituto Nacional de Salud Pública (INSP), Secretaría de Salud (SS), Reynales-Shigematsu, L. M., Guerrero-López, C. M., Lazcano-Ponce, E., Villatoro-Velázquez, J. A., Medina-Mora, M. E., Fleiz-Bautista, C., Téllez-Rojo, M. M., Mendoza-Alvarado, L. R., Romero-Martínez, M., Gutiérrez-Reyes, J. P., Castro-Tinoco, M., Hernández-Ávila, M., Tena-Tamayo, C., Alvear-Sevilla C., & Guisa-Cruz, V. (2011). *Encuesta Nacional de Adicciones 2011: tabaco*. Ciudad de México: Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Instituto Nacional de Salud Pública; Secretaría de Salud.

Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (INPRF), Instituto Nacional de Salud Pública (INSP), Secretaría de Salud (SS), Villatoro-Velázquez, J., Medina-Mora, M., Fleiz-Bautista, C., Téllez-Rojo, M. M., Mendoza-Alvarado, L. R., Romero-Martínez, M., Gutiérrez-Reyes, J. P., Castro-Tinoco, M., Hernández-Ávila, M., Tena-Tamayo, C., Alvear-Sevilla C., & Guisa-Cruz, V. (2011). *Encuesta Nacional de Adicciones 2011: drogas*. Ciudad de México: Instituto Nacional de Psiquiatría Ramón de la Fuente

Muñiz; Instituto Nacional de Salud Pública; Secretaría de Salud.

Instituto Nacional de Salud Pública (INSP), & Secretaría de Salud (SS). (2016). *Encuesta Nacional de Salud y Nutrición de Medio Camino 2016*. Ciudad de México: Secretaría de Salud. Recuperado de <http://bit.ly/2tslQIU>

International Society for the Advancement of Kinanthropometry (ISAK). (2010). *Kinanthropometry X Proceedings of the 10th International Society for the Advancement of Kinanthropometry Conference, held in Conjunction With the 13th Commonwealth International Sport Conference*. (M. J. Marfell-jones & T. Olds, Eds.). Routledge.

IPAQ Research Committee. (2002). Cuestionario Internacional de Actividad Física. Recuperado de <http://bit.ly/2to6mQA>

IPAQ Research Committee. (2005). Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. Recuperado de <http://bit.ly/2tsBnU2>

Japkowicz, N. (2000). The Class Imbalance Problem: Significance and Strategies. En *Proceedings of the 2000 International Conference on Artificial Intelligence* (Vol. 1, pp. 111-117).

Jo, T., & Japkowicz, N. (2004). Class imbalances versus small disjuncts. *ACM SIGKDD Explorations Newsletter*, 6(1), 40. DOI:10.1145/1007730.1007737

Kass, G. V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2), 119. DOI:10.2307/2986296

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An introduction to cluster analysis*. En Kaufman, L., & Rousseeuw, P. J. Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc. DOI:10.1002/9780470316801

KIDSCREEN Group Europe. (2006). Los cuestionarios KIDSCREEN - cuestionarios de calidad de vida para los niños y adolescentes. Recuperado de <http://bit.ly/2tsSS6H>

- King, M. (1986). At risk drinking among general practice attenders: validation of the CAGE questionnaire. *Psychological Medicine*, 16(1), 213-217.
- Knoll, U., Nakhaeizadeh, G., & Tausend, B. (1994). Cost-sensitive pruning of decision trees. En *Proceedings of the European conference on machine learning on Machine Learning* (pp. 383-386). Catania, Italy: Springer-Verlag.
- Kontogianni, M. D., Vidra, N., Farmaki, A. E., Koinaki, S., Belogianni, K., Sofrona, S., Magkanari, F., & Yannakoulia, M. (2008). Adherence rates to the Mediterranean diet are low in a representative sample of Greek children and adolescents. *The Journal of Nutrition*, 138(10), 1951-1956.
- Kotler, P. (1988). *Marketing management: analysis, planning, implementation, and control* (6th ed.). New Jersey: Prentice Hall.
- Kubat, M., & Matwin, S. (1997). Addressing the Curse of Imbalanced Training Sets: One-Sided Selection. En *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 179-186).
- Lara-Rivas, G., Ramírez-Venegas, A., Sansores-Martínez, R., Espinosa, A. M., & Regalado-Pineda, J. (2007). Indicadores de síntomas de abstinencia en un grupo de fumadores mexicanos. *Salud Pública de México*, 49(Suppl. 2), S257-S262.
- Lauro, N., & D'Ambra, L. (1984). Analyse Non Symétrique des Correspondances. En Diday E., & Coll. (Eds.), *Data Analysis and Informatics III*. (pp. 433-446). Amsterdam: Elsevier.
- Lee, P. H., Macfarlane, D. J., Lam, T., & Stewart, S. M. (2011). Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1), 115. DOI:10.1186/1479-5868-8-115

- Lewis, G., & Araya, R. I. (1995). Is the General Health Questionnaire (12 item) a culturally biased measure of psychiatric disorder? *Social Psychiatry and Psychiatric Epidemiology*, 30(1), 20-25.
- Light, R. J., & Margolin, B. H. (1971). An Analysis of Variance for Categorical Data. *Journal of the American Statistical Association*, 66(335), 534-544. DOI:10.1080/01621459.1971.10482297
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.
- Lin, N., Dean, A., & Ensel, W. M. (1979). Development of social support scales. *Center for Health Services Research, Washington, DC*.
- Lin, N., Dean, A., & Ensel, W. M. (2013). *Social Support, Life Events, and Depression*. Elsevier Science.
- Linage-Rivadeneira, M., & Lucio-Gómez-Maqueo, M. E. (2012). Asociación entre consumo de sustancias y estrategias de afrontamiento en jóvenes. *Journal of Behavior, Health & Social Issues*, 4(1), 55-66. DOI:10.5460/jbhsi.v4.1.32971
- Liu, Y., Yu, X., Huang, X., & An, A. (2009). Blog Data Mining: The Predictive Power of Sentiments. En *Data Mining for Business Applications* (pp. 183-195). Boston, MA: Springer US. DOI:10.1007/978-0-387-79420-4\_13
- López-Castedo, A., & Fernández, L. (2005). Psychometric Properties of the Spanish Version of the 12-Item General Health Questionnaire in Adolescents. *Perceptual and Motor Skills*, 100(3), 676-680. DOI:10.2466/pms.100.3.676-680
- López P., A. M., Solé B., A., & Cortés M., I. (2008). Percepción de satisfacción-insatisfacción de la imagen corporal en una muestra de adolescentes de Reus (Tarragona). *Zainak. Cuadernos de Antropología-Etnografía*, (30), 125-146.

- Luque H., L., Ruíz G., J., Gómez A., C., & Rodríguez O., A. R. (2008). Prevalencia de tabaquismo y grado de adicción a la nicotina en adolescentes de una escuela preuniversitaria urbana de Morelia. *Revista Cubana de Pediatría*, 80(4).
- Maddaleno, M. (1986). Enfoque familiar y los problemas de salud del adolescente. En Florenzano, R., Horwitz, N., Rigeling, Y., & Venturini, G. (Eds.), *Salud Familiar* (pp. 125-145). Chile: Universidad de Chile.
- Maddaleno H, M., Horwitz C, N., Jara V., C., Florenzano Urzua, R., & Zalazar R., D. (1987). Aplicación de un instrumento para calificar el funcionamiento familiar en la atención de adolescentes. *Rev. Chil. Pediatr*, 58(3), 246-249.
- Mahoney, M. W., & Drineas, P. (2009). CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(3), 697-702. DOI:10.1073/pnas.0803205106
- Mahoney, M. W., Maggioni, M., & Drineas, P. (2006). Tensor-CUR decompositions for tensor-based data. En *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06* (p. 327). New York: ACM Press. DOI:10.1145/1150402.1150440
- Mallo F., F. (2011). *Modelos Multivariantes Internos De Medición De Riesgos De Crédito, Acordes Con BASILEA II* (Tesis doctoral). Universidad de Salamanca, Salamanca, España.
- Mantilla T., S. C., & Gómez-Conesa, A. (2007). El Cuestionario Internacional de Actividad Física. Un instrumento adecuado en el seguimiento de la actividad física poblacional. *Revista Iberoamericana de Fisioterapia Y Kinesiología*, 10(1), 48-52. DOI:10.1016/S1138-6045(07)73665-1
- Mari, J., & Williams, P. (1985). Una comparación de la validez de dos cuestionarios de seguimiento psiquiátrico (GHQ-12 y CAR-20) en Brasil, con relativo análisis de características operativas (ROC). *Psychological Medicine*, 15(2), 651-659.

- Mariscal-Arcas, M., Rivas, A., Velasco, J., Ortega, M., Caballero, A. M., & Olea-Serrano, F. (2009). Evaluation of the Mediterranean Diet Quality Index (KIDMED) in children and adolescents in Southern Spain. *Public Health Nutrition*, 12(9), 1408. DOI:10.1017/S1368980008004126
- Markowitz, M. (1984). Alcohol misuse as a response to perceived powerlessness in the organization. *Journal of Studies on Alcohol*, 45(3), 225-227.
- Marrodán, M. D., López-Ejeda, N., González-Montero de Espinosa, M., Cabañas A., M. D., Villarino M., A., García A., R. M., Prado M., C., & Carmenate M., M. (2014). Adhesión a la dieta mediterránea y presión arterial en la infancia y adolescencia. En Martínez Á., J. R., & Villarino M., A. (Eds.), *Dieta mediterranea. Avances en alimentación, nutrición y dietética* (pp. 67-78). Madrid: Fundación Alimentación Saludable.
- Marrodán, M. D., Montero-Roblas, V., Mesa, M. S., Pachecho, J. L., González, M., Bejarano, I., Lomaglio, D. B., Verón, J. A., & Carmenate, M. (2008). Realidad, percepción y atractivo de la imagen corporal: condicionantes biológicos y socioculturales. *Zainak, Cuadernos de Antropología-Etnografía*, 30, 15-28.
- Martín P., C. (2005). *Prevalencia y atención a los trastornos mentales en una comunidad rural* (Tesis doctoral). Universidad de Granada, Granada, España.
- Martínez R., E. (1984). Aspectos teóricos del Análisis de Clúster y Aplicación a la caracterización del electorado potencial de un partido. En Sánchez C., J. J. (Ed.), *Introducción a las técnicas de Análisis Multivariable aplicadas a las Ciencias Sociales* (p. 165). Madrid: CIS.
- Masur, J., & Monteiro, M. G. (1983). Validation of the "CAGE" alcoholism screening test in a Brazilian psychiatric inpatient hospital setting. *Brazilian Journal of Medical and Biological*, 16(3), 215-218.

- Mayfield, D., McLeod, G., & Hall, P. (1974). The CAGE questionnaire: validation of a new alcoholism screening instrument. *The American Journal of Psychiatry*, 131(10), 1121-1123. DOI:10.1176/ajp.131.10.1121
- Meila, M., & Heckerman, D. (1998). An Experimental Comparison of Several Clustering and Initialization Methods. *Microsoft Research Technical Report*.
- Mendoza, N. J. (2007). *Efectos de un programa social sobre el desarrollo social, los estilos de vida y la calidad de vida relacionada con la salud en población rural venezolana validación transcultural de la medida de salud SF-36 en población rural de Venezuela* (Tesis doctoral). Universidad de Alicante, Alicante, España.
- Mendoza E., H. (2011). Los estudios sobre la juventud en México. *Espiral (Guadalajara)*, 18(52), 193-224.
- Meneses-Gaya, I. C. de, Zuardi, A. W., Loureiro, S. R., & Crippa, J. A. de S. (2009). Psychometric properties of the Fagerström Test for Nicotine Dependence. *Jornal Brasileiro de Pneumologia*, 35(1), 73-82.
- Ming-Yi, S., Jar-Wen, J., & Lien-Fu, L. (2010). A Two-Step Method for Clustering Mixed Categorical and Numeric Data. *Tamkang Journal of Science and Engineering*, 13(1), 11-19. DOI:10.6180/jase.2010.13.1.02
- Morgan, J. N., & Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58(301), 415-434.
- Motulsky, H., & Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression : a practical guide to curve fitting*. Oxford: Oxford University Press.
- Munera V., R., Zuluaga O., M. T., & Vélez L. M., N. (1994). Prevalencia del consumo de alcohol en la población escolarizada del municipio de Sabaneta 1992-1993. *CES Medicina*, 8(2), 73-80.

- Muracén D., I., Martínez J., A., Aguilar M., J. M., & González R., M. R. (2001).  
Pesquisaje de alcoholismo en un área de salud. *Revista Cubana de Medicina  
General Integral*, 17(1), 62-67.
- Musitu, G., & Cava, M. J. (2003). El rol del apoyo social en el ajuste de los  
adolescentes. *Intervención Psicosocial*, 12(2), 179-192.
- Niquille, M., Burnand, B., & Yersin, B. (1990). Alcoholism and the practicing  
physician. *Therapeutische Umschau. Revue Therapeutique*, 47(9), 717-  
723.
- Ochoa, S., Villareal, L., Molina, N., & Bayle, I. (2008). Validación y confiabilidad  
del Body Shape Questionnaire (BSQ) en adolescentes de la Ciudad de  
Colima. *Revista Científica Electrónica de Psicología*, 4, 204-216.
- OMS. (2009). *Marco para el seguimiento y evaluación de la aplicación.  
Estrategia mundial de la OMS sobre régimen alimentario, actividad física  
y salud*. Ginebra: OMS.
- OMS. (2013a). Actividad física. Recuperado de <http://bit.ly/1lCcwCr>
- OMS. (2013b). La actividad física en los jóvenes. Recuperado de  
<http://bit.ly/1g36wL6>
- OMS. (2015). *Desarrollo en la adolescencia*. World Health Organization.  
Recuperado de <http://bit.ly/1WldmmP>
- OMS. (2017). Adolescentes: riesgos para la salud y soluciones. WHO.  
Recuperado de <http://bit.ly/1B9NMX3>
- Organización Panamericana de la Salud (OPS). (2016). *Informe sobre el Control  
del Tabaco en la Región de las Américas. A 10 años del Convenio Marco de  
la Organización Mundial de la Salud para el Control del Tabaco*.  
Washington, DC: Organización Panamericana de la Salud. Recuperado de  
<http://bit.ly/2r9n6sl>

- Orosz, K., Farkas, I. J., & Pollner, P. (2016). Quantifying the changing role of past publications. *Scientometrics*, 108(2), 829-853. DOI:10.1007/s11192-016-1971-9
- Padmaja, M. T., Krishna, R. P., & Bapi, R. S. (2008). Majority filter-based minority prediction (MFMP): An approach for unbalanced datasets. En *TENCON 2008-2008 IEEE Region 10 Conference* (pp. 1-6). IEEE. DOI:10.1109/TENCON.2008.4766705
- Padrón B., F., Gutiérrez H., C. Y., & Medina C., M. A. (2015). Propiedades Psicométricas de la escala de satisfacción con la vida (SWLS) de Diener en población de Michoacán (México). *Avances En Psicología Latinoamericana*, 33(2), 223-232. DOI:10.12804/apl33.02.2015.04
- Palacios D., J. R., Sánchez T., B., & Andrade P., P. (2010). Intento de suicidio y búsqueda de sensaciones en adolescentes. *Revista Intercontinental de Psicología Y Educación*, 12(1), 53-75.
- Páramo, M. Á. (2011). Factores de Riesgo y Factores de Protección en la Adolescencia: Análisis de Contenido a través de Grupos de Discusión. *Terapia Psicológica*, 29(1), 85-95. DOI:10.4067/S0718-48082011000100009
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5(2), 164-172. DOI:10.1037/1040-3590.5.2.164
- Pavot, W., & Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3(2), 137-152. DOI:10.1080/17439760701756946
- Pazzani, M., Merz, C., Murphy, P., & Ali, K. (1994). Reducing misclassification costs. En *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 217-225).

- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302), 157-175. DOI:10.1080/14786440009463897
- Pearson, K. (1901). Mathematical Contributions to the Theory of Evolution. X. Supplement to a Memoir on Skew Variation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 197, 443-459.
- Pedret Y., R. (1986). *Técnicas cuantitativas al servicio del marketing: métodos descriptivos de análisis multivariable*. Universitat de Barcelona, Barcelona.
- Pérez D. B., C. (2012). Habilidades para la vida y consumo de drogas en adolescentes escolarizados mexicanos. *Adicciones*, 24(2), 153-160. DOI:10.20882/adicciones.108
- Piña L., J. A., & Rivera I., B. M. (2007). Validación del cuestionario de apoyo social funcional en personas seropositivas al VIH del noroeste de México. *Ciencia Y Enfermería*, 13(2), 53-63. DOI:10.4067/s0717-95532007000200007
- Pook, M., Tuschen-Caffier, B., & Brähler, E. (2008). Evaluation and comparison of different versions of the Body Shape Questionnaire. *Psychiatry Research*, 158(1), 67-73. DOI:10.1016/j.psychres.2006.08.002
- Prado M., C., Fernández D. O., R., & Anuncibay H., J. (2007). Evaluación de la calidad de la dieta y su relación con el estatus nutricional en niños y adolescentes de 9 a 15 años de la ciudad de Madrid. *Antropo*, 14, 61-73.
- R Development Core Team. (2013). R: The R Project for Statistical Computing. Vienna, Austria.

- Raich, R. M., Deus, J., Muñoz, M. J., Pérez, O., & Requena, A. (1991). Evaluación de la preocupación por la figura en una muestra de adolescentes catalanas. *Revista de Psiquiatría de La Facultad de Medicina de Barcelona*, 5, 210-220.
- Raich, R. M., Mora, M., Soler, A., Avila, C., Clos, I., & Zapater, L. (1996). Adaptación de un instrumento de evaluación de la insatisfacción corporal. *Clínica Y Salud*, 1(7), 51-66.
- Ramos H., S. (2015). *TAID versus CHAID Búsqueda de perfiles de mujeres trabajadoras en el servicio doméstico* (Trabajo de fin de Máster). Universidad de Salamanca, Salamanca, España.
- Räty, L. K. A., Larsson, G., Söderfeldt, B. A., & Larsson, B. M. W. (2005). Psychosocial aspects of health in adolescence: the influence of gender, and general self-concept. *The Journal of Adolescent Health*, 36(6), 530.
- Ravens-Sieberer, U., Erhart, M., Rajmil, L., Herdman, M., Auquier, P., Bruil, J., Power, M., Duer, W., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Haquist, C., Kilroe, J., & European KIDSCREEN Group. (2010). Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Quality of Life Research*, 19(10), 1487-1500. DOI:10.1007/s11136-010-9706-5
- Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Duer, W., Auquier, P., Power, M., Abel, T., Czemy, L., Mazur, J., Czimbalmos, A., Tountas, Y., Haquist, C., Kilroe, J., & KIDSCREEN Group, E. (2005). KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Review of Pharmacoeconomics & Outcomes Research*, 5(3), 353-364. DOI:10.1586/14737167.5.3.353
- Real Academia Española. Diccionario Usual. (2001a). Adolescencia. Recuperado de <http://bit.ly/2umW1Rw>

- Real Academia Española. Diccionario Usual. (2001b). Etiología. Recuperado de <http://bit.ly/2umWzHo>
- Real Academia Española. Diccionario Usual. (2001c). Entropía. Recuperado de <http://bit.ly/2spFuLr>
- Regidor, E. (2001). La clasificación de clase social de Goldthorpe: Marco de referencia para la propuesta de medición de la clase social del grupo de trabajo de la Sociedad Española de Epidemiología. *Revista Española de Salud Pública*, 75(1), 13-22.
- Rocha, K. B., Pérez, K., Rodríguez-Sanz, M., Borrell, C., & Obiols, J. E. (2011). Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española. *International Journal of Clinical and Health Psychology*, 11(1), 125-139.
- Rodrigo, M. J., Máiquez, M. L., García, M., Mendoza, R., Rubio, A., Martínez, A., & Martín, J. C. (2004). Relaciones padres-hijos y estilos de vida en la adolescencia. *Psicothema*, 16(2), 203-210.
- Rodríguez M., A., Pinzón P., S., Mallada G., E., Amaya O., F., Sánchez Z., J., & Máiquez P., A. (2004). ¿Tienen apoyo social y familiar los drogodependientes que participan en el programa “libre de drogas” en prisión? *Medicina de Familia*, 5(1), 16-21.
- Rodríguez R., F. J., Espinoza O., L. R., Gálvez C., J., Macmillan K., N. G., & Solis U., P. (2013). Estado nutricional y estilos de vida en estudiantes universitarios de la Pontificia Universidad Católica de Valparaíso. *Universidad Y Salud*, 15(2), 123-135.
- Rubio-Hurtado, M.-J., & Vilà-Baños, R. (2017). El análisis de conglomerados bietápico o en dos fases con SPSS. *Revista d’Innovació I Recerca En Educació*, 10(1), 118-126. DOI:10.1344/reire2017//

- Rueff, B., Crnac, J., & Darne, B. (1989). Detection of alcoholic patients using the systematic CAGE autoquestionnaire. In out patients. *Presse Medicale*, 18(33), 1654-1656.
- Ruiz-Risueño A., J., Ruiz-Juan, F., & Zamarripa R., J. I. (2012). Alcohol y tabaco en adolescentes españoles y mexicanos y su relación con la actividad físico-deportiva y la familia. *Rev Panam Salud Publica*, 31(3), 211-220.
- Salazar, C. M., Juárez L., R., Andrade S., A. I., Peña V, C. S., Arrellano C., A. C., & Hernández M., J. A. (2016). Percepción del beneficio de los deportes y actividades recreativas en habilidades para la vida en niños y adolescentes de Ciudad Juárez, México. *Sportis. Scientific Journal of School Sport, Physical Education and Psychomotricity*, 2(3), 356-378. DOI:10.17979/sportis.2016.2.3.1524
- Salazar C., C. M. (2012). *Estilos y calidad de vida en jóvenes preuniversitarios del estado de Colima* (Tesis doctoral). Universidad de Extremadura, Badajoz, España.
- Salazar C., C. M., Manzo L., E. G., & Vargas E., M. G. (2010). Juventud y obesidad. Un acercamiento a la noción de cuerpo en jóvenes universitarios de Colima. En Salazar C., C. M., Gómez F., J. A., Del Río V., J. E., Hernández L., S., Velasco L., M. J., Wong M., S., Quintana R., A., Quintana R., L., & Manzo L., E. G. (Eds.), *Usos y representaciones de las prácticas físicas-deportivas de los jóvenes mexicanos* (1st ed., pp. 10-26). Guadalajara: Universidad de Colima y Universidad Veracruzana.
- Sánchez-Cervantes, F. S., Serrano-González, R. E., & Márquez-Caraveo, M. E. (2015). Suicidios en menores de 20 años. México 1998-2011. *Salud Mental*, 38(5), 379-389. DOI:10.17711/SM.0185-3325.2015.051
- Santos-Preciado, J. I., Villa-Barragán, J. P., García-Avilés, M. A., León-Álvarez, G., Quezada-Bolaños, S., & Tapia-Conyer, R. (2003). La transición epidemiológica de las y los adolescentes en México. *Salud Pública de México*, 45(Supl 1), S140-S152. DOI:10.1590/S0036-36342003000700018

- Santos M., C. (2015). "Two-step Cluster" en SPSS y técnicas relacionadas (Trabajo de fin de Máster). Universidad de Salamanca, Salamanca, España.
- Seixas, F. A., Washburn, S., & Eisen, S. V. (1988). Alcoholism, Alcoholics Anonymous attendance, and outcome in a prison system. *The American Journal of Drug and Alcohol Abuse*, 14(4), 515-524.
- Sepúlveda, R. (2004). *Contribuciones al análisis de clases latentes en presencia de dependencia local* (Tesis doctoral). Universidad de Salamanca, Salamanca, España.
- Serra-Majem, L., Ribas, L., García, A., Pérez-Rodrigo, C., & Aranceta, J. (2003). Nutrient adequacy and Mediterranean Diet in Spanish school children and adolescents. *European Journal of Clinical Nutrition*, 57(Suppl 1), S35-S39. DOI:10.1038/sj.ejcn.1601812
- Serra-Majem, L., Ribas, L., Ngo, J., Ortega, R. M., García, A., Pérez-Rodrigo, C., & Aranceta, J. (2004). Food, youth and the Mediterranean diet in Spain. Development of KIDMED, Mediterranean Diet Quality Index in children and adolescents. *Public Health Nutrition*, 7(7), 931-935.
- Serra-Majem, L., Ribas, L., Ngo, J., Ortega, R. M., Pérez-Rodrigo, C., & Aranceta, J. (2002). Alimentación, jóvenes y dieta mediterránea en España. Desarrollo del KIDMED, índice de calidad de la dieta mediterránea en la infancia y la adolescencia. En Serra-Majem, L., & Aranceta, J. (Eds.), *Alimentación infantil y juvenil. Estudio enKid* (p. 51). Barcelona: Masson.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.
- Shevlin, M., & Adamson, G. (2005). Alternative Factor Models and Factorial Invariance of the GHQ-12: A Large Sample Analysis Using Confirmatory Factor Analysis. *Psychological Assessment*, 17(2), 231-236. DOI:10.1037/1040-3590.17.2.231

- Siciliano, R., & Mola, F. (1998). Ternary Classification Trees: A Factorial Approach. En Blasius, J., & Greenacre, M. J. (Eds.), *Visualization of Categorical Data* (1st ed., pp. 311-323). London: Academic Press.
- Siddiqi N. (2006). *Credit risk scorecards. Developing and implementing intelligent credit scoring*. New Jersey: John Wiley & Sons, Inc.
- Sigmund, E., Sigmundová, D., Šnoblová, R., & Gecková, A. M. (2014). ActiTrainer-determined segmented moderate-to-vigorous physical activity patterns among normal-weight and overweight-to-obese Czech schoolchildren. *European Journal of Pediatrics*, 173(3), 321-329. DOI:10.1007/s00431-013-2158-5
- Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society*, 13(2), 238-241.
- Smilkstein, G. (1978). The family APGAR: a proposal for a family function test and its use by physicians. *The Journal of Family Practice*, 6(6), 1231-1239.
- Smilkstein, G., Ashworth, C., & Montano, D. (1982). Validity and reliability of the family APGAR as a test of family function. *The Journal of Family Practice*, 15(2), 303-311.
- Smith, W. R. (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, 21(1), 3. DOI:10.2307/1247695
- Snedecor, G. W. (1938). *Statistical methods applied to experiments in agriculture and biology*. Ames, Iowa: Collegiate Press, Inc.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201. DOI:10.2307/1412107

- Stewart, G. W. (1999). Four algorithms for the the efficient computation of truncated pivoted QR approximations to a sparse matrix. *Numerische Mathematik*, 83(2), 313-323. DOI:10.1007/s002110050451
- Stunkard, A. J., & Stellar, E. (1990). Eating and its disorders. En Cash, T., & Pruzinsky, T. (Eds.), *Body Images* (pp. 3-20). Nueva York: Guilford Press.
- Suarez C., M. A., & Alcalá E., M. (2014). APGAR familiar: una herramienta para detectar disfunción familiar. *Revista Médica La Paz*, 20(1), 53-57.
- Theodoridis, S., & Koutroumbas, K. (1999). *Pattern recognition* (1st ed.). San Diego: Elsevier.
- Thoits, P. A. (1982). Conceptual, Methodological, and Theoretical Problems in Studying Social Support as a Buffer Against Life Stress. *Journal of Health and Social Behavior*, 23(2), 145-159. DOI:10.2307/2136511
- Trejo R., M. (2015). *Distribución de la grasa corporal y su relación con la percepción e insatisfacción de la imagen corporal en mujeres de 19 a 25 años de edad* (Tesis de maestría). Universidad Autónoma de Nuevo León, San Nicolás de los Garza, México.
- Trujano R., P., De García B., M., Nava Q., C., & Limón A., G. R. (2014). Sobrepeso y obesidad en preadolescentes mexicanos: estudio descriptivo, variables correlacionadas y directrices de prevención. *Acta Colombiana de Psicología*, 17(1), 119-130. DOI:10.14718/ACP.2014.17.1.12
- Trujillo-Hernández, B., Vásquez, C., Almanza-Silva, J. R., Jaramillo-Virgen, M. E., Mellin-Landa, T. E., Valle-Figueroa, O. B., Pérez-Ayala, R., Millán-Guerrero, R. O., Prieto-Díaz-Chávez, E., & Newton-Sánchez, O. (2010). Frecuencia y factores de riesgo asociados a sobrepeso y obesidad en universitarios de Colima, México. *Rev. Salud Pública*, 12(2), 197-207.
- United Nations Office on Drugs and Crime (UNODC). (2015). *World Drug Report*. Nueva York: United Nations publications.

- Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2007). Experimental perspectives on learning from imbalanced data. En *Proceedings of the 24th international conference on Machine learning - ICML '07* (pp. 935-942). New York, New York, USA: ACM Press. DOI:10.1145/1273496.1273614
- Vázquez A., R., Galán, J., López A., X., Alvarez R., G. L., Mancilla-Díaz, J. M., Caballero R., A., & Unikel S., C. (2011). Validez del Body Shape Questionnaire (BSQ) en mujeres mexicanas. *Revista Mexicana de Trastornos Alimentarios/Mexican Journal of Eating Disorders*, 2(1), 42-52.
- Verdugo-Lucero, J. C., Ponce de León-Pagaza, B. G., Guardado-Llamas, R. E., Meda-Lara, R. M., Uribe-Alvarado, J. I., & Guzmán-Muñiz, J. (2013). Estilos de afrontamiento al estrés y bienestar subjetivo en adolescentes y jóvenes. *Revista Latinoamericana de Ciencias Sociales, Niñez Y Juventud*, 11(1), 79-91. DOI:10.11600/1692715x.1114120312
- Vermunt, J. K., & Magidson, J. (2005). *Latent GOLD ® Choice 4.0 User's Manual*. Belmont: Statistical Innovations, Inc.
- Villa G, I. C., Zuluaga A., C., & Restrepo R., L. (2013). Propiedades psicométricas del Cuestionario de Salud General de Goldberg GHQ-12 en una institución hospitalaria de la ciudad de Medellín. *Avances En Psicología Latinoamericana*, 31(3), 532-545.
- Wilkinson, R., & Marmot, M. (Eds.). (2003). *Social determinants of health: the solid facts*. (2nd ed.). Copenhagen: WHO.
- Yang, J., Rübél, O., Prabhat, Mahoney, M. W., & Bowen, B. P. (2015). Identifying Important Ions and Positions in Mass Spectrometry Imaging Data Using CUR Matrix Decompositions. *Analytical Chemistry*, 87(9), 4658-4666. DOI:10.1021/ac5040264
- Yen, S. J., & Lee, Y. S. (2009). Cluster-based under-sampling approaches for imbalanced data distributions. *Expert Systems with Applications*, 36(3), 5718-5727. DOI:10.1016/j.eswa.2008.06.108

- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM Sigmod Record*, 25(2), 103-114.
- Zhang, T., Ramakrishnan, R., & Livny, M. (1997). BIRCH: A New Data Clustering Algorithm and Its Applications. *Data Mining and Knowledge Discovery*, 1(2), 141-182. DOI:10.1023/A:1009783824328
- Zhao, H., Chen, X., Nguyen, T., Huang, J. Z., Williams, G., & Chen, H. (2016). Stratified Over-Sampling Bagging Method for Random Forests on Imbalanced Data. En Chau, M., Wang, A., & Chen, H. (Eds.), *Pacific-Asia Workshop on Intelligence and Security Informatics* (pp. 63-72). Springer, Cham. DOI:10.1007/978-3-319-31863-9\_5
- Zheng, Z., & Srihari, R. (2003). Optimally Combining Positive and Negative Features for Text Categorization. *Workshop on Learning from Imbalanced Datasets II*.

---

# ANEXO

---

## 1.1 Instrumento de estudio

Folio:

Fecha: \_\_\_/\_\_\_/\_\_\_

**Instrucciones:** Lee atentamente cada una de las preguntas y marca con una equis la opción que consideres adecuada. Recuerda que no hay respuestas correctas e incorrectas, que el cuestionario es anónimo por lo que no podrás ser identificado y tus respuestas serán utilizadas exclusivamente para tareas de investigación científica por lo que te pedimos respondas honestamente. Por favor evita dejar preguntas sin responder.

### I. Datos sociodemográficos

I.1. Género:                      I.2. Edad: \_\_\_\_\_ I.3. Peso: \_\_\_\_\_ I.4. Talla: \_\_\_\_\_

Hombre

Mujer

I.5 ¿Cuál es la profesión del principal sustentador del hogar? (Si no trabaja, pon una línea)

\_\_\_\_\_

### II. Hábitos de vida

#### II.1. KIDMED

	Sí	No
1. ¿Tomas una fruta o jugo de frutas cada día?	<input type="checkbox"/>	<input type="checkbox"/>
2. ¿Tomas una segunda de fruta cada día?	<input type="checkbox"/>	<input type="checkbox"/>
3. ¿Tomas verduras/hortalizas frescas (ensaladas) o cocinadas regularmente una vez al día?	<input type="checkbox"/>	<input type="checkbox"/>
4. ¿Tomas verduras frescas o cocinadas más de una vez al día?	<input type="checkbox"/>	<input type="checkbox"/>
5. ¿Tomas pescado con regularidad (al menos 2 o 3 veces por semana)?	<input type="checkbox"/>	<input type="checkbox"/>
6. ¿Acudes más de una vez a la semana a un restaurante de comida rápida (hamburguesas, tortas, tacos, etcétera)?	<input type="checkbox"/>	<input type="checkbox"/>
7. ¿Te gustan las legumbres y las comes más de una vez a la semana (frijol, habas, ejotes, lentejas, garbanzos, cacahuates, etcétera)?	<input type="checkbox"/>	<input type="checkbox"/>

	Sí	No
8. ¿Tomas pasta, arroz casi a diario (5 días o más a la semana)?	<input type="checkbox"/>	<input type="checkbox"/>
9. ¿Desayunas cereales o derivados (pan, etcétera)	<input type="checkbox"/>	<input type="checkbox"/>
10. ¿Tomas frutos secos con regularidad (al menos 2 o 3 veces por semana)?	<input type="checkbox"/>	<input type="checkbox"/>
11. ¿Utilizas aceite de oliva en casa?	<input type="checkbox"/>	<input type="checkbox"/>
12. ¿Te saltas el desayuno?	<input type="checkbox"/>	<input type="checkbox"/>
13. ¿Desayunas productos lácteos (leche, yogurt, etcétera)?	<input type="checkbox"/>	<input type="checkbox"/>
14. ¿Desayunas/comes bollería industrial o comercial (pasteles, galletas, etcétera)?	<input type="checkbox"/>	<input type="checkbox"/>
15. ¿Tomas dos yogures y/o un poco de queso (40g) cada día?	<input type="checkbox"/>	<input type="checkbox"/>
16. ¿Tomas dulces y golosinas varias veces al día?	<input type="checkbox"/>	<input type="checkbox"/>

## II.2. IPAQ-Corto

*Por favor responde a cada pregunta aún si no te consideras una persona activa. Piensa en aquellas actividades que haces como parte de la escuela, el trabajo, en el jardín, en la casa, para ir de un lugar a otro o en tus tiempos libres, ejercicio o deporte.*

<p>Piensa acerca de todas aquellas actividades <b>VIGOROSAS</b> que realizaste en los últimos 7 días. Actividades vigorosas son las que requieren un esfuerzo físico fuerte y le hacen respirar mucho más fuerte que lo normal. Piensa solamente en esas actividades que hiciste por lo menos 10 minutos continuos.</p>		
<p>1. Durante los últimos 7 días ¿cuántos días realizaste actividades físicas vigorosas como levantar objetos pesados, excavar, aeróbicos o pedalear rápido en bicicleta?</p>		
<input type="text"/> días por semana	<input type="checkbox"/> Ninguna actividad física vigorosa → pasa a la pregunta 3	
<p>2. ¿Cuánto tiempo en total usualmente te tomó realizar actividades físicas vigorosas en uno de esos días que las realizaste?</p>		
<input type="text"/> horas por día	<input type="text"/> minutos por día	<input type="checkbox"/> No sé/No estoy seguro(a)

Piensa acerca de todas aquellas actividades **MODERADAS** que realizaste en los últimos 7 días. Actividades moderadas son las que requieren un esfuerzo físico moderado y le hacen respirar algo más fuerte que lo normal. Piensa solamente en esas actividades que hiciste por lo menos 10 minutos continuos.

3. Durante los últimos 7 días ¿cuántos días hiciste actividades físicas moderadas tal como cargar objetos livianos, pedalear en bicicleta a paso regular o jugar dobles de tenis? No incluyas caminatas.

_____ días por semana	<input type="checkbox"/> Ninguna actividad física moderada → pase a la pregunta 5
-----------------------	---

4. Usualmente ¿Cuánto tiempo dedicas en uno de esos días haciendo actividades físicas moderadas?

_____ horas por día	_____ minutos por día	<input type="checkbox"/> No sé/No estoy seguro(a)
---------------------	-----------------------	---

Piensa acerca del tiempo que dedicaste a **CAMINAR** en los últimos 7 días. Esto incluye trabajo en la casa, caminatas para ir de un sitio a otro, o cualquier otra caminata que hiciste únicamente por recreación, deporte, ejercicio o placer.

5. Durante los últimos 7 días ¿cuántos días caminaste por al menos 10 minutos continuos?

_____ días por semana	<input type="checkbox"/> Ninguna actividad física moderada → pase a la pregunta 7
-----------------------	---

6. Usualmente ¿cuánto tiempo pasaste en uno de esos días caminando?

_____ horas por día	_____ minutos por día	<input type="checkbox"/> No sé/No estoy seguro(a)
---------------------	-----------------------	---

La última pregunta se refiere al tiempo que permaneciste **SENTADO** en la semana en los últimos 7 días. Incluye el tiempo sentado(a) en la escuela, el trabajo, la casa y en tu tiempo libre. Esto puede incluir tiempo sentado(a) en un escritorio, visitando amigos(as), leyendo o permanecer sentado(a) o acostado(a) mirando televisión.

7. Durante los últimos 7 días ¿Cuánto tiempo permaneciste sentado(a) en un día a la semana?

_____ horas por día	_____ minutos por día	<input type="checkbox"/> No sé/No estoy seguro(a)
---------------------	-----------------------	---

### II.3. Test de Fagerström

1. ¿Fumas? (Se considera que fumas si al menos fumas un cigarrillo al día)

Sí

No

Ex-fumador

## 2. Si fumas contesta a estas preguntas:

2.1 ¿Cuánto tiempo pasa desde que te levantas hasta que fumas el primer cigarrillo?

5 min                       6-30 min                       31-60 min                       Más de 60 min

2.2 ¿Encuentras difícil no fumar en lugares donde está prohibido (bares, cine, escuela, biblioteca,...)?

Sí     No

2.3 ¿Qué cigarrillo te molesta más dejar de fumar?

El primero de la mañana                       Cualquier otro

2.4 ¿Cuántos cigarrillos fumas cada día?

10 o menos                       11 a 20                       21 a 30                       31 o más

2.5 ¿Fumas más durante las primeras horas al levantarte que el resto del día?

Sí     No

2.6 ¿Fumas aunque estés tan enfermo que tengas que guardar cama la mayor parte del día?

Sí     No

## II.4. CAGE

	Sí	No
1. ¿Piensas que comes demasiados dulces?	<input type="checkbox"/>	<input type="checkbox"/>
2. ¿Te han ofrecido alguna vez un “churro” o una dosis de cocaína?	<input type="checkbox"/>	<input type="checkbox"/>
3. ¿Te ha molestado alguna vez la gente que te critica tu forma de beber?	<input type="checkbox"/>	<input type="checkbox"/>
4. ¿Te has planteado en alguna ocasión hacer algo de ejercicio con carácter semanal?	<input type="checkbox"/>	<input type="checkbox"/>
5. ¿Consideras que duermes las horas necesarias para estar a tono durante el día?	<input type="checkbox"/>	<input type="checkbox"/>
6. ¿Has tenido alguna vez la impresión que deberías beber menos?	<input type="checkbox"/>	<input type="checkbox"/>
7. ¿Te has planteado seriamente en alguna ocasión dejar de fumar?	<input type="checkbox"/>	<input type="checkbox"/>
8. ¿Te han comentado alguna vez que deberías comer más frutas y verduras?	<input type="checkbox"/>	<input type="checkbox"/>
9. ¿Te has sentido alguna vez mal o culpable por tu costumbre de beber?	<input type="checkbox"/>	<input type="checkbox"/>

	Sí	No
10. ¿Te han comentado en alguna ocasión que deberías fumar menos?	<input type="checkbox"/>	<input type="checkbox"/>
11. ¿Alguna vez lo primero que has hecho por la mañana ha sido beber para calmar los nervios o para librarte de la resaca?	<input type="checkbox"/>	<input type="checkbox"/>
12. En caso de necesitarlas ¿cambiarías el consumo de pastillas para dormir por técnicas de relajación?	<input type="checkbox"/>	<input type="checkbox"/>

### III. Percepción

#### III.1. DUKE-UNC-11

	Mucho menos de lo que deseo	Menos de lo que deseo	Ni mucho ni poco	Casi como deseo	Tanto como deseo
1. Recibo visitas de mis amigos y familiares	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Recibo ayuda en asuntos relacionados con mi casa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Recibo elogios y reconocimientos cuando hago bien mi trabajo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Cuento con personas que se preocupan de lo que me sucede	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Recibo amor y afecto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Tengo la posibilidad de hablar con alguien de mis problemas en el trabajo o en la casa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Tengo la posibilidad de hablar con alguien de mis problemas personales y familiares	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Tengo la posibilidad de hablar con alguien de mis problemas económicos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Recibo invitaciones para distraerme y salir con otras personas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Recibo consejos útiles cuando me ocurre algún acontecimiento importante en mi vida	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Recibo ayuda cuando estoy enfermo en la cama	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## III.2. APGAR

	Casi nunca	A veces	Casi siempre
1. ¿Estás satisfecho con la ayuda que recibes de tu familia cuando tienes un problema?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. En tu familia ¿conversan entre ustedes los problemas que tienen en casa?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. ¿Las decisiones importantes se toman en conjunto en la casa?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. ¿Está satisfecho con el tiempo que tú y tu familia pasan juntos?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. ¿Sientes que tu familia te quiere?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## III.3. SWLS-C

	No	Más bien no	Ni sí, ni no	Más bien sí	Sí
1. En la mayoría de aspectos mi vida es como a mí me gustaría	<input type="checkbox"/>				
2. Las cosas en mi vida son excelentes	<input type="checkbox"/>				
3. Soy feliz con mi vida	<input type="checkbox"/>				
4. Hasta ahora he conseguido todas las cosas importantes que quiero en la vida	<input type="checkbox"/>				
5. Si pudiera vivir mi vida de nuevo, lo haría de la misma manera	<input type="checkbox"/>				

## III.4. KIDSCREEN-10

Durante los últimos 7 días:

	Nada	Un poco	Moderadamente	Mucho	Muchísimo
1. ¿Te has sentido bien y en forma?	<input type="checkbox"/>				
2. ¿Te has sentido lleno de energía?	<input type="checkbox"/>				
3. ¿Te has sentido triste?	<input type="checkbox"/>				
4. ¿Te has sentido solo(a)?	<input type="checkbox"/>				
5. ¿Has tenido suficiente tiempo para ti?	<input type="checkbox"/>				
6. ¿Has podido hacer las cosas que querías en tu tiempo libre?	<input type="checkbox"/>				

Durante los últimos 7 días:	Nada	Un poco	Moderadamente	Mucho	Muchísimo
7. ¿Tus padres te han tratado de forma justa?	<input type="checkbox"/>				
8. ¿Te has divertido con tus amigos(as)?	<input type="checkbox"/>				
9. ¿Te ha ido bien en la escuela? ( <i>Si no vas a clase en estos días, piensas en la última semana que has ido a la preparatoria</i> )	<input type="checkbox"/>				
10. ¿Has podido prestar atención?	<input type="checkbox"/>				

III.5. GHQ-12

1. ¿Has podido concentrarte bien en lo que haces?	<input type="checkbox"/> Mejor que lo habitual	<input type="checkbox"/> Igual que lo habitual	<input type="checkbox"/> Menos que lo habitual	<input type="checkbox"/> Mucho menos que lo habitual
2. ¿Has sido capaz de disfrutar de tus actividades normales de cada día?	<input type="checkbox"/> Más que lo habitual	<input type="checkbox"/> Igual que lo habitual	<input type="checkbox"/> Menos que lo habitual	<input type="checkbox"/> Mucho menos que lo habitual
3. ¿Tus preocupaciones te han hecho perder mucho sueño?	<input type="checkbox"/> Mejor que lo habitual	<input type="checkbox"/> Igual que lo habitual	<input type="checkbox"/> Menos que lo habitual	<input type="checkbox"/> Mucho menos que lo habitual
4. ¿Has sido capaz de hacer frente adecuadamente a tus problemas?	<input type="checkbox"/> Más capaz que lo habitual	<input type="checkbox"/> Igual de capaz que lo habitual	<input type="checkbox"/> Menos capaz que lo habitual	<input type="checkbox"/> Mucho menos capaz que lo habitual
5. ¿Has sentido que está desempeñando un papel útil en la vida?	<input type="checkbox"/> Más útil que lo habitual	<input type="checkbox"/> Igual de útil que lo habitual	<input type="checkbox"/> Menos útil que lo habitual	<input type="checkbox"/> Mucho menos útil que lo habitual
6. ¿Te has sentido poco feliz o deprimido?	<input type="checkbox"/> No, en absoluto	<input type="checkbox"/> No más que lo habitual	<input type="checkbox"/> Algo más que lo habitual	<input type="checkbox"/> Mucho más que lo habitual

7. ¿Te has sentido poco capaz de tomar decisiones?

Mejor que lo habitual     Igual que lo habitual     Menos que lo habitual     Mucho menos que lo habitual

8. ¿Has perdido confianza en ti mismo?

No, en absoluto     No más que lo habitual     Algo más que lo habitual     Mucho más que lo habitual

9. ¿Te has notado constantemente agobiado y en tensión?

No, en absoluto     No más que lo habitual     Algo más que lo habitual     Mucho más que lo habitual

10. ¿Has pensado que eres una persona que no vale para nada?

No, en absoluto     No más que lo habitual     Algo más que lo habitual     Mucho más que lo habitual

11. ¿Has tenido la sensación de que no puedes superar las dificultades?

No, en absoluto     No más que lo habitual     Algo más que lo habitual     Mucho más que lo habitual

12. ¿Te sientes razonablemente feliz considerando todas las circunstancias?

Mejor que lo habitual     Igual que lo habitual     Menos que lo habitual     Mucho menos que lo habitual

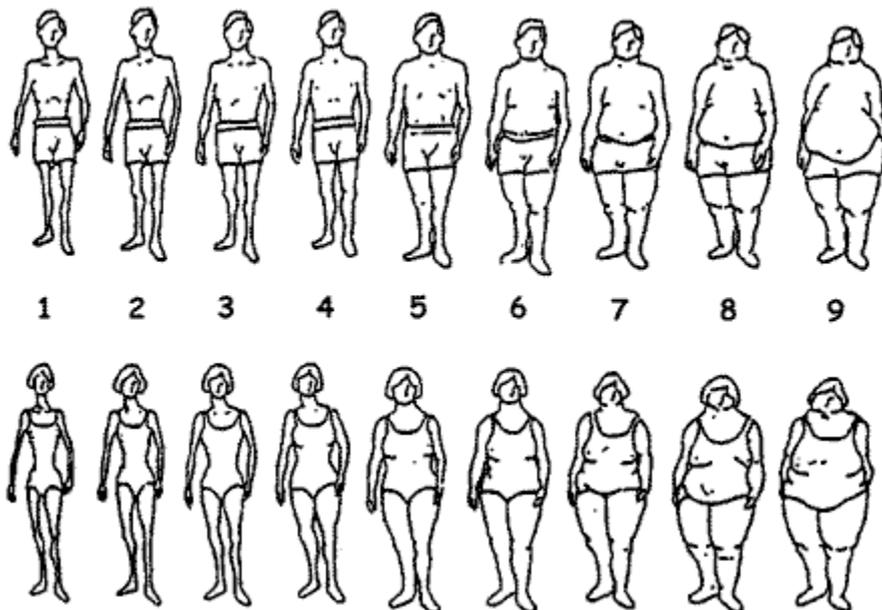
III.6. BSQ-8C

	Nunca	Raramente	A veces	A menudo	Muy a menudo
1. ¿Has tenido miedo a engordar?	<input type="checkbox"/>				
2. Sentirte lleno(a) (después de una gran comida) ¿te ha hecho sentirte gordo(a)?	<input type="checkbox"/>				
3. Pensar en tu figura ¿ha interferido en tu capacidad de concentración (cuando miras la televisión, lees o mantienes una conversación)?	<input type="checkbox"/>				
4. ¿Te has imaginado cortando partes gruesas de tu cuerpo?	<input type="checkbox"/>				
5. ¿Te has sentido excesivamente gordo(a) o redondeado(a)?	<input type="checkbox"/>				

	Nunca	Raramente	A veces	A menudo	Muy a menudo
6. ¿Has pensado que la figura que tienes es debida a tu falta de autocontrol?	<input type="checkbox"/>				
7. Verte relegado(a) en un espejo o en un escaparate ¿te ha hecho sentirte mal por tu figura?	<input type="checkbox"/>				
8. ¿Te has fijado más en tu figura estando en compañía de otras personas?	<input type="checkbox"/>				

III.7. Figuras de Stunkard

Mira bien las siguientes figuras y contesta las siguientes preguntas



1. Realmente ¿a cuál de las figuras te pareces? Pon el número: \_\_\_\_\_

2. ¿A cuál de las figuras te gustaría parecer? Pon el número: \_\_\_\_\_

## 1.2 Formas de evaluar el instrumento

En el capítulo 4 de esta tesis se presentan los instrumentos que se utilizaron para llevar a cabo la investigación. Para cada uno de ellos se detalla su origen, objetivo, así como forma de medición y puntos de corte. En este apartado se presentan de forma sintetizada las clasificaciones a partir de los resultados.

### Peso y talla

El Índice de Masa Corporal (IMC) se calcula a través de dividir el peso entre la altura elevada al cuadrado. El resultado obtenido se clasifica según la tabla A. Para más detalle ir a la página 27.

$$IMC = \frac{\text{peso}}{(\text{altura})^2}$$

Tabla A. Puntos de corte internacional para el IMC por sexo

Edad	IMC: 25kg/m <sup>2</sup>		IMC: 30kg/m <sup>2</sup>	
	Hombres	Mujeres	Hombres	Mujeres
15	23.29	23.94	28.30	29.11
16	23.90	24.37	28.88	29.43
17	24.46	24.70	29.41	29.69
18	25	25	30	30

Fuente: Elaboración propia con base en Cole et al. (2000).

## Esquema de Goldthorpe

De acuerdo la ocupación a la que se dedique el principal proveedor del hogar, puede clasificarse según la tabla B. Para más detalle ir a la página 31.

Tabla B. Clasificación de las clases sociales según el Esquema de Goldthorpe

Clase elitaria	Clasificación	Significado
Clases de servicio	I	Profesionales superiores; directivos de grandes establecimientos y grandes empleadores (más de 25 empleados).
	II	Profesionales de nivel medio e inferior; técnicos superiores; directivos de pequeños establecimientos (menos de 25 empleados); supervisores de empleados no manuales.
Clases intermedias	III a	Empleados no manuales de rutina en la administración y el comercio.
	III b	Trabajadores de servicios personales y de seguridad.
	IV a	Pequeños propietarios, artesanos, etc., con empleados (menos de 25).
	IV b	Pequeños propietarios, artesanos, sin empleados.
	IV c	Agricultores, pescadores, etc.
	V	Supervisores de trabajadores manuales, técnicos de nivel inferior etc.
Clases obreras	VI	Trabajadores manuales cualificados.
	VII a	Trabajadores semicualificados y sin cualificar no agrarios.
	VII b	Trabajadores agrarios.

Fuente: Elaboración propia con base en (Goldthorpe, 1997).

## KIDMED

Los ítems 6, 12, 14 y 16 tienen connotación negativa por lo que se les asigna un valor de -1. Al resto de los ítems se les asigna un valor de +1. Las sumas de los valores de la prueba se clasificaron en tres niveles (Serra-Majem et al., 2004) o cuatro (Prado et al., 2007) como puede observarse en la tabla C. Para más detalle ir a la página 33.

Tabla C. Puntos de corte para la clasificación de la dieta según el índice de KIDMED

Autor	Puntuación	Clasificación
Serra-Majem et al. (2004)	$\geq 8$	Dieta óptima
	Entre 4 y 7	Es necesario mejorar y ajustar
	$\leq 3$	Muy baja calidad
Prado et al. (2007)	$\geq 8$	Óptima
	6 y 7	Media/alta
	4 y 5	Media/baja
	$\leq 3$	Muy baja

Fuente: Elaboración propia con base en Serra-Majem et al. (2004) y Prado et al. (2007).

## IPAQ-Corto

El protocolo para calificar las respuestas al IPAQ utilizan la referencia de Ainsworth et al. (2000)<sup>25</sup> e indican que caminar tiene una tasa metabólica de 3.3 METs; la actividad física moderada, 4 METs; y la actividad física vigorosa, 8 METs. Teniendo en cuenta lo anterior debe multiplicarse el valor en METs de la actividad realizada por la frecuencia y la duración de la actividad para obtener el índice de actividad física y después clasificarlo en alguna de las tres

<sup>25</sup> Las actividades en el Compendio de Actividades Físicas van desde 0.9 METs (por dormir) hasta 18 METs (por hacer una carrera a 17.5 Km/h) (Ainsworth et al., 2000).

categorías (IPAQ Research Committee, 2005) (ver tabla D). Para más detalle ir a la página 35.

Tabla D. Clasificaciones a las respuestas del IPAQ-Corto

Categoría	Criterio
Bajo	Aquellas personas que no cumplen con los criterios para las categorías 2 o 3 se consideran que tienen un nivel de actividad física "bajo".
Moderado <sup>26</sup>	<ul style="list-style-type: none"> <li>d) 3 o más días de actividad vigorosa de al menos 20 minutos al día.</li> <li>e) 5 o más días de actividad de intensidad moderada y/o caminar al menos 30 minutos por día.</li> <li>f) 5 o más días de cualquier combinación de actividades de caminata, de intensidad moderada o intensa, alcanzando un mínimo de actividad física total de al menos 600 MET-min/semana.</li> </ul>
Elevado <sup>26</sup>	<ul style="list-style-type: none"> <li>c) Actividad de intensidad vigorosa en al menos 3 días logrando al menos 1500 MET-minutos / semana.</li> <li>d) 7 o más días de cualquier combinación entre caminar y actividades de intensidad moderada o de intensidad vigorosa, logrando un mínimo de actividad física total de al menos 3000 MET-min/semana.</li> </ul>

Fuente: Elaboración propia con base en IPAQ Research Committee (2005).

Otra forma de evaluar el IPAQ-Corto es a través del algoritmo presente en la figura 2 de la página 39.

## Test de Fagerström

Para evaluar la adicción a la nicotina, cada pregunta del FTND tiene un puntaje distinto las cuales se detallan en la tabla E. Para más detalles del test ir a la página 39.

Tabla E. Puntaje otorgado a las respuestas del FTND

<sup>26</sup> Si cumple al menos uno de los criterios.

Pregunta	Respuesta	Puntaje
2.1	5 min	3
	6-30 min	2
	31-60 min	1
	Más de 60 min	0
2.2	Sí	1
	No	0
2.3	El primero de la mañana	1
	Cualquier otro	0
2.4	10 o menos	0
	11 a 20	1
	21 a 30	2
	31 o más	3
2.5	Sí	1
	No	0
2.6	Sí	1
	No	0

Fuente: Elaboración propia con base en Fagerström et al. (1996).

La suma de las respuestas otorgadas puede clasificarse, según Fagerström et al. (1996) según la tabla F.

Tabla F. Valores del FTND

Puntaje	Clasificación
0 a 3 puntos	Dependencia leve
Entre 4 y 6 puntos	Dependencia moderada
7 puntos o más	Dependencia intensa

Fuente: Elaboración propia con base en Fagerström et al. (1996).

## CAGE

El número de respuestas positivas a cualquiera de estas preguntas se suman y se clasifican según la tabla G. Para más detalle ir a la página 41.

3. ¿Te ha molestado alguna vez la gente que te critica tu forma de beber?
6. ¿Has tenido alguna vez la impresión que deberías beber menos?
9. ¿Te has sentido alguna vez mal o culpable por tu costumbre de beber?
11. ¿Alguna vez lo primero que has hecho por la mañana ha sido beber para calmar los nervios o para librarte de la resaca?

Tabla G. Dictamen a las respuestas del test de CAGE

Núm. de respuestas positivas	Clasificación
0 a 1	Consumo social o bebedor social
2	Consumo de riesgo
3	Consumo perjudicial
4	Dependencia alcohólica

Fuente: Elaboración propia con base en (Muracén, Martínez, Aguilar & González, 2001).

## DUKE-UNC-11

Cada ítem tiene cinco opciones de respuesta tipo Likert, que va desde 1 (menos de lo que quisiera) hasta 5 (tanto como me gustaría). La suma de los puntos obtenidos se clasifica según la tabla H. Para más detalle ir a la página 29.

Tabla H. Puntos de corte para el DUKE-UNC-11

Puntaje	Clasificación
≥33 puntos	Apoyo Social Funcional
<33 puntos	Escaso Apoyo Social

Fuente: Elaboración propia con base en Bellón et al. (1996a).

## APGAR-Familiar

Cada ítem tiene tres opciones de respuesta que va de 0 a 2, en la que 0 equivale a “casi nunca”, 1 a “a veces” y 2 a “casi siempre”. La suma de los puntos obtenidos se clasifica según la tabla I. Para más detalle ir a la página 43.

Tabla I. Valores del Test APGAR-Familiar

Autor	Escala	Puntaje	Clasificación
Smilkstein (1978)	0 a 2	0 a 3 puntos	Disfunción grave
		4 a 6 puntos	Disfunción leve
		7 a 10 puntos	Normofuncional

Fuente: Elaboración propia con base en Smilkstein (1978).

## SWLS-C

Cada ítem tiene cinco opciones de respuesta tipo Likert, que va desde 1 (No) hasta 5 (Sí). La suma de los puntos obtenidos se clasifica según la tabla J. Para más detalle ir a la página 46.

Tabla J. Puntos de corte para la SWLS-C

Puntuación	Clasificación
21 a 25	Satisfecho
16 a 20	Ligeramente satisfecho
11 a 15	Satisfacción media
6 a 10	Ligeramente insatisfecho
0 a 5	Insatisfecho

Fuente: Elaboración propia con base en Atienza et al. (2000).

## KIDSCREEN-10

Cada ítem tiene cinco opciones de respuesta tipo Likert. Para hacer la comparación con países europeos, de acuerdo con Salazar (2012) se utiliza una escala de 0 a 100 puntos (nunca = 20; un poco = 40; moderadamente = 60; mucho = 80; y, muchísimo = 100), por lo que, una calificación arriba de 50 puntos significa alta CVRS como puede verse en la tabla K. Para más detalle ir a la página 48.

Tabla K. Puntos de corte para el KIDSCREEN-10

Puntaje	Clasificación
≥50 puntos	Alta CVRS
<50 puntos	Baja CVRS

Fuente: Elaboración propia con base en Kidscreen Group.

## GHQ-12

Cada ítem tiene cuatro opciones de respuesta tipo Likert, la suma de los puntos obtenidos se clasifica según la forma de puntuar como se presenta en la tabla L. Para más detalle ir a la página 51.

Tabla L. Puntos de corte para el GHQ-12

Autor	Forma de puntuar	Significado
Goldberg et al. (1997)	Método usual	≤ 2 ausencia de psicopatología
	(0-0-1-1)	> 2 presencia de psicopatología
	C-GHQ	≤ 5 ausencia de psicopatología
	(0-1-1-1)	> 5 presencia de psicopatología
	Likert	≤ 12 ausencia de psicopatología
	(0-1-2-3)	> 12 presencia de psicopatología

Fuente: Elaboración propia con base en Goldberg et al. (1997).

## BSQ-8C

Cada ítem tiene cinco opciones de respuesta tipo Likert, que va desde 1 (Nunca) hasta 5 (Muy a menudo). La suma de los puntos obtenidos se clasifica según la tabla M. Para más detalle ir a la página 54.

Tabla M. Valores del Test BSQ-8

Puntos	Clasificación
8	No preocupado / Satisfacción corporal
9 a 22	Leve preocupación / Miedo a engordar
23 a 28	Moderada preocupación / Baja autoestima
29 a 40	Extrema preocupación Deseo de perder peso

Fuente: Elaboración propia con base en López, Solé & Cortés (2008).

## Figuras de Stunkard

La diferencia entre la talla actual y la que considera ideal es llamado el índice de discrepancia y es considerado para representar la insatisfacción corporal. Este se puede clasificar según la tabla N. Para más detalle ir a la página 57.

Tabla N. Clasificación del índice de discrepancia

Índice de discrepancia	Clasificación
Diferencia nula	Satisfacción con la imagen corporal.
Diferencia positiva	Insatisfacción en cuanto al deseo de una figura más delgada.
Diferencia negativa	Insatisfacción en cuanto al deseo de una figura más gruesa.

Fuente: García (2015).