
Propuesta de índices multidimensionales para la evaluación del programa “Ser Pilo Paga”

Proposal of multidimensional indexes for the evaluation of the program “Ser Pilo Paga”

Edwin Leonardo Buitrago Chinchilla^a
edwinbuitrago@usantotomas.edu.co

José Giovany Babativa Márquez^b
josebabativa@usantotomas.edu.co

Resumen

En diferentes estudios las unidades estadísticas se describen en dos tipos de variables: cuantitativas y cualitativas. Un ejemplo común donde se observa este tipo de variables son los instrumentos de recolección de información conocidos como encuestas, por lo general, en estos instrumentos existen numerosas preguntas cerradas que suelen ser reunidas en varios conjuntos que pueden ser de naturaleza cuantitativa o cualitativa (Bécue-Bertaut & Pagès 2008). Una técnica multivariada que analiza simultáneamente este tipo de variables es el análisis factorial para datos mixtos (AFDM), esta metodología consiste en transformar las variables cuantitativas en cualitativas, desglosando su intervalo de variación en clases produciendo una tabla homogénea resultante que permite implementar un análisis de correspondencia múltiple (ACM) (Pagès 2014), sin embargo, la noción de agrupación de variables no es una propiedad del AFDM. Si el instrumento cumple esta característica (la agrupación de variables en los sujetos del análisis) se considera necesario hacer uso del análisis factorial múltiple (AFM) ya que este permite analizar grupos de variables cuantitativas y/o cualitativas siendo ampliamente útil cuando el número de variables de cada grupo es muy diferente (Escofier & Pagès 2008).

Extender el AFM a grupos de variables cualitativas o mixtas (de ambas naturalezas) se conoce como un análisis factorial múltiple para datos mixtos (AFMDM), el cual combina tanto el ACM, el AFDM y el AFM, esta técnica resalta los principales factores de variabilidad de los individuos, descritos estos, de manera equilibrada por los grupos de variables. Aplicando el AFMDM al instrumento de la línea base del programa Ser Pilo Paga, se crearon cuatro índices multidimensionales que resumen 9 grupos temáticos compuestos de 111 variables de tipo cuantitativo y cualitativo, caracterizando un total de 1.487 jóvenes, los cuales 682 son elegibles para hacer parte del programa y 805 no lo son. Se propone para la evaluación de impacto, un estimador de diferencia en diferencia aplicado a los cuatro índices multidimensionales.

Palabras clave: Análisis factorial múltiple para datos mixtos, Ser Pilo Paga, Análisis multivariado de datos, Regresión discontinua, Línea base, Índices multidimensionales.

Abstract

In different studies statistical units are described in two types of variables: quantitative and qualitative. A common example where these types of variables are observed are the information collection instruments known as surveys, in general, in these instruments there are closed questions that are usually gathered together in sets that may be of a quantitative or qualitative nature (Bécue-Bertaut & Pagès 2008). A multivariate technique that analyzes this type of variables is the factorial analysis for mixed data (FAMD), this methodology consist in transform the quantitative variables into qualitative variables,

^aEstudiante de la facultad de Estadística. Universidad Santo Tomás

^bDocente de la facultad de Estadística. Universidad Santo Tomás

breaking down the range of variation in the classes producing a homogeneous table that allows you to implement an analysis Multiple Correspondence (MCA) (Pagès 2014), however, the notion of grouping variables is not an FAMD property. If the instrument complies with this characteristic (the grouping of variables in the individuals of the analysis) it is considered necessary to use multiple factorial analysis (MFA). Because this kind of analysis it allows to analyze quantitative and qualitative variables being broadly useful when the number of variables from each group is very different (Escofier & Pagès 2008).

Extending the MFA to groups of qualitative or mixed variables (of both natures) is known as a multiple factorial analysis for mixed data (MFAMD), which combines both the MCA, the FAMD and the MFA, this technique highlights the main variability factors of the individuals, described these, in a balanced way by the groups of variables. Applying the MFAMD to the instrument of the baseline of the program Ser Pilo Paga, four multidimensional indices were created that summarize 9 thematic groups composed of 111 variables of quantitative and qualitative type, characterizing a total of 1,487 young people, which 682 are eligible to be part of the program and 805 are not. An estimator of difference in difference applied to the four multidimensional indices is proposed for the impact evaluation.

Keywords: Multiple factorial analysis for mixed data, Ser Pilo Paga, multivariate data analysis, discontinuous regression, Baseline, multidimensional indexes.

Introducción

El gobierno nacional de Colombia tiene como meta ser el país más educado de Latinoamérica para el año 2025 (Colombia 2015), el Ministerio de Educación Nacional (MEN) planteó las líneas estratégicas que determinarán el rumbo de la educación para los próximos años en diversos temas, siendo el acceso a la educación superior de calidad uno de los puntos más importantes en la línea estratégica. De acuerdo con Álvarez et al. (2017) en el año 2014, el MEN lanzó el programa Ser Pilo Paga (SPP) como la herramienta que permitirá alcanzar la meta educativa propuesta para el año 2025, este programa tiene como objetivo principal mejorar las condiciones de acceso a la educación superior de los jóvenes bachilleres que obtienen un excelente rendimiento en la prueba Saber 11, otorgando la posibilidad de acceder a un programa educativo de interés en instituciones de educación superior acreditadas de alta calidad. Según Álvarez et al. (2017), en la primera fase del programa los beneficiarios obtienen un crédito 100% condonable que permite cursar el programa académico de su preferencia si cumple los siguientes requisitos:

1. Haber presentado la prueba Saber 11 a partir del 3 de agosto de 2014 con un puntaje de 310 puntos o más.
2. Ser beneficiario del Sisbén¹ con puntaje igual o inferior al establecido de acuerdo al lugar de residencia del estudiante. En este sentido, los puntajes establecidos por el DNP² para la focalización del programa a través del puntaje Sisbén son los siguientes:
 - Un puntaje máximo de 57.21 para jóvenes residentes en las 14 ciudades principales sin sus áreas metropolitanas: Bogotá, Medellín, Cali, Barranquilla, Cartagena, Cúcuta, Bucaramanga, Ibagué, Pereira, Villavicencio, Pasto, Montería, Manizales y Santa Marta.
 - Un puntaje máximo de 56.32 para jóvenes residentes en zona urbana diferente a las 14 principales ciudades, los centros poblados y la zona rural dispersa de las 14 ciudades principales.
 - Un puntaje máximo de 40.75 para jóvenes residentes en área rural.
3. Presentarse y haber sido aceptado en una de las 33 universidades acreditadas de alta calidad.

Para la segunda fase del programa se realizaron unos cambios en los requisitos del programa:

¹Sistema de Selección de Beneficiarios Para Programas Sociales

²Departamento Nacional de Planeación

1. Ser colombiano.
2. Haber presentado las pruebas Saber 11 a partir del 2 de agosto de 2015 y haber obtenido un puntaje global igual o superior 318.
3. Haber cursado y aprobado el grado 11 a partir del año 2015.
4. Tener un puntaje específico individual de Sisbén según ubicación geográfica y de acuerdo con el corte definido por el DNP de la base Sisbén a partir del 19 de junio de 2015 (mismos puntajes de corte de Ser Pilo Paga 1) o en caso de ser indígena estar registrado en Base Censal del Ministerio del Interior con corte a 30 de junio de 2015.
5. Ser admitido en uno de los programas ofertados por alguna de las 39 instituciones de educación superior acreditadas en alta calidad.

En el año 2016 se modificó el puntaje mínimo a 348 en la prueba Saber 11, los demás requisitos como ser colombiano, presentación de las pruebas, haber cursado y aprobado el grado 11 y el puntaje Sisbén, se mantuvieron de forma similar con fechas establecidas para dicho año.

Además, a los beneficiarios del programa se les otorga un subsidio económico para su sostenimiento durante el tiempo que dure el programa, cuyo monto dependerá del lugar elegido de realización del programa como se presentan a continuación:

- 1 S.M.M.L.V³ semestral para estudiantes que no requieran desplazarse del lugar de residencia de su núcleo familiar para cursar sus estudios.
- 1.5 S.M.M.L.V semestral para estudiantes cuyo lugar de residencia de su núcleo familiar es diferente al lugar donde se oferta el programa académico que cursa el beneficiario y estos municipios están agrupados dentro de la misma región.
- 4 S.M.M.L.V semestral para estudiantes cuyo lugar de residencia de su núcleo familiar es diferente al lugar donde se oferta el programa académico que cursa el beneficiario y estos municipios no están agrupados dentro de la región.

Basado en los criterios anteriores, se determina la focalización de la población beneficiaria en las regiones de elegibilidad compuestas por las puntuaciones de la prueba $P_{saber11} \geq u_p$ y las puntuaciones asignadas por el $P_{sisben} \leq u_r$ siendo $P_{saber11}$ el puntaje obtenido por el joven en la prueba Saber 11 y P_{sisben} el puntaje determinado por el Sisbén. u_p es el umbral de la prueba Saber 11 y u_r el umbral de riqueza definido por el Sisbén.

En el contexto metodológico del programa SPP, se propuso realizar el diseño muestral sobre los municipios en donde se encuentran las instituciones de educación superior que poseen acreditación de alta calidad, otorgada por el MEN. En esta propuesta se consideraron los siguientes municipios para la realización de encuestas: Bogotá, Medellín, Cali, Barranquilla, Cartagena, Manizales, Pereira, Tunja, Popayán y Bucaramanga. Dadas las condiciones de la evaluación de los municipios seleccionados, se debía extraer la información de los jóvenes que cumplieran con los criterios de elegibilidad del programa SPP y de los individuos que servirían como controles en la evaluación, es decir, aquellos individuos que cumplen con el puntaje Sisbén exigido para ser beneficiarios, pero que no cumplen con el criterio de elegibilidad del puntaje en la prueba Saber 11. Bajo estas condiciones, la propuesta técnica contemplaba encuestar únicamente a aquellos individuos que residían en dichos municipios al momento de presentar la prueba Saber 11 en el 2015, para garantizar la consistencia en el impacto encontrado de implementar el programa.

Para los jóvenes no elegibles, se aclara que estos deben contar con un puntaje Sisbén igual o inferior al establecido por el programa para cada una de las áreas, con un puntaje Saber 11 lo suficientemente alto para estar dentro del ancho de banda óptimo elegido para la evaluación de impacto, y quienes pueden

³Salario Mínimo Mensual Legal Vigente

haber ingresado o no a la educación superior y haber completado sus estudios de educación media en alguno de los 793 municipios donde se encontró al menos un elegible dentro del ancho de banda.

En el caso de los elegibles, también es preciso aclarar que fueron objeto de selección todos aquellos individuos que tuvieron un puntaje en la prueba Saber 11 igual o superior al definido como puntaje de corte para ser beneficiario de SPP 310 para la primera corte, y 318 para la segunda y que están dentro del ancho de banda, independientemente de que hayan ingresado o no a la educación superior. Tanto elegibles como no elegibles debían haber completado sus estudios de educación media en alguno de los 793 municipios donde se encontraba al menos un elegible.

La metodología utilizada en el programa SPP hace parte de los diseños cuasi-experimentales de estimación de impacto causal conocida como regresión discontinua, aprovecha que existe una variable de focalización que determina quiénes son o no elegibles a un programa según un punto de corte establecido, se puede profundizar esta metodología en (Bernal & Peña 2011). Este tipo de metodologías define el concepto de ancho de banda como la región en donde se encuentran los jóvenes que cumplen los requisitos de Sisbén establecidos en el programa con puntajes en la prueba saber 11 cercanos al umbral definido, tanto por la izquierda como a la derecha del punto focal. Supone además que los individuos son muy similares entre si, con la única diferencia, de que unos eran elegibles para recibir SPP y los otros no. Existen otros tipos de enfoques como los propuestos por Imbens & Lemieux (2008), los cuales sugieren estimar dos regresiones discontinuas que permitan medir el impacto de forma separada entre el grupo de elegibles y el de no elegibles, otra propuesta para la evaluación de impacto se puede encontrar en Lee & Munk (2008), los cuales sugieren definir una única ecuación que permita capturar los efectos de los impactos promedio del programa y los errores aleatorios para cada individuo.

Las metodologías de regresión discontinua son una propuesta al esfuerzo de evaluar los impactos del programa SPP. Las consideraciones iniciales de evaluar los impactos de una política pública puede conllevar a un sin número de esfuerzos y estrategias que permitan su medición y acreditación, para el caso del programa de SPP, se deberá esperar que los primeros beneficiarios cumplan sus planes de estudio para poder determinar cuál fue el impacto de participar en el programa, no solo como política o estrategia gubernamental, sino en las diversas temáticas evaluadas en esta investigación como el déficit de la vivienda, los gastos del joven, las expectativas educativas, entre otras. Es posible que un joven evaluado en la línea base⁴ haya sido impactado en todas las temáticas de interés por haber participado en el programa, o solo se haya dado en algunos temas y variables en particular, o en el peor de los casos, no se haya generado ningún impacto por tomar el programa de SPP.

Con lo expuesto anteriormente, el objetivo de este trabajo es medir el impacto del programa de SPP con un estimador de diferencia en diferencia, aplicado a cuatro índices multidimensionales que permiten medir el comportamiento de cada variable, y además, analiza toda una temática en particular a partir del AFMDM caracterizando a la población total de la línea base.

En la sección 1: Marco Teórico, se describe los fundamentos teóricos del análisis factorial múltiple para datos mixtos, en la sección 2: Metodología, se encuentra la descripción de la fuente de información, la construcción de algunas variables dicotómicas y la conformación de los 9 grupos de variables, en la sección 3: Resultados, se presentan los hallazgos de aplicar la técnica factorial, y se presenta la construcción de los cuatro índices multidimensionales con la propuesta del estimador de diferencia en diferencia, en la sección 4: Conclusiones, se harán comentarios generales de los resultados obtenidos, en la sección 5: Discusión, se reseñaran hallazgos, posibles limitaciones y recomendaciones para trabajos futuros.

⁴Es una primera etapa de medición de los indicadores planteados en una investigación, habitualmente, en proyectos de desarrollo social, lo que permite conocer el valor de los indicadores planeados, estos valores iniciales son el punto de partida o referencia del proyecto o intervención

1. Marco Teórico

1.1. Análisis factorial múltiple para datos mixtos

En diferentes estudios las unidades estadísticas se describen en dos tipos de variables: cuantitativas y cualitativas. Un ejemplo común donde se observa este tipo de variables son los instrumentos de recolección de información conocidos como encuestas, por lo general, en estos instrumentos existen numerosas preguntas cerradas que suelen ser reunidas en varios conjuntos que pueden ser de naturaleza cuantitativa o cualitativa (Bécue-Bertaut & Pagès 2008).

Tratar simultáneamente variables de tipo cuantitativo y cualitativo fue abordado por Escofier (1979) el cual sugiere la introducción de variables cuantitativas en el ACM bajo una codificación apropiada, otro caso es el propuesto por Saporta (1990) donde se considera la introducción de información de variables cualitativas en el análisis de componentes principales (ACP), ya que creía, que sí era posible la implementación de una métrica especial, esta permitiría analizar de forma simultánea los dos tipos de variables. Sin embargo, el abordaje de los datos mixtos ya había sido estudiado por Gower (1971), el cual propuso por primera vez una solución para equilibrar variables cuantitativas y cualitativas cualquiera sea su tipo. Las distancias específicas se utilizan para variables cualitativas y cuantitativas donde el rango de variación de cada distancia se estandariza a 1 antes de la agregación en una distancia global. Además, al utilizar un conjunto de ponderaciones a priori pueden favorecer aquellas variables consideradas importantes.

Es preciso resaltar que el AFM, es una técnica multivariada que permite analizar grupos de variables cuantitativas y/o cualitativas, por lo cual, es necesario generar una extensión que permita que el análisis factorial múltiple a grupos se pueda realizar simultáneamente a variables cuantitativas o mixtas (Pagès 2014), luego el método de referencia que permite procesar individuos por tabla de variables cualitativas, es el análisis de correspondencias múltiple (ACM), de este modo se tendrá una combinación de ACM, AFMD y AFM siendo la ponderación de los grupos de variables implementada en el AFM la clave para tratar el instrumento de interés.

1.1.1. AFM de grupos mixtos

Manteniendo la notación propuesta por (Pagès 2014) del AFM, define a j como el índice de un grupo específico y J la cantidad total de grupos, análogamente q define una variable cualitativa y Q el número total de variables cualitativas como se explica a continuación:

- Sea Y una matriz disyunta completa con y_{ik} las variables en la matriz, en donde $y_{ik} \in \{0, 1\}$.
En la matriz disyunta completa, los indicadores son agrupados en grupos que incluyen varias variables, de modo que los indicadores de una variable siempre pertenecen a un solo grupo para el caso de variables cualitativas.
- K_j es el número de categorías en el grupo j .
- $K = \sum_j K_j$ es el total de categorías.
- Q_j es el número de variables en el grupo j .
- Q es el total de variables $Q = \sum_j Q_j$.

Un grupo j se descompone ahora en grupos j_1 de variables cuantitativas, en j_2 para variables cualitativas y en j_3 para variables mixtas, y de forma simultánea se analizan variables mixtas por el AFDM, pero AFDM no tiene en cuenta la noción de grupos de variables. Es necesario realizar la extensión del AFM a datos mixtos ya que esta técnica factorial permite el análisis simultáneo de grupos de variables de diferente naturaleza.

Según Pagès (2014) para hacer referencia al peso de las variables, los grupos se equilibran con la ponderación del AFM $\frac{1}{\lambda_1^j}$, y el peso inicial de la variable k del grupo j se divide por:

- λ_1^j si el grupo j es cuantitativo o mixto por lo que λ_1^j es el primer valor propio del ACP o del AFDM para el grupo j .
- $\lambda_1^j Q_j$ si el grupo j es categórico λ_1^j es el primer valor propio del grupo del MCA para el grupo j incluyendo las variables Q_j .
- λ_s es el mayor valor propio del análisis global.

El mismo resultado puede obtenerse reemplazando los datos brutos con los factores de los análisis factoriales por separado de los grupos (ACP, MCA o AFDM según el caso). Estos factores no están estandarizados o atribuidos a un peso inicial igual a su valor propio asociado. Esta propiedad sugiere que las contribuciones a las inercias de los ejes, debe calcularse tanto para las variables iniciales como para los factores de los análisis por separado.

1.1.2. Representación de las nubes parciales de individuos

Al igualar las inercias máximas de las nubes parciales, es posible su comparación de forma simultánea (Pagès 2014), esta representación se beneficia de las relaciones de transición parciales idénticas a las del AFM aplicadas a grupos de variables del mismo tipo. Ahora, el individuo parcial i del grupo cualitativo j se encuentra en el baricentro de las categorías del grupo j que posee, de la siguiente forma:

$$F_s(i^j) = \frac{1}{\lambda_s} \frac{J}{\lambda_1^j Q_j} \sum_{k \in K_j} y_{ik} F_s(g_k) = \frac{1}{\sqrt{\lambda_s}} \frac{J}{\lambda_1^j Q_j} \sum_{k \in K_j} y_{ik} G_s(k)$$

El individuo parcial i del grupo cuantitativo j se encuentra del lado de las variables del grupo j que tienen valores altos y opuestos a los del grupo j , los cuales se caracterizan por tener valores bajos; así:

$$F_s(i^j) = \frac{1}{\sqrt{\lambda_s}} \frac{J}{\lambda_1^j} \sum_{k \in K_j} x_{ik} G_s(k)$$

En resumen, un individuo se encuentra tanto del lado de las variables cuantitativas para las cuales tiene un valor alto como del lado de las categorías que posee. En las relaciones previas, el valor J asegura que el individuo i se encuentre en el isobaricentro⁵ de sus puntos parciales i^j , para cualquier tipo de variable de la siguiente manera:

$$F_s(i) = \frac{1}{J} \sum_{j \in J} F_s(i^j)$$

Las categorías parciales se pueden definir a partir de los individuos parciales. La categoría parcial k del grupo j se denota k^j ; se encuentra situada en el isobaricentro de los individuos parciales i^j asociada con los individuos que poseen la categoría k . Esta definición se aplica para cualquiera que sea la naturaleza del grupo j .

⁵es un baricentro con todas las masas iguales entre sí

1.1.3. Análisis multicanónico

Para las variables cuantitativas según (Pagès 2014), es posible considerar el AFM como un análisis multicanónico con la condición de que la relación entre una variable z y un grupo de variables $K_j = v_k; k = 1, K_j$ es medido por Lg que proyecta la inercia de las variables del grupo K_j a lo largo de z . De la siguiente manera:

$$Lg(z, K_j) = \sum_{k \in K_j} \text{inercia de la proyección de } v_k, z$$

Cuando una variable j está compuesta por indicadoras de un elemento cualitativo de la variable V la siguiente medida es adecuada para z centrado como sigue:

$$Lg(z, K_j) = \eta^2(z, V)$$

En donde $\eta^2(z, V)$ es la relación de correlación cuadrada entre z y la variable V si el grupo j incluye Q_j variables cualitativas V_q , esta medida se puede expresar como:

$$Lg(z, K_j) = \frac{1}{\lambda_1^j} \frac{1}{Q_j} \sum_{q \in Q_j} \eta^2(z, V_q)$$

Esta medida es proporcional a la media de la correlación al cuadrado de las relaciones entre la variable z y las variables que componen el grupo j . Esta media se divide por su valor máximo contabilizando los datos del grupo j respecto al primer valor propio del ACM de este grupo j . De hecho, en el ACM, el valor propio de rango s es igual a la media de las razones de correlación al cuadrado entre factor F_s y las Q variables cualitativas. Finalmente, la medición de Lg vale 0 si la variable z tiene una relación de correlación cero con cada variable del grupo j , y valdrá 1 si z coincide con la dirección principal de inercia del grupo j .

La cantidad maximizada por los factores del AFM puede expresarse en términos de un análisis canónico, con estos factores como variables generales. Esta cantidad se da a continuación, y es limitada a tres grupos: K_1 (grupo cuantitativo compuesto por variables $K_1 v_k$), K_2 (grupo cualitativo compuesto por variables $Q_2 v_q$) y K_3 (grupo mixto compuesto por variables cuantitativas $K_3 v_k$ y variables cualitativas $Q_3 v_q$). La variable general de rango s (denotada v_s) maximiza a:

$$Lg(v_s, K_1) + Lg(v_s, K_2) + Lg(v_s, K_3) =$$

$$\frac{1}{\lambda_1^1} \sum_{k \in K_1} r^2(v_s, v_k) + \frac{1}{\lambda_1^2 Q_2} \sum_{q \in Q_2} \eta^2(v_s, v_q) + \frac{1}{\lambda_1^3} \left[\sum_{k \in K_3} r^2(v_s, v_k) + \sum_{q \in Q_3} \eta^2(v_s, v_q) \right]$$

Con las restricciones habituales de norma y ortogonalidad para v_s y denotando λ se considera j_1 como el primer valor propio del análisis factorial (ACP, ACM o AFDM) del grupo j . Como en AFDM, esta cantidad significa que el coeficiente de correlación al cuadrado y la razón de correlación al cuadrado juegan el mismo papel, esta aumenta cuanto más estrechamente es la correlación con las variables cuantitativas, y es más fuerte esta relación con la correlación entre las variables cualitativas. De esta forma, v_s está vinculado a K_1 , K_2 y K_3 . La ponderación de los grupos de variables está limitada a la contribución máxima de cada uno de los grupos a la cantidad de 1. Por lo tanto, esta cantidad vale un máximo de 3 (caso general: J), cuando v_s coincide con el primer factor del ACP de K_1 y con el ACM de K_2 y con el de AFDM de K_3 .

1.1.4. Coeficientes Rv y Lg

El coeficiente Rv definido por Escoufier (1973) es considerado una medida de similitud entre dos matrices que se caracterizan por ser simétricas, especialmente semidefinidas positivas. Algunos autores consideran este coeficiente como una extensión matricial del coeficiente de Spearman. De este modo se define X y Y matrices reales, y $S = XX'$ y $T = YY'$, con S y T semidefinidas positivas, entonces el coeficiente Rv se define como:

$$Rv = \frac{\text{Traza}(S'T)}{\sqrt{\text{Traza}(S'S)\text{Traza}(T'T)}}$$

Es necesario considerar que en el AFM estas matrices tienen el mismo número de filas y de columnas permitiendo calcular la homotecia a partir de las matrices de covarianzas de S y de T .

El coeficiente Lg es considerado como una medida de relación entre dos matrices que indica el número de características compartidas entre ellas en sentido de homotecia⁶, si el número de variables que se relacionan entre las matrices es alta, el coeficiente Lg será más grande, este coeficiente varía entre 0 y 1. Se define este coeficiente de la siguiente manera:

$$Lg = \frac{\text{Traza}(S'T)}{\alpha_1^2 \lambda_1^2}$$

Donde α_1^2 y λ_1^2 son los primeros valores propios de S y de T (Abdi et al. 2013).

2. Metodología

2.1. Fuente de la información

La información de la línea base del programa SPP fue recolectada a través de una encuesta realizada directamente a los jóvenes elegibles y no elegibles del programa, permitiendo la caracterización de los 1.487 jóvenes que cumplían el puntaje de focalización del Sisbén de SPP y que se encontraban alrededor del punto de corte de Saber 11. De estos jóvenes, 682 cumplían el requisito de Saber 11 y 805 no lo hacían. Para la elección de los participantes se utilizaron las bases del Sistema para la Prevención de la Deserción de la Educación Superior SPADIES, la base de Saber 11 y la base de ICETEX⁷.

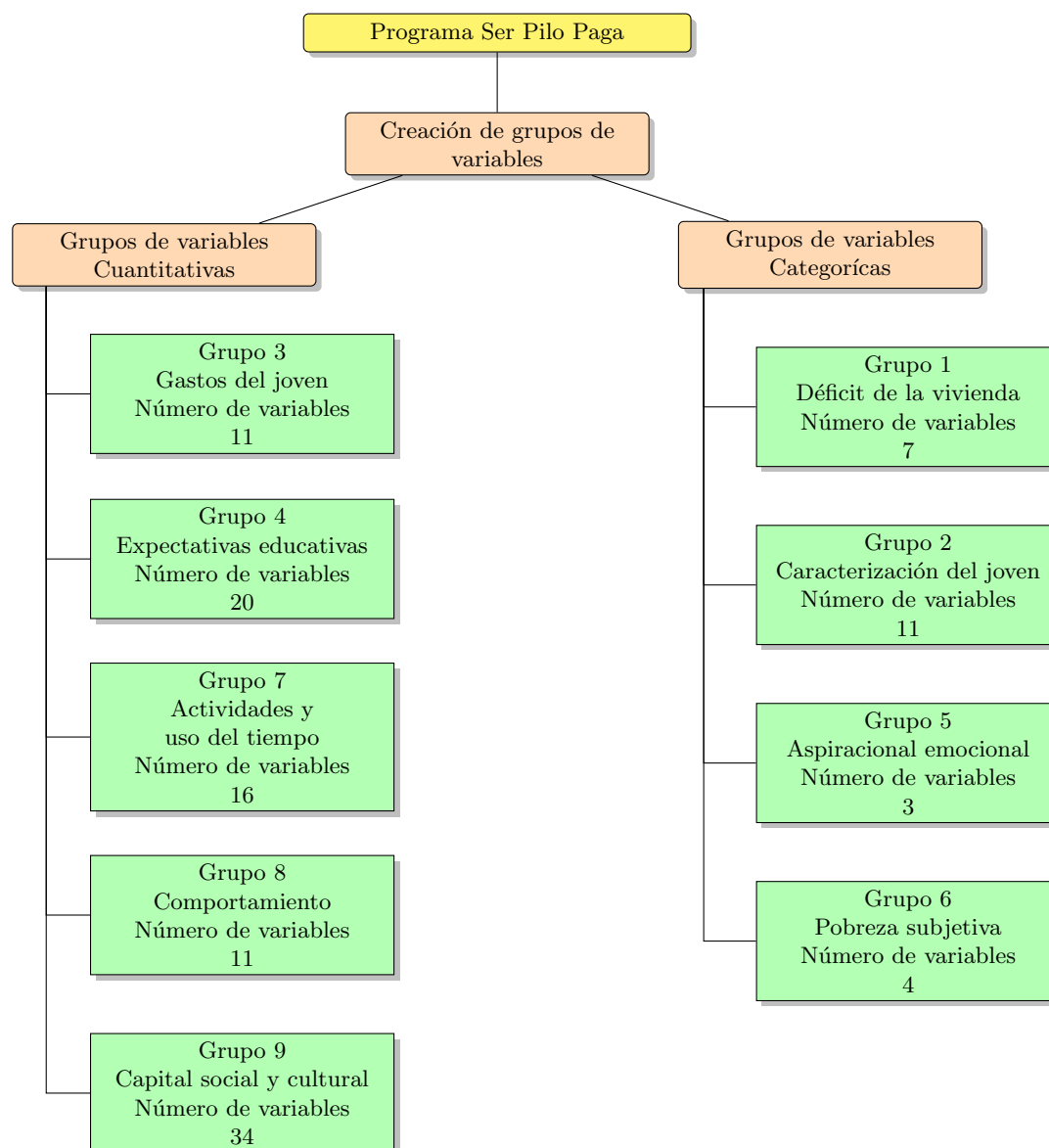
2.2. Conformación de los grupos de variables

Con el propósito de cumplir los objetivos de esta investigación, se seleccionaron 117 variables (ver el anexo A.1) de las 706 de la línea base del programa SPP, estas variables fueron distribuidas en 9 grupos que hacen referencia a una temática en particular. Las variables pertenecientes a un grupo tienen una sola naturaleza (cuantitativa o cualitativa), descartando por completo que un grupo tenga variables de ambas naturalezas.

A continuación se detalla la naturaleza del grupo, su nombre temático y el número de variables que lo conforman como se muestra a continuación:

⁶El algunas definiciones se considera como una proyección o transformación respecto a un punto de una figura la cual es multiplicada todas sus distancias por un mismo factor generando proporciones mayores o menores de la figura

⁷El Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior



2.3. Construcción de variables

Con las variables seleccionadas es necesaria la construcción de algunas variables nuevas, con el fin de conformar un grupo que represente el déficit de la vivienda y la caracterización del joven. Su elaboración busca brindar mayor información desde la combinación con otras variables naturales del instrumento, reestructuraciones o colapsos de categorías.

La mayoría de las variables son dicotómicas, en donde el valor de uno representa la existencia y cero la ausencia de una característica. En el grupo del déficit de la vivienda se evalúa por ejemplo el tipo de vivienda, la composición de las paredes, los pisos de la vivienda, los servicios públicos, la alimentación del joven, entre otras. En este grupo se tiene en cuenta el hacinamiento de la vivienda, según el Departamento Administrativo Nacional de Estadística (DANE) en (Colombia 2009) se considerará un hogar con hacinamiento mitigable si habitan entre tres a cuatro personas por cuarto (sin incluir cocina, baños y garajes), un hogar se considerará con hacinamiento no mitigable, si habitan cinco o más personas por cuarto (sin incluir cocina, baños y garajes), en caso contrario, el hogar evaluado no tendrá hacinamiento.

Esta nueva variable es la única del grupo del déficit de vivienda, que no es dicotómica ya que posee tres categorías, si el hacinamiento es mitigable, no mitigable o no hay hacinamiento.

El grupo de la caracterización del joven está conformado por tres variables dicotómicas, que evalúan el estado civil del joven, su tipo de afiliación a la EPS⁸ y si recibe ingresos actualmente, a diferencia de los ejemplos del grupo del déficit de la vivienda, el colapsos de las categorías del estado civil por ejemplo definen estar soltero o tener otro estado civil, mientras que el tipo de afiliación define tener o no afiliación a una EPS, mientras que percibir ingresos evalúa una serie de ítems en donde se indica si recibe o no ingresos por esos conceptos.

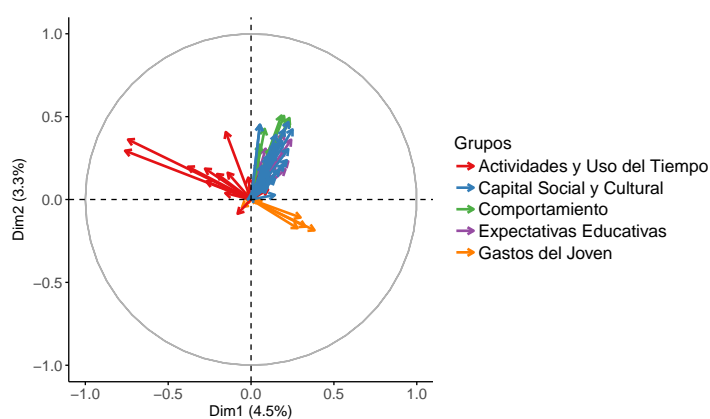
3. Resultados

Una vez realizada la revisión de la documentación existente sobre el programa de SPP, la selección y creación de los grupos variables de interés, la verificación de las metodologías existentes para el tratamiento simultaneo de grupos de variables de diferente naturaleza, se procede hacer uso del análisis factorial múltiple para datos mixtos implementada en el paquete FactoMineR (Lê et al. 2008) de (R Core Team 2017).

3.1. Descripción multivarida de los datos

De forma general, se desarrolla un AFMDM con el conjunto de variables mencionado en el anexo A.1 y las nuevas variables creadas, usando un total de 117 variables activas, con un análisis previo se permite identificar que algunas variables no aportan significativamente a los resultados esperados, preguntas que evaluaron el estado civil del joven, si tiene alguna discapacidad, o si fue admitido a la universidad, entre otras, fueron sacadas del conjunto de datos final por lo que el conjunto de trabajo con el que se desarrollaron los análisis es de 111 variables que se encuentran contenidas en 9 grupos con distinta naturaleza.

Figura 1: Direcciones de los constructos de las variables cuantitativas



Fuente: Elaboración propia

Después de la depuración de variables mencionada, la figura 1 presenta los resultados del comportamiento de las variables de tipo cuantitativo para cada uno de los grupos compuestos por este tipo de variables, se resalta que todas estas tienen una dirección definida intuyendo un concepto global, este comporta-

⁸Entidad Prestadora de Salud

miento es un resultado importante para el análisis, ya que las variables seleccionadas logran recoger los constructos que se desean explicar.

Tabla 1: Correlaciones de los grupos resultados globales

Dimensiones	Dim1	Dim2	Dim3	Dim4
Déficit de la vivienda	0,33	0,25	0,58	0,38
Caracterización del joven	0,79	0,45	0,44	0,46
Gastos del joven	0,53	0,30	0,45	0,42
Expectativas educativas	0,37	0,58	0,38	0,23
Aspiracional emocional	0,70	0,24	0,39	0,56
Pobreza subjetiva	0,49	0,25	0,61	0,73
Actividades y uso del tiempo	0,83	0,52	0,46	0,31
Comportamiento	0,28	0,70	0,26	0,13
Capital social y cultural	0,37	0,68	0,40	0,43

Fuente: Elaboración propia

La tabla 1 presenta las correlaciones que tiene cada uno de los grupos con cada una de las dimensiones retenidas del análisis global. La dimensión 1 se correlaciona principalmente con los grupos de actividades y uso del tiempo, caracterización del joven, aspiracional emocional y el grupo de los gastos del joven, mientras la dimensión 2 se correlaciona con los grupos del comportamiento, el capital social y cultural y con el grupo de las expectativas educativas, la dimensión 3 se correlaciona con el déficit de la vivienda, mientras que el grupo de la pobreza subjetiva esta altamente correlacionado con la dimensión 4.

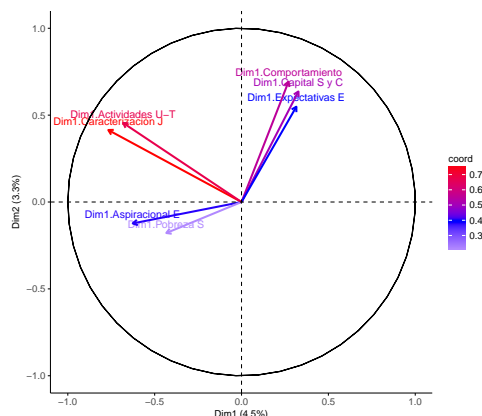
Considerando inicialmente que los grupos de variables se encuentran correlacionados con 4 dimensiones del análisis global, la tabla 3 del anexo A.2 presenta los coeficientes Lg , según Corzo (2017) estos coeficientes son un “indicador del grado de similitud o deformación con respecto a un foco (homotecia) entre los conjuntos de indicadores”, luego el valor máximo encontrado entre un grupo y el AFMDM es $Lg_{(AFM, PobrezaS)} = 3.42$, lo que permite concluir que las 4 dimensiones retenidas en el análisis global permiten describir las agrupaciones de interés. Por otra parte, la tabla 4 del anexo A.2 presenta los resultados del coeficiente Rv para cada uno de los 9 grupos, se puede concluir que aun no siendo parecidos entre grupos y presentando asociaciones entre el AFMDM y los grupos (medio-baja), su representación sobre los planos generados puede considerarse adecuada.

3.2. Interpretación de las dimensiones globales

Dimensión 1: La dimensión 1 de actividades y uso del tiempo, la dimensión 1 de la caracterización del joven, la dimensión 1 de aspiracional emocional y la dimensión 1 del grupo de la pobreza subjetiva del análisis separado, poseen una alta correlación con la parte negativa de esta dimensión, mientras la dimensión 3 de los gastos del joven del análisis separado se contraponen en su parte positiva (Ver eje horizontal figura 2), lo que significa que la dimensión 1 global en su parte negativa, se caracteriza por los jóvenes que gastan poco tiempo en actividades académicas como estudiar y hacer tareas, ya que no asisten a un programa de educación superior o de formación para el trabajo, la mayor parte del tiempo estuvieron trabajando, buscando trabajo o en oficios del hogar (ver Figura 8, Tabla 7 del anexo A.2), en sus aspiraciones educativas, creen que alcanzarán el nivel universitario y les gustaría realmente alcanzar este nivel educativo (ver Tabla 8 del anexo A.2), se consideran pobres y desean pertenecer a la clase media cuando tengan 30 años. Los jóvenes descritos anteriormente, se contraponen de aquellos que hacen uso de sus recursos económicos para cubrir sus gastos en transporte y materiales (ver Tabla 7 del anexo A.2).

Dimensión 2: La dimensión 1 del comportamiento, la dimensión 1 del capital social y cultural, y la dimensión 1 de las expectativas educativas del análisis separado poseen una alta correlación con la parte positiva de esta dimensión (ver eje vertical de la Figura 2), luego esta dimensión se caracteriza por los jóvenes que planean su futuro en cuanto al estudio o el trabajo, consideran que el esfuerzo en estos campos será recompensado con un mejor trabajo, terminan todo lo que comienzan y trabajan con compromiso

Figura 2: Dimensión 1 y 2 del análisis global

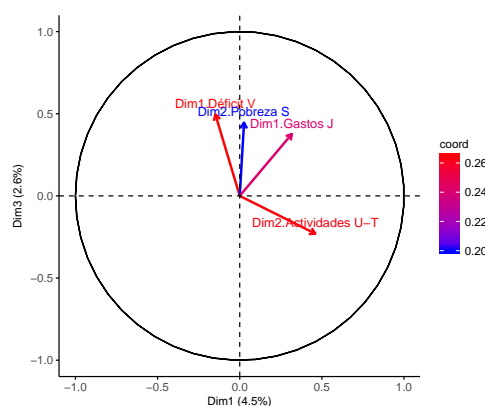


Fuente: Elaboración propia

y dedicación para alcanzar sus metas (ver Figura 8 del anexo A.2), estos jóvenes se caracterizan además por su núcleo familiar, ya que comparten mucho tiempo con sus familias, compartían lecturas con él, dialogaban sobre sus opciones para el futuro, son jóvenes que sienten que tienen muchas cosas por enseñarle a los demás (ver Figura 9 del anexo A.2), les gusta estudiar porque creen que a través de él desarrollarán habilidades útiles para el trabajo, permitiéndoles aumentar su confianza (ver Figura 7 del anexo A.2).

Dimensión 3: La dimensión 1 del déficit de la vivienda, la dimensión 2 de la pobreza subjetiva, y la dimensión 1 de los gastos del joven del análisis separado poseen una correlación moderada con esta dimensión en su parte positiva, mientras que en la parte negativa se correlaciona con la dimensión 2 de las actividades y uso del tiempo del análisis separado. Los jóvenes que se encuentran en la parte positiva de esta dimensión se caracterizan por presentar múltiples deficiencias en su hogar, en temas como los servicios públicos, los pisos y las paredes del hogar, además, con un posible hacinamiento mitigable (Tabla 6 del anexo A.2), son jóvenes que se consideran pobres y de clase social baja (Tabla 9) del anexo A.2), que direccionan todos sus gastos en alimentación, arriendo, internet para el teléfono y en el pago de los servicios del hogar (ver Figura 7 del anexo A.2), en su parte negativa se encuentran los jóvenes que gastan una gran parte de su tiempo en asistir a eventos culturales y pasar tiempo con sus amigos en actividades de recreación o diversión (ver Figura 8) del anexo A.2).

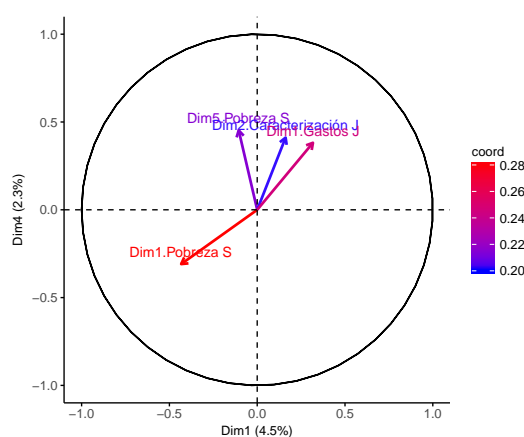
Figura 3: Dimensión 3 (Vertical) del análisis global



Fuente: Elaboración propia

Dimensión 4: La dimensión 5 de la pobreza subjetiva, la dimensión 2 de la caracterización del joven, y la dimensión 1 de los gastos del joven del análisis separado, poseen una correlación moderada con esta dimensión en su parte positiva, mientras que en la parte negativa se correlaciona con la dimensión 1 de la pobreza subjetiva del análisis separado, luego esta dimensión se caracteriza por los jóvenes que consideran que seguirán perteneciendo a la clase social baja cuando tenga una edad de 30 años, esperan alcanzar en esta edad un salario entre 0.7 y 1.4 millones de pesos (ver Tabla 9 del anexo A.2), son jóvenes que tienen personas a cargo (Ver Tabla 7 del anexo A.2) y que además gastan todos sus recursos en temas principalmente en alimentación, arriendo e internet para el teléfono celular (ver Tabla 7 del anexo A.2), en su parte negativa esta dimensión se caracteriza por los jóvenes que desean pertenecer a la media cuando tengan 30 años y recibir ingresos entre 2,1 y 3,5 millones de pesos (ver Tabla 9 del anexo A.2).

Figura 4: Dimensión 4 (Vertical) del análisis global



Fuente: Elaboración propia

3.3. Índices multidimensionales

En esta sección se hace uso del término índice para hacer referencia a una de las cuatro dimensiones retenidas en el análisis global. Cada índice observado, recoge una serie de cualidades que permitieron caracterizar a la población de interés, a través de unos atributos específicos evaluados para cada uno de los jóvenes del programa.

Estos índices tienen las siguientes características:

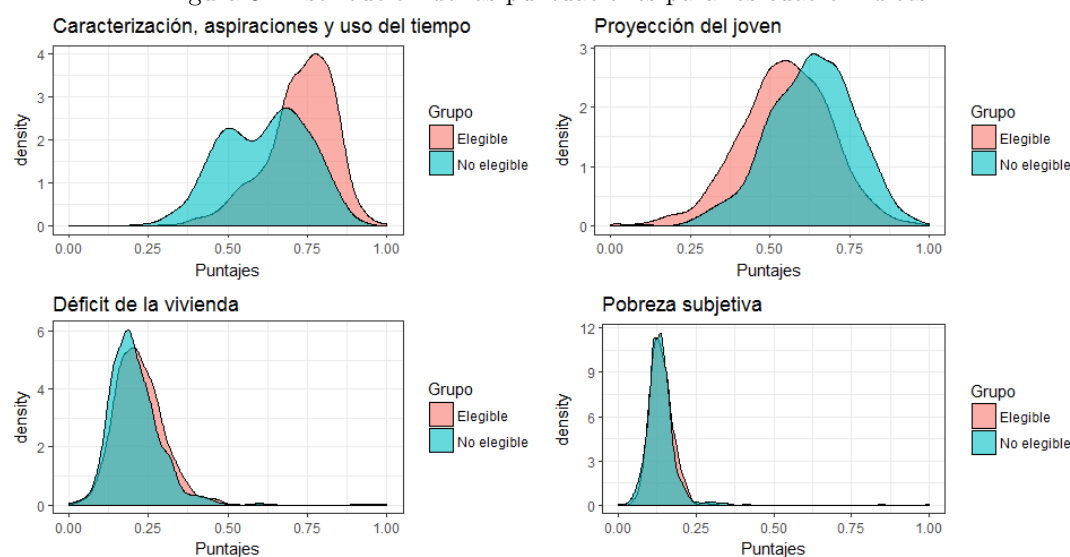
1. I_{Acu} : **Caracterización, aspiraciones y uso del tiempo:** los jóvenes con puntajes altos en este índice, son personas alejadas del ámbito académico, pobres, obligados a trabajar o en búsqueda de trabajo, creen y desean poder alcanzar al menos un título universitario y mejorar su nivel social, con puntajes bajos se encuentran los jóvenes que actualmente gastan su dinero en temas de transporte y materiales, estos gastos pueden estar relacionados directamente por actividades académicas o de formación universitaria.
2. I_{Pj} : **Proyección del joven:** los jóvenes con puntajes altos en este índice tienen un concepto positivo de trabajo y esfuerzo en lo académico o laboral, con un notorio acompañamiento formativo de la familia.
3. I_{Dv} : **Déficit de la vivienda:** los jóvenes con puntajes altos en este índice se caracterizan por ser pobres, con condiciones de pobreza considerables, sus gastos están enfocados a la subsistencia.

Los jóvenes con puntajes bajos son aquellos que gastan una cantidad de horas considerables de su tiempo en la televisión, películas y demás actividades de ocio.

4. I_{Ps} : **Pobreza subjetiva**: los jóvenes con puntajes altos en este índice se caracterizan por tener personas a cargo, consideran que seguirán siendo pobres con niveles salariales bajos, gastan sus recursos en su subsistencia. Los jóvenes con puntajes bajos son aquellos que se consideran pobres, sin embargo disfrutaban estudiar, ya que el estudio les permite desarrollar habilidades para encontrar un trabajo, aumentando su confianza.

Como cada joven presenta una puntuación factorial en el análisis global, la tabla 5 del anexo A.2, presenta el reescalamiento de dichas puntuaciones para cada individuo, entre 0 a 100 y la figura 5 su distribución para cada uno de los índices 4 índices multidimensionales, este proceso se elaboró para construir la tabla 2 que cuantifica la media de las puntuaciones de los jóvenes elegibles y no elegibles del programa para cada índice, luego, este valor es una referencia de partida con la que se pretende comparar los resultados del programa con los observados en la nueva evaluación.

Figura 5: Distribución de las puntuaciones para los cuatro índices



Fuente: Elaboración propia

Los índices tanto para elegibles como no elegibles, logran recoger una serie de características dadas por varias variables que reproducen unos temas en particular, y que para efectos de la línea base, no tienen una diferencia significativa, de modo tal, que pueden ser utilizados para posteriormente hacer la evaluación del impacto del programa, particularmente para las temáticas agrupadas en cada uno de los índices.

Tabla 2: Índices multidimensionales propuestos

Grupo	I_{Cau} : Caracterización, aspiraciones y uso del tiempo	I_{Pj} : Proyección del joven	I_{Dv} : Déficit de la vivienda	I_{Ps} : Pobreza subjetiva
Elegibles	0,73	0,55	0,22	0,14
No elegibles	0,62	0,63	0,21	0,14

Fuente: Elaboración propia

Un posible planteamiento que permita medir el impacto una vez se cuente con el estudio de seguimiento, puede ser a través de un estimador de diferencia en diferencia como lo indica (Bernal & Peña 2011). A

continuación se presenta dicho estimador, poniendo como ejemplo el cálculo del impacto para el índice $I_{(Acu)}$:

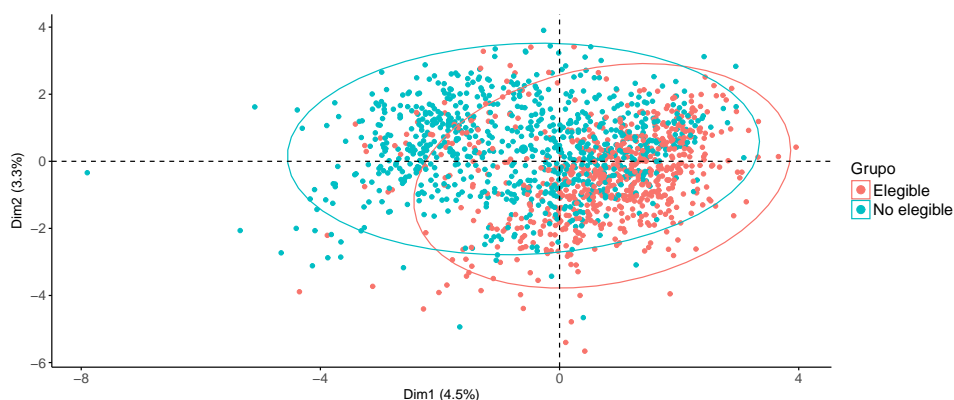
$$D = (I_{(Acu,Elegible,t_1)} - I_{(Acu,Noelegible,t_1)}) - (I_{(Acu,Elegible,l_b)} - I_{(Acu,Noelegible,l_b)})$$

Si $D > 0$ concluye que el impacto del programa es positivo para el índice I_{Acu} , entre elegibles y no elegibles, en las características que resaltan a I_{Acu} , entre la nueva medición (t_1) y las puntuaciones actuales de línea base (l_b), si $D < 0$ entonces el impacto en este índice es negativo, y si $D = 0$ no existe impactos del programa entre elegibles y no elegibles en las características de este índice. De forma similar se puede realizar los cálculos del estimador de diferencia en diferencia para los otros índices.

3.4. Mapa de individuos

La figura 6 permite observar que la población de interés (jóvenes elegibles y no elegibles) comparten similitudes desde el análisis propuesto impidiendo encontrar una clara división de las poblaciones, esto es concluyente desde el comportamiento de las elipses, ya que estas no logran separar a los jóvenes respecto a su grupo focal, luego, la focalización multidimensional no está discriminando por un tema en particular a los jóvenes del programa dentro del ancho de banda, siendo este un hallazgo positivo por la conformación de los datos de la línea base. Según Bernal & Peña (2011) cuando se parte de metodologías de regresión discontinua “el estimador se basa en que los individuos justamente de la izquierda del umbral deben ser muy parecidos a los individuos justamente a la derecha del umbral”, la única diferencia entre los jóvenes elegibles y no elegibles es que unos pertenecerán al programa y los otros no.

Figura 6: Mapa de individuos elegibles y no elegibles



Fuente: Elaboración propia

La regresión discontinua considera un ancho de banda solamente teniendo en cuenta los resultados de la prueba Saber 11, luego, las condiciones iniciales sobre la selección de la muestra de los jóvenes elegibles y no elegibles mantienen sus similitudes bajo las variables de estudio, lo que significa, que la metodología propuesta logra recoger multidimensionalmente que los individuos no se diferencian respecto a las variables seleccionadas en este estudio.

4. Conclusiones

Dada la selección de las 111 variables cuantitativas y cualitativas de línea base del programa SPP y la conformación de los 9 grupos temáticos, la implementación de la técnica factorial del AFMDM reduce la interpretación de los 9 grupos a cuatro índices multidimensionales que caracterizan al total de la población del programa.

La tabla 2 y la figura 5 presenta una característica importante en los cuatro índices planteados sobre la línea base, estos no se diferencian, lo que significa que podrán realizar las mediciones de la evaluación de impacto sobre cada una de las temáticas descritas a través del estimador de diferencia en diferencia, obteniendo un ahorro significativo en tiempo sobre futuros abordajes de la línea base del programa, ya que los análisis de interés se deben desarrollar en los índices propuestos y no en la totalidad de las variables evaluadas en el instrumento.

5. Discusión

La selección de variables y la generación de los grupos de variables estuvieron basadas en criterios estadísticos que generaran buenas asociaciones y/o correlaciones. Previo a los resultados obtenidos, se elaboró una depuración de variables a través de técnicas multivariadas que permitieron eliminar variables del análisis por su poco aporte, sin embargo, se debe tener en cuenta que la elección y conformación de los grupos de variables no es único, un experto por ejemplo podría tener distintos criterios en la elección y agrupación de variables y obtener resultados distintos a los observados en esta investigación. Es importante resaltar que se considera adecuada la inclusión del criterio de un experto si se implementa esta metodología, ya que el enfoque metodológico está orientado a la forma en cómo se deben construir los índices multidimensionales para la evaluación de impacto del programa SPP.

Aunque los resultados de la tabla 4 presentan coeficientes que se consideran bajos entre los grupos y el AFMDM, investigaciones como la propuesta por (Adams 2016) permiten observar que el coeficiente Rv se ve altamente afectado por temas de tamaño de muestra y el número de variables en el análisis, tamaños de muestra bajos hacen que el coeficiente sea alto y tamaños de muestra altos que este disminuya, para el caso del número de variables, entre más variables se tengan, mayor es el coeficiente. Por lo que el coeficiente Rv se ve adversamente afectado por los atributos de los datos que no caracterizan la estructura de covarianza entre los conjuntos de variables. Como consecuencia, el autor considera que es difícil determinar si los valores de Rv pequeños son el resultado de una mayor independencia entre los módulos, Además, no se evidencia la eficiencia o problemática del coeficiente con variables de naturaleza tanto cuantitativa como cualitativa, por lo que se recomienda hacer uso del coeficiente con precaución y cuantificar su importancia a través del criterio de un experto.

Para los trabajos futuros: En instrumentos de recolección de información como las encuestas, es posible conseguir variables de tipo frecuencia, para este tipo de variables (Bécue-Bertaut & Pagès 2008) proponen realizar simultáneamente este análisis teniendo en cuenta la inclusión de este tipo de variables. La línea base del programa contaba con variables de tipo frecuencia, sin embargo, su comportamiento no otorgo resultados coherentes y confiables que permitiera tenerlas en cuenta para esta investigación.

El estimador de diferencia en diferencia es uno de los tantos estimadores que se pueden plantear para medir el impacto del programa de SPP, por lo que en una investigación futura se debería comparar la eficiencia de otros estimadores con el planteado en esta investigación.

Referencias

- Abdi, H., Williams, L. J. & Valentin, D. (2013), 'Multiple factor analysis: principal component analysis for multitable and multiblock data sets', *Wiley Interdisciplinary reviews: computational statistics* **5**(2), 149–179.
- Adams, D. C. (2016), 'Evaluating modularity in morphometric data: challenges with the rv coefficient and a new test measure', *Methods in Ecology and Evolution* **7**(5), 565–572.
- Álvarez, M. J., Castro, C., Corredor, J., Londoño, J., Maldonado, C., Rodríguez, C., Sánchez, F., Pulido, X. et al. (2017), El programa ser pilo paga: Impactos iniciales en equidad en el acceso a la educación superior y el desempeño académico, Technical report, UNIVERSIDAD DE LOS ANDES-CEDE.
- Bécue-Bertaut, M. & Pagès, J. (2008), 'Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data', *Computational Statistics & Data Analysis* **52**(6), 3255–3268.
- Bernal, R. & Peña, X. (2011), *Guía práctica para la evaluación de impacto*, Ediciones Uniandes-Universidad de los Andes.
- Colombia, D. (2009), 'Metodología déficit de vivienda', Bogotá, Departamento Administrativo Nacional de Estadística DANE .
- Colombia, D. (2015), 'la mejor educada en el 2025', *Líneas estratégicas de la política educativa del Ministerio de Educación Nacional* .
- Corzo, J. A. (2017), 'Análisis factorial múltiple para clasificación de universidades latinoamericanas', *Comunicaciones en Estadística* **10**(1), 57–82.
- Escoufier, B. (1979), 'Traitement simultané de variables qualitatives et quantitatives en analyse factorielle', *Les cahiers de l'analyse des données* **4**, 137–146.
- Escoufier, B. & Pagès, J. (2008), *Analyses factorielles simples et multiples: objectifs, méthodes et interprétation*, Dunod.
- Escoufier, Y. (1973), 'Le traitement des variables vectorielles', *Biometrics* pp. 751–760.
- Gower, J. C. (1971), 'A general coefficient of similarity and some of its properties', *Biometrics* pp. 857–871.
- Imbens, G. W. & Lemieux, T. (2008), 'Regression discontinuity designs: A guide to practice', *Journal of econometrics* **142**(2), 615–635.
- Lê, S., Josse, J. & Husson, F. (2008), 'FactoMineR: A package for multivariate analysis', *Journal of Statistical Software* **25**(1), 1–18.
- Lee, H. & Munk, T. (2008), Using regression discontinuity design for program evaluation, in 'Proceedings of the 2008 Joint Statistical Meeting', American Statistical Association Alexandria, VA, pp. 3–7.
- Pagès, J. (2014), *Multiple factor analysis by example using R*, Chapman and Hall/CRC.
- R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<https://www.R-project.org/>
- Saporta, G. (1990), 'Probability', *Analysis of Data and Statistics, Technip* .

A. Anexos

A.1. Descripción y selección de variables

La línea base del programa SPP cuenta con un total de 1.487 jóvenes y un total de 706 variables distribuidas en los siguientes componentes:

- Identificación
- Vivienda y Hogar
- Caracterización del joven
- Preguntas para los que no solicitaron admisión a ningún programa
- Expectativas educativas, de salarios y profesionales
- Proyectos de vida de los jóvenes
- Actividades y uso del tiempo
- Orientación socio - ocupacional
- Preguntas específicas de ser pilo paga
- Datos de contacto adicionales
- (Formulario auto-reportado) Agencia y otros
- Capital social y cultural
- Preguntas específicas para todos los jóvenes que ingresaron a educación superior

De las anteriores dimensiones del instrumento se seleccionaron un total de 147 variables de interés que permitirán dar respuesta a los objetivos propuestos en esta investigación, las siguientes variables son:

1. Identificación

- Registro
- Control

2. Vivienda y hogar

- De la P14: La vivienda que habita actualmente Ud. es, P16 hasta la P22: En la vivienda que habita actualmente, dispone de:

3. Caracterización del joven

- De la P32: Su estado civil es, hasta la P36: ¿En qué actividad ocupo Ud. la mayor parte del tiempo LA SEMANA PASADA?, luego se seleccionó la P46, P49, P51, P52, P58, P65, P66 y por ultimo P72: Durante el segundo semestre de 2015, ¿Solicitó admisión a algún programa de educación superior?.

4. Expectativas educativas, de salarios y profesionales

- De la P83: Utilice la siguiente escala para evaluar las siguientes afirmaciones, P86 hasta la P88: Sobre las expectativas de su entorno respecto a la educación: Utilice la siguiente escala para evaluar las siguientes afirmaciones.

5. Proyectos de vida de los jóvenes (Aspiracional emocional)

- De la P89: Teniendo en cuenta todos los factores, ¿qué tan satisfecho se encuentra con su vida en estos momentos de acuerdo a la siguiente escala?, hasta la P94: Cuando usted tenga 30 años, ¿A qué clase social cree que va a pertenecer?

6. Actividades y uso del tiempo

- Solo se seleccionó la P95: A continuación leeré una serie de actividades y una escala de tiempos, para que me indique que tanto tiempo dedica a cada una de ellas un día normal: En un día normal entre semana (Lunes a Viernes) cuanto tiempo dedica a.

7. Agencias y otros (Comportamiento)

- De la P201: Utilice la siguiente escala para evaluar las siguientes afirmaciones, hasta la P202: Abajo encontrará una serie de declaraciones que le pueden o no aplicar, utilice la siguiente escala para evaluarlas.

8. Capital social y cultural

- De la P203: Sobre su familia: Utilice la siguiente escala para evaluar las siguientes afirmaciones, hasta la P206: Utilice la siguiente escala para evaluar las siguientes afirmaciones.

A.2. Tablas y gráficas

Tabla 3: Matriz Lg del AFMDM

Matriz Lg	Deficit de la vivienda	Caracterización del joven	Gastos del joven	Expectativas educativas	Aspiracional emocional	Pobreza subjetiva	Actividades y uso del tiempo	Comportamiento	Capital social y cultural	MFA
Deficit de la vivienda	4.70	0.06	0.11	0.04	0.05	0.10	0.08	0.01	0.07	2.01
Caracterización del joven	0.06	3.43	0.19	0.03	0.21	0.16	0.76	0.02	0.05	1.90
Gastos del joven	0.11	0.19	3.36	0.03	0.06	0.07	0.24	0.02	0.04	1.59
Expectativas educativas	0.04	0.03	0.03	2.30	0.09	0.09	0.07	0.15	0.15	1.14
Aspiracional emocional	0.05	0.21	0.06	0.09	5.69	0.34	0.22	0.05	0.11	2.64
Pobreza subjetiva	0.10	0.16	0.07	0.09	0.34	7.82	0.11	0.05	0.11	3.42
Actividades y uso del tiempo	0.08	0.76	0.24	0.07	0.22	0.11	3.47	0.04	0.13	1.98
Comportamiento	0.01	0.02	0.02	0.15	0.05	0.05	0.04	1.45	0.22	0.78
Capital social y cultural	0.07	0.05	0.04	0.15	0.11	0.11	0.13	0.22	2.16	1.17
MFA	2.01	1.90	1.59	1.14	2.64	3.42	1.98	0.78	1.17	6.42

Fuente: Elaboración propia

Tabla 4: Matriz de correlaciones Rv del AFMDM

Matriz de correlaciones RV	Deficit de la vivienda	Caracterización del joven	Gastos del joven	Expectativas educativas	Aspiracional emocional	Pobreza subjetiva	Actividades y uso del tiempo	Comportamiento	Capital social y cultural	MFA
Deficit de la vivienda	1.00	0.02	0.03	0.01	0.01	0.02	0.02	0.00	0.02	0.37
Caracterización del joven	0.02	1.00	0.06	0.01	0.05	0.03	0.22	0.01	0.02	0.41
Gastos del joven	0.03	0.06	1.00	0.01	0.01	0.01	0.07	0.01	0.01	0.34
Expectativas educativas	0.01	0.01	0.01	1.00	0.02	0.02	0.02	0.08	0.07	0.30
Aspiracional emocional	0.01	0.05	0.01	0.02	1.00	0.05	0.05	0.02	0.03	0.44
Pobreza subjetiva	0.02	0.03	0.01	0.02	0.05	1.00	0.02	0.01	0.03	0.48
Actividades y uso del tiempo	0.02	0.22	0.07	0.02	0.05	0.02	1.00	0.02	0.05	0.42
Comportamiento	0.00	0.01	0.01	0.08	0.02	0.01	0.02	1.00	0.12	0.25
Capital social y cultural	0.02	0.02	0.01	0.07	0.03	0.03	0.05	0.12	1.00	0.32
MFA	0.37	0.41	0.34	0.30	0.44	0.48	0.42	0.25	0.32	1.00

Fuente: Elaboración propia

Tabla 5: Puntuación de los jóvenes elegibles y no elegibles en cada índice

Registro	Control	Caracterización, aspiraciones y uso del tiempo	Proyección del joven	Déficit del hogar	Pobreza subjetiva	Promedio global
505	No elegible	0,68	0,67	0,14	0,17	0,41
1689	No elegible	0,49	0,86	0,33	0,06	0,43
770	Elegible	0,66	0,57	0,20	0,10	0,39
837	No elegible	0,63	0,70	0,22	0,15	0,43
1179	No elegible	0,73	0,73	0,09	0,15	0,42
633	Elegible	0,73	0,48	0,23	0,13	0,39
...
1414	No elegible	0,67	0,64	0,15	0,11	0,39
1265	No elegible	0,38	0,55	0,20	0,11	0,31
938	No elegible	0,76	0,68	0,20	0,13	0,45
1267	No elegible	0,53	0,77	0,16	0,10	0,39
1272	No elegible	0,54	0,87	0,27	0,10	0,45
1254	No elegible	0,65	0,49	0,13	0,21	0,37

Fuente: Elaboración propia

Tabla 6: Correlaciones para el grupo Déficit de la vivienda

Dim 1	R2	Dim 2	R2	Dim 3	R2
d_servicios_publicos	0,26	d_falimentos	0,27	d_vivienda	0,42
d_pisos	0,23	d_vivienda	0,21	d_paredes	0,24
d_hacinamiento	0,23	d_hacinamiento	0,20	d_hacinamiento	0,15
d_paredes	0,19	d_servicios_publicos	0,17	d_accesoatics	0,13
d_falimentos	0,18	d_pisos	0,15	d_servicios_publicos	0,06
d_accesoatics	0,18	d_accesoatics	0,10	d_pisos	0,01
d_vivienda	0,05	d_paredes	0,09	d_falimentos	0,00
Categoría	Estimate	Categoría	Estimate	Categoría	Estimate
d_servicios_publicos_1	0,5	d_falimentos_0	0,3	d_vivienda_1	0,60
d_pisos_1	1,8	d_vivienda_0	0,5	d_paredes_1	0,97
d_hacinamiento_1	0,4	d_servicios_publicos_1	0,4	d_hacinamiento_0	0,40
d_paredes_1	1,0	d_pisos_1	13,7	d_accesoatics_0	0,20
d_falimentos_1	0,2	d_accesoatics_0	0,2	d_servicios_publicos_0	0,21
d_accesoatics_1	0,3	d_paredes_1	0,6	d_pisos_1	0,34
d_vivienda_1	0,2	d_hacinamiento_0	0,4	d_falimentos_1	0,03
d_hacinamiento_2	0,1	d_paredes_0	-0,6	d_falimentos_0	-0,03
d_vivienda_0	-0,2	d_accesoatics_1	-0,2	d_pisos_0	-0,34
d_accesoatics_0	-0,3	d_pisos_0	-13,7	d_hacinamiento_1	-0,01
d_falimentos_0	-0,2	d_servicios_publicos_0	-0,4	d_servicios_publicos_1	-0,21
d_paredes_0	-1,0	d_hacinamiento_2	-0,8	d_hacinamiento_2	-0,39
d_hacinamiento_0	-0,5	d_vivienda_1	-0,5	d_accesoatics_1	-0,20
d_pisos_0	-1,8	d_falimentos_1	-0,3	d_paredes_0	-0,97
d_servicios_publicos_0	-0,5			d_vivienda_0	-0,60

Fuente: Elaboración propia

Tabla 7: Correlaciones para el grupo de Caracterización del Joven

Dim1	R2	Dim2	R2	Dim3	R2
c_ocupacionSp	0,70	c_hijos	0,59	c_getnico	0,42
c_edsuperior	0,68	c_personasacargo	0,57	c_planes	0,27
c_recibeingresos	0,35	c_edsuperior	0,10	c_discapacidad	0,21
c_personasacargo	0,19	c_ocupacionSp	0,08	c_afiliadosalud	0,20
c_hijos	0,17	c_recibeingresos	0,06	c_edsuperior	0,00
c_getnico	0,03	c_planes	0,02	c_recibeingresos	0,00
Categoría	Estimate	c_getnico	0,02	c_hijos	0,00
c_edsuperior_1	0,43	c_discapacidad	0,00	c_personasacargo	0,00
c_ocupacionSp_1	0,44	Categoría	Estimate	Categoría	Estimate
c_recibeingresos_1	0,29	c_hijos_1	1,28	c_planes_1	0,31
c_personasacargo_1	0,82	c_personasacargo_1	11,81	c_getnico_4	0,97
c_hijos_1	0,82	c_edsuperior_0	0,14	c_discapacidad_1	0,31
c_getnico_1	0,04	c_ocupacionSp_0	0,13	c_afiliadosalud_1	0,48
c_getnico_5	0,10	c_recibeingresos_0	0,10	c_getnico_5	0,30
c_getnico_3	-0,14	c_planes_1	0,10	c_edsuperior_0	0,03
c_hijos_0	-0,82	c_getnico_77	0,13	c_recibeingresos_1	0,02
c_personasacargo_0	-0,82	c_getnico_4	0,16	c_hijos_0	0,09
c_recibeingresos_0	-0,29	c_discapacidad_0	0,04	c_personasacargo_0	0,08
c_edsuperior_0	-0,43	c_discapacidad_1	-0,04	c_personasacargo_1	-0,08
c_ocupacionSp_0	-0,44	c_getnico_5	-0,29	c_hijos_1	-0,09
		c_planes_0	-0,10	c_recibeingresos_0	-0,02
		c_recibeingresos_1	-0,10	c_edsuperior_1	-0,03
		c_ocupacionSp_1	-0,13	c_getnico_3	-0,22
		c_edsuperior_1	-0,14	c_getnico_77	-0,01
		c_personasacargo_0	-11,81	c_getnico_2	-0,85
		c_hijos_0	-1,28	c_afiliadosalud_0	-0,48
				c_discapacidad_0	-0,31
				c_planes_0	-0,31

Fuente: Elaboración propia

Tabla 8: Correlaciones para el grupo Aspiracional Emocional

Dim1	R2	Dim2	R2	Dim3	R2
A_nivelmaximoG	0,81	A_nivelmaximoG	0,82	A_nivelmaximoG	0,80
A_nivelmaximoC	0,84	A_nivelmaximoC	0,83	A_nivelmaximoC	0,80
Emo_satisfecho,vida	0,15	Emo_satisfecho,vida	0,01	Emo_satisfecho,vida	0,01
Categoría	Estimate	Categoría	Estimate	Categoría	Estimate
A_nivelmaximoC_4	0,1	A_nivelmaximoG_8	50,9	A_nivelmaximoC_6	11,7
A_nivelmaximoG_4	0,7	A_nivelmaximoC_2	31,1	A_nivelmaximoG_6	0,9
A_nivelmaximoG_8	17,8	A_nivelmaximoC_3	17,6	A_nivelmaximoG,NA	1,7
A_nivelmaximoC_2	15,6	A_nivelmaximoG_4	0,0	Emo_satisfecho,vida_3	-0,1
A_nivelmaximoC_3	1,0	A_nivelmaximoC_1	30,1	A_nivelmaximoG_4	-0,6
Emo_satisfecho,vida_3	0,4	Emo_satisfecho,vida_3	0,2	A_nivelmaximoC_4	-0,3
Emo_satisfecho,vida_4	0,1	Emo_satisfecho,vida_4	0,0	A_nivelmaximoG_7	-0,5
Emo_satisfecho,vida_2	0,3	A_nivelmaximoC_6	-22,0	A_nivelmaximoC_7	-0,3
A_nivelmaximoC_1	0,6	A_nivelmaximoC_4	-14,0	A_nivelmaximoC_5	-0,8
A_nivelmaximoC_6	-0,9	A_nivelmaximoG_6	-12,5	A_nivelmaximoG_5	-12,3
A_nivelmaximoG_6	-0,5	A_nivelmaximoG_7	-0,6		
A_nivelmaximoC_5	-0,6	A_nivelmaximoC_7	-15,7		
Emo_satisfecho,vida_5	-0,5	A_nivelmaximoG_5	-17,7		
A_nivelmaximoG_5	-0,1	A_nivelmaximoC_5	-27,2		
A_nivelmaximoC_7	-17,5				
A_nivelmaximoG_7	-13,6				

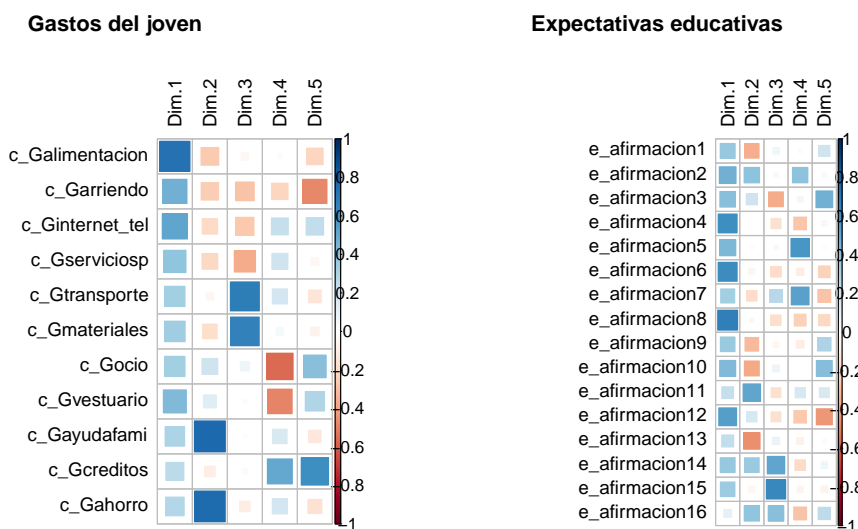
Fuente: Elaboración propia

Tabla 9: Correlaciones para el grupo de Pobreza Subjetiva

Dim1	R2	Dim2	R2	Dim3	R2	Dim5	R2
Emo_clasnew	0,58	Emo_clasS	0,52	Emo_clasnew	0,46	Emo_30Aingresos	0,53
Emo_30Aingresos	0,46	Emo_considerapobre	0,44	Emo_30Aingresos	0,43	Emo_clasnew	0,37
Emo_clasS	0,32	Emo_clasnew	0,17	Emo_clasS	0,23	Emo_clasS	0,18
Emo_considerapobre	0,11	Emo_30Aingresos	0,18	Emo_considerapobre	0,07	Emo_considerapobre	0,01
Categoría	Estimate	Categoría	Estimate	Categoría	Estimate	Categoría	Estimate
Emo_clasnew_3	0,33	Emo_considerapobre_1	0,56	Emo_clasnew_4	1,6	Emo_clasnew_5	6,86
Emo_considerapobre_1	0,29	Emo_clasS_5	0,47	Emo_clasS_1	2,4	Emo_30Aingresos_6	1,07
Emo_30Aingresos_5	0,16	Emo_clasnew_4	0,87	Emo_30Aingresos_7	1,2	Emo_30Aingresos_4	0,67
Emo_clasS_5	0,9	Emo_30Aingresos_1	0,49	Emo_considerapobre_0	0,2	Emo_clasS_2	0,85
Emo_clasS_4	0,65	Emo_clasS_1	0,98	Emo_30Aingresos_8	2,6	Emo_considerapobre_0	0,07
Emo_clasS_3	0,3	Emo_30Aingresos_2	0,33	Emo_30Aingresos_2	-0,7	Emo_clasS_5	0,03
Emo_30Aingresos_7	0,87	Emo_30Aingresos_7	0,68	Emo_clasS_2	-0,4	Emo_considerapobre_1	-0,07
Emo_30Aingresos_6	0,21	Emo_30Aingresos_3	0,22	Emo_clasS_5	-0,7	Emo_clasS_1	-0,67
Emo_clasnew_5	0,83	Emo_clasS_4	-0,4	Emo_clasS_3	-0,5	Emo_30Aingresos_8	-1,70
Emo_clasnew_4	-0,1	Emo_30Aingresos_8	-1,4	Emo_30Aingresos_4	-0,8	Emo_30Aingresos_7	-0,28
Emo_30Aingresos_4	-0,06	Emo_clasS_2	-0,26	Emo_clasS_4	-0,7	Emo_clasS_3	-0,18
Emo_clasS_1	-1,31	Emo_30Aingresos_6	-0,2	Emo_30Aingresos_6	-0,3	Emo_30Aingresos_5	-0,10
Emo_30Aingresos_2	-0,6	Emo_clasnew_1	-0,13	Emo_30Aingresos_1	-0,4	Emo_30Aingresos_3	-0,08
Emo_clasnew_2	-0,16	Emo_30Aingresos_5	-0,21	Emo_30Aingresos_5	-0,4		
Emo_considerapobre_0	-0,29	Emo_clasnew_2	-0,55	Emo_considerapobre_1	-0,2		
Emo_clasS_2	-0,53	Emo_clasS_3	-0,79	Emo_clasnew_2	-0,8		
Emo_30Aingresos_1	-1,08	Emo_considerapobre_0	-0,56	Emo_30Aingresos_3	-1,1		
Emo_clasnew_1	-0,9			Emo_clasnew_3	0		

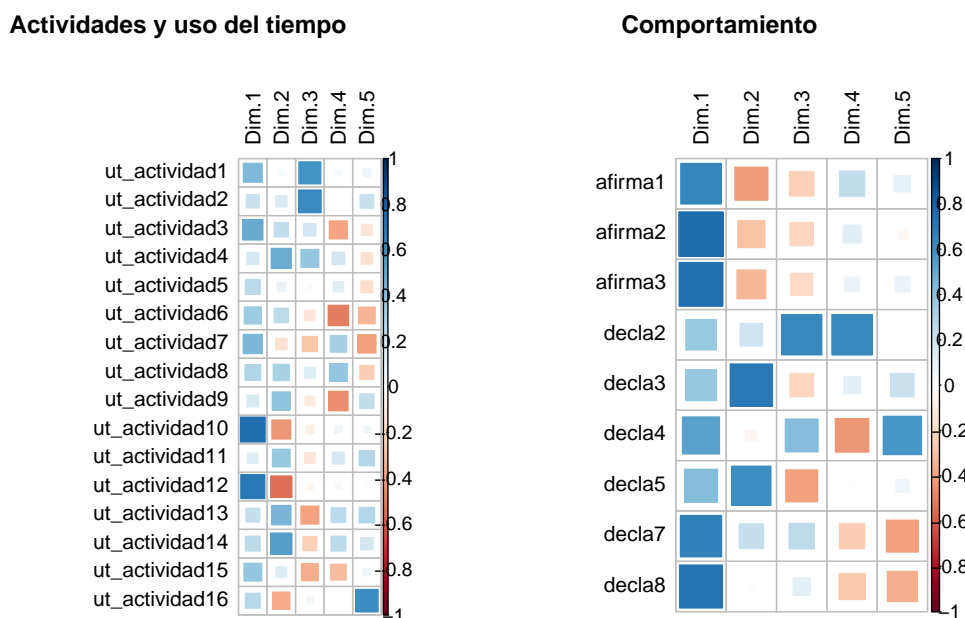
Fuente: Elaboración propia

Figura 7: Correlaciones grupo 3 y grupo 4



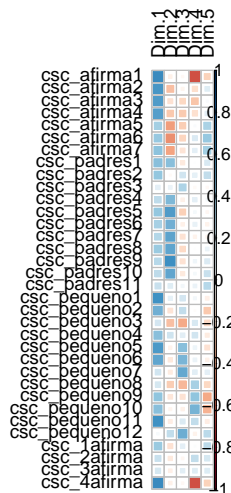
Fuente: Elaboración propia

Figura 8: Correlaciones grupo 7 y grupo 8



Fuente: Elaboración propia

Figura 9: Correlaciones grupo 9
Capital social y cultural



Fuente: Elaboración propia

A.3. Código Computacional

```
#####
##### Código Análisis Factorial Múltiple para datos mixtos - SPP #####
#####

#####
### Librerías #####
#####

library(haven)
library(dplyr)
library(FactoMineR)
library(factoextra)
library(ggplot2)

#####
## Directorio de trabajo ##
#####

setwd("../")

SPP <- read_dta("../\\LINEA BASE SPP - ANONIMIZADA.dta")

#####
##### Selección de variables definidas por dirección del constructo y aporte #####
#####

base_final <- data.frame(registro=SPP$registro, Control=SPP$P1, d_vivienda,
```

```

d_paredes, d_pisos, d_servicios_publicos, d_falimentos, d_hacinamiento,
d_accesoaotics, c_getnico, c_hijos, c_afiliadosalud, c_discapacidad,
c_ocupacionSp, c_personasacargo, c_edsuperior, c_planes,
c_recibeingresos,c_Galimentacion, c_Garriendo, c_Ginternet_tel,
c_Gserviciosp, c_Gtransporte, c_Gmateriales, c_Gocio, c_Gvestuario,
c_Gayudafami, c_Gcreditos, c_Gahorro, e_afirmacion1, e_afirmacion2,
e_afirmacion3, e_afirmacion4, e_afirmacion5,e_afirmacion6, e_afirmacion7,
e_afirmacion8, e_afirmacion9, e_afirmacion10, e_afirmacion11,
e_afirmacion12, e_afirmacion13, e_afirmacion14, e_afirmacion15,
e_afirmacion16, A_nivelmaximoG, A_nivelmaximoC, Emo_satisfecho.vida,
Emo_30Aingresos, Emo_claseS, Emo_considerapobre, Emo_clasenew, ut_actividad1,
ut_actividad2, ut_actividad3, ut_actividad4, ut_actividad5, ut_actividad6,
ut_actividad7, ut_actividad8,ut_actividad9, ut_actividad10, ut_actividad11,
ut_actividad12, ut_actividad13, ut_actividad14, ut_actividad15,ut_actividad16,
afirma1, afirma2, afirma3,decla2, decla3, decla4, decla5, decla7, decla8,
csc_afirma1, csc_afirma2, csc_afirma3, csc_afirma4,csc_afirma5, csc_afirma6,
csc_afirma7, csc_padres1,csc_padres2, csc_padres3, csc_padres4, csc_padres5,
csc_padres6, csc_padres7, csc_padres8, csc_padres9, csc_padres10, csc_padres11,
csc_pequeno1, csc_pequeno2, csc_pequeno3, csc_pequeno4, csc_pequeno5,
csc_pequeno6, csc_pequeno7, csc_pequeno8, csc_pequeno9, csc_pequeno10,
csc_pequeno11, csc_pequeno12, csc_1afirma, csc_2afirma, csc_3afirma, csc_4afirma
)

```

```

# Deficit de la vivienda
base_final[,3:9] <- lapply(base_final[,3:9], as.factor)
# Caracterización del joven
base_final[,10:18] <- lapply(base_final[,10:18], as.factor)
# Gastos del joven
base_final[,19:29] <- lapply(base_final[,19:29], as.numeric)
# Expectativas educativas
base_final[,30:45] <- lapply(base_final[,30:45], as.numeric)
# Aspiracional emocional
base_final[,46:48] <- lapply(base_final[,46:48], as.factor)
# Pobreza subjetiva
base_final[,49:52] <- lapply(base_final[,49:52], as.factor)
# Actividades y uso del tiempo
base_final[,53:68] <- lapply(base_final[,53:68], as.numeric)
# Comportamiento
base_final[,69:77] <- lapply(base_final[,69:77], as.numeric)
# Capital social y cultural
base_final[,78:111] <- lapply(base_final[,78:111], as.numeric)

```

```

#####
#### Análisis Factorial Múltiple para datos mixtos ##
#####

```

```

base_final <- base_final[, !colnames(base_final)=="Control"]
base_general <- base_final[,-1]
rownames(base_general)<-base_final[,1]

```

```

afm_global <- MFA(base_general, group = c(7,9,11,16,3,4,16,9,34),
type=c("n","n","s","s","n","n","s","s","s"),

```

```
name.group = c("Deficit de la vivienda",  
              "Caracterización del joven",  
              "Gastos del joven",  
              "Expectativas educativas",  
              "Aspiracional emocional",  
              "Pobreza subjetiva",  
              "Actividades y uso del tiempo",  
              "Comportamiento",  
              "Capital social y cultural"),  
graph = T)
```