



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado

**CARRERA DE ESPECIALIZACIÓN EN MÉTODOS
CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE
DATOS EN ORGANIZACIONES**

TRABAJO FINAL DE ESPECIALIZACIÓN

Pobreza multidimensional en Argentina. Análisis
factorial de la Encuesta Permanente de Hogares.

AUTOR: NÉLIDA MÓNICA CANTONI RABOLINI

[DICIEMBRE 2019]



Resumen

Este trabajo se realiza para caracterizar la pobreza multidimensional en Argentina desde el enfoque del análisis factorial de la Encuesta Permanente de Hogares (EPH).

El objetivo general de este trabajo es identificar las principales características relacionadas con la población bajo la línea de pobreza en Argentina utilizando análisis factorial. Los objetivos específicos son: clasificar a los individuos en su situación de pobreza, determinar las tasas de pobreza e indigencia en la población a partir de los microdatos de la EPH, caracterizar el perfil sociodemográfico de la población bajo la línea de pobreza, aplicar el método de componentes principales, análisis discriminante y análisis factorial en lenguaje R para determinar los factores relacionados con la pobreza.

La hipótesis que se pretende contrastar es si las características sociodemográficas, las condiciones de habitabilidad, el nivel educativo alcanzado, los ingresos y la ocupación son los factores que se relacionan con la población bajo la línea de pobreza en Argentina.

Se realizará un estudio descriptivo y correlacional de diseño no experimental, longitudinal y retrospectivo. La unidad de análisis son los hogares compuestos por individuos que respondieron a la EPH. Se hicieron dos mediciones de cada hogar una en el primer trimestre y otra en el segundo trimestre de 2019.

Para el cálculo de la pobreza en Argentina actualmente se utiliza el indicador línea de pobreza. Dado que el abordaje de la pobreza es complejo, se propone determinar qué otros aspectos están relacionados con la línea de pobreza además de los ingresos de los individuos y hogares. Para dar respuesta a este problema se realizará un estudio multidimensional de las variables de la EPH aplicando el método de componentes principales, análisis discriminante y análisis factorial.

Con estos resultados, se podrá determinar qué variables están relacionadas con la situación de los individuos en relación con la línea de pobreza, también se podrán realizar clasificaciones y predicciones aplicando nuevos casos a cada uno de los modelos utilizados en este trabajo.

Palabras clave: Pobreza, EPH, pobreza multidimensional, análisis discriminante, análisis factorial.



Estructura

Introducción.....	4
1. Un análisis factorial para medir la pobreza multidimensional	6
1.1. Métodos de medición de pobreza	6
1.2. Factores causales que influyen en la condición de pobreza	8
1.3. Indicadores para medir pobreza.....	9
2. Desarrollo de un modelo para caracterizar la pobreza e indigencia.....	10
2.1. Tipo de estudio, población y muestra.....	10
2.2. Instrumento de recolección de datos	11
2.3. Base de datos y programación en R.....	12
3. Pobreza e indigencia primer semestre 2019	13
3.1. Caracterización de la pobreza e indigencia	13
3.2. Componentes principales. Perfiles para categorizar a los individuos	15
3.3. Análisis discriminante. Clasificación y predicción	16
3.4. Análisis factorial. Factores que influyen en la pobreza.....	18
Conclusión.....	20
Referencias bibliográficas	22
Apéndice 1. Código en R Caracterización de la pobreza e indigencia.....	24
Apéndice 2. Código en R Componentes Principales.....	34
Apéndice 3. Código en R Análisis Discriminante.....	35
Apéndice 4. Código en R Análisis Factorial	37



Introducción

Argentina posee un alto índice de pobreza (23,4% hogares pobres en el segundo semestre 2018). Para medir la pobreza se utiliza el método de medición indirecto denominado línea. El indicador denominado línea de pobreza (LP) se calcula a través de datos proporcionados por la Encuesta Permanente de Hogares (EPH) que realiza el INDEC y las canastas básicas. Esta medición determina si los hogares y las personas, a partir de sus ingresos, tienen la capacidad para satisfacer las necesidades de consumos alimentarios mínimos y básicos no alimentarios. (INDEC, 2019)

Se han encontrado diversos estudios que abordan la medición de la pobreza. Un estudio realizado en Honduras en 2015 con el objetivo de identificar los principales factores que explican la pobreza aplicó el análisis de componentes principales y la creación de un modelo factorial. Los resultados obtenidos indican que la pobreza en Honduras está caracterizada además de los factores económicos por las características demográficas, las condiciones de habitabilidad de los hogares, la formación de capital humano y el mercado del trabajo. En Guatemala en 2009, se realizó un análisis multifactorial desde la perspectiva económica para identificar pobreza. Se utilizó la metodología de análisis factorial y de componentes principales para construir un índice de capacidades, para luego compararlo con otras medidas convencionales de la pobreza, entre ellas las basadas en los ingresos o en el acceso a infraestructura básica. Se concluyó que existen múltiples formas de usar los instrumentos existentes para analizar la pobreza y las capacidades, pero es fundamental mejorar la información estadística, reforzar los instrumentos, y asegurar su comparabilidad en el tiempo. Por lo tanto, uno de los desafíos es como medir la pobreza multidimensional, esto es identificar los factores multicausales que influyen en las condiciones de pobreza en Argentina. (Ruano, 2015; Gammage, 2009)

La caracterización de la población según la Encuesta Permanente de Hogares en Argentina, en relación con la línea de pobreza, a través de la utilización de la metodología de análisis factorial, es un aporte al concepto de pobreza multidimensional. En este sentido los aspectos a considerar son: las características sociodemográficas de la población estudiada, condiciones de habitabilidad, nivel educativo alcanzado, ingresos y ocupación.



Este trabajo se realiza para conocer: ¿Cuáles son las principales características relacionadas con la población bajo la línea de pobreza en Argentina?

El objetivo general es identificar las principales características relacionadas con la población bajo la línea de pobreza en Argentina utilizando análisis factorial.

Los objetivos específicos son: categorizar a los individuos según su situación de pobreza, determinar las tasas de pobreza e indigencia en la población a partir de los microdatos de la EPH, caracterizar el perfil sociodemográfico de la población bajo la línea de pobreza, aplicar el método de componentes principales, análisis discriminante y análisis factorial en R para determinar los factores relacionados con la pobreza.

La hipótesis que se pretende contrastar es si las características sociodemográficas, las condiciones de habitabilidad, el nivel educativo alcanzado, los ingresos y la ocupación son los factores que se relacionan con la población bajo la línea de pobreza en Argentina.

En el primer apartado de este trabajo, se abordará el concepto de pobreza, explicando el cálculo de la línea de pobreza. En cuanto al concepto de pobreza multidimensional, se abordará el enfoque a través de los métodos de análisis multivariado, considerando el análisis de componentes principales, análisis discriminante y análisis factorial.

En el segundo apartado se presentará un modelo para caracterizar la pobreza e indigencia, se especificará la metodología a utilizar, se describirán los instrumentos de recolección de datos, las técnicas de análisis de datos y programación en lenguaje R.

En el tercer apartado se presentarán los resultados de la pobreza e indigencia para el primer semestre de 2019, caracterizando a la población bajo la línea de pobreza e indigencia, especificando los factores que influyen en la situación de pobreza y proponiendo formas de clasificación y predicción.

Finalmente se presentarán las conclusiones del trabajo, los aportes, las propuestas y líneas de trabajo futuro.



1. Un análisis factorial para medir la pobreza multidimensional

En las últimas décadas ha tenido gran relevancia las mediciones de pobreza y las políticas tendientes a combatir esta problemática. Uno de los Objetivos de Desarrollo del Milenio insta a los países adscriptos a reducir a la mitad el nivel de pobreza, en este aspecto muchas políticas públicas, como programas y planes en diferentes países de Latino América, se han orientado a reducir el porcentaje de personas en condiciones de vida precarias. Considerando la importancia que tiene la pobreza como problema social en el mundo, un análisis de este fenómeno aportaría un conocimiento para mejorar las condiciones de vida de las personas. (Conconi, 2011; Torres, 2019)

1.1. Métodos de medición de pobreza

Las tradicionales mediciones de pobreza basadas únicamente en privaciones de ingreso o necesidades básicas insatisfechas (NBI) resultan limitadas, puesto que la pobreza va más allá de estas definiciones. Estas estimaciones se fijan en términos de reducir el porcentaje de hogares (o individuos) por debajo de determinado umbral. Se considera que un individuo puede enfrentar privaciones en diversas dimensiones de su vida, como en su educación, capacidad de obtener un empleo y condiciones de salud, que no son contempladas en las mediciones tradicionales de pobreza. (Conconi, 2011)

La medición de la pobreza tiene dos dimensiones. La primera tiene que ver con la concepción de la pobreza, que pone el énfasis en el ingreso, y la segunda con la determinación de la línea de pobreza, que se refiere a las capacidades y realizaciones. Una medida es más acotada que un concepto, en lo referente al indicador de pobreza. Se presenta la inquietud de si es posible ir más allá de estas dimensiones, formulando indicadores multidimensionales que describan las condiciones sociales de las personas. Actualmente, existe cierto consenso de que la pobreza es multidimensional y que las futuras mediciones deberían orientarse a ampliar la noción de pobreza. (González, 2011; Conconi, 2011)

En este sentido, diferentes estudios mencionados por González hacen referencia a que variables se incluyeron en los estudios multidimensionales de la pobreza. En Estados Unidos se estudiaron variables como ingreso, salud, servicio médico y escolaridad. En Indonesia consideraron las variables gasto, índice de masa corporal, escolaridad, provisión



de agua potable y saneamiento básico. En seis países de América Latina que abordaron pobreza multidimensional, las variables estudiadas fueron ingreso, asistencia escolar de los niños, educación, agua, saneamiento, vivienda. (González, 2011)

En Colombia se calculó un índice que incluye 15 variables relacionadas con educación, trabajo, salud, vivienda y servicios públicos. Se considera que un hogar es pobre cuando tiene cinco carencias según esas 15 variables. Los distintos enfoques de la pobreza multidimensional tienen un elemento en común: que el ingreso es necesario, pero no es suficiente para conseguir el bienestar. (González, 2011)

En Argentina, el indicador de pobreza se mide desde 2016 con una nueva metodología. Los resultados del segundo semestre de 2018, muestran que el porcentaje de hogares por debajo de la línea de pobreza es 23,4%; estos comprenden el 32% de las personas. Dentro de este conjunto se distingue un 4,8% de hogares indigentes que incluyen el 6,7% de las personas. (INDEC, 2019)

Actualmente para el cálculo de la pobreza se usa el método de medición indirecto denominado línea. La línea de pobreza (LP) se define por medio de datos proporcionados a través de una encuesta permanente que se realiza a hogares y personas. Esta encuesta, denominada Encuesta Permanente de Hogares (EPH), se utiliza para medir indicadores en la población de referencia. Para el indicador LP, la medición se realiza para determinar si la población, a partir de los ingresos totales, tiene la capacidad para satisfacer – por la compra de bienes y servicios- las necesidades de consumos alimentarios mínimos y los consumos básicos no alimentarios. (INDEC, 2016)

Para determinar si las personas pueden satisfacer sus necesidades, se establece la canasta básica alimentaria (CBA) que es normalizada de acuerdo con las necesidades nutricionales de la población de referencia, donde se han incluido últimamente las canastas regionales, y que marca un mínimo al que se debe alcanzar para no calificar en la línea de pobreza. Para expandir el valor de CBA se utiliza el Coeficiente de Engel (CdE). Al mismo tiempo se considera la Canasta Básica Total (CBT) que es la que engloba tanto los hábitos de consumo alimentario (CBA) como los no alimentario de la población de referencia.

Las fuentes utilizadas para elaborar el indicador línea de pobreza (LP) son las bases de datos de la (EPH). Los datos sociodemográficos se cruzan con otras encuestas como los Gastos en los Hogares (ENGHo), para determinar perfiles de consumo, con las tablas de



requerimientos nutricionales según sexo y edad, con las canastas CBA y CBT y el Índice de Precios al Consumidor (IPC). Con el análisis de toda la información vinculada se determinan la situación de cada individuo con relación a la línea de pobreza y se calculan las tasas de pobreza e indigencia. (INDEC, 2016).

En Argentina la medición de la pobreza e indigencia se realiza por medio de los ingresos totales de los hogares. En este sentido, considerando las probables inconsistencias que pueden darse cuando un individuo responde sobre sus ingresos, resulta necesario plantear otros enfoques adicionales de medición de pobreza. Por ejemplo, análisis que observen, además de los ingresos, las características individuales de las personas y de los hogares donde viven, con el objetivo de conocer los factores que influyen en la condición de pobreza.

1.2. Factores causales que influyen en la condición de pobreza

El concepto de pobreza multidimensional se puede abordar a través de la identificación de factores multicausales que influyen en las condiciones de pobreza en los hogares y los individuos de Argentina.

En Colombia se elaboró un documento con el objetivo central de proponer una medición alternativa de la pobreza multidimensional, para ello se utilizará el indicador de Bourguignon y Chakravarty, identificando a los hogares pobres a partir de dimensiones, variables y parámetros de estimación no utilizados en la metodología convencional. La pobreza va más allá de la carencia de ingresos. La pobreza multidimensional resulta ser un enfoque que permite identificar los déficits que tiene la población en ciertas dimensiones, y proponer políticas que permitan superar esa vulnerabilidad social. De este modo, se plantea la imperiosa necesidad de generar políticas que den oportunidades para que la población vulnerable supere los mínimos umbrales en todas las dimensiones básicas del bienestar. Mayor acceso y calidad de la educación, garantizar el acceso de la población al sistema de salud, mejorar el entorno ambiental, programas de educación sexual y reproductiva, políticas transversales y focalizadas de empleo, alternativas de empleo asociativo en entornos populares y rurales que logren reducir la vulnerabilidad económica, son algunas opciones que podrían contribuir al mejoramiento de los niveles de vida de la población y la generación de más y mejores oportunidades. (Torres, 2019)



1.3. Indicadores para medir pobreza

Como indicadores para medir la pobreza se puede mencionar el modelo de análisis factorial, que es un modelo de regresión múltiple que relaciona variables latentes con variables observadas. Tiene muchos puntos en común con el análisis de componentes principales, y busca esencialmente nuevas variables o factores que expliquen los datos. Cabe destacar la diferencia entre componentes principales y análisis factorial. Mientras que en el análisis de componentes principales las variables se agrupan para formar los componentes, por medio de combinaciones lineales ponderadas de las variables, reduciendo la dimensionalidad del problema, en el análisis factorial exploratorio los factores buscan las variables que mejor explican el comportamiento de cada factor. El análisis factorial exploratorio intenta detectar si hay variables latentes (no observadas) que explican porque los indicadores están correlacionados entre sí y si pueden agruparse para reducir los datos. Los factores muestran el comportamiento latente frente a lo observado de las variables. (Manzano y Jiménez, 2017)

Un estudio realizado en España en 2010 con el objetivo de construir un indicador de pobreza multidimensional que permita detectar los principales focos de riesgo de pobreza en España. Al estructurar las diversas posibilidades metodológicas que ofrece el estudio de la pobreza multidimensional, concluyó que la utilización del análisis factorial ha proporcionado un indicador de dimensión objetiva compuesta por las variables relacionadas con la renta individual, del hogar, actual y pasada y otras subjetivas como la satisfacción laboral, la situación económica, la vivienda y el ocio. El indicador posee ventajas: ofrece las tres vertientes fundamentales de los indicadores de pobreza: extensión, intensidad y desigualdad; utiliza un sistema de ponderación no arbitrario; tiene carácter multidimensional y latente –incorporando dimensiones objetivas y subjetivas–, por lo que permite acercarnos en gran medida al abstracto concepto de la pobreza. Como restricciones: se debe ser preciso a la hora de interpretarlo, por lo que no muestra el grado de pobreza de una persona sino el riesgo que tiene de incurrir en ella. (Poza Lara, 2010)

A continuación, se presenta la metodología utilizada para relacionar la situación de cada individuo con relación a la pobreza y las características del hogar donde vive. En primer lugar, se realizará un estudio descriptivo de los individuos categorizados como indigentes,



pobres (no indigentes) y no pobres relacionado con sus características individuales. Posteriormente se trabajará con el análisis de componentes principales, un análisis discriminante y finalmente el análisis factorial, utilizando como soporte informático R.

2. Desarrollo de un modelo para caracterizar la pobreza e indigencia.

En este apartado se describe el tipo de investigación, la población y la muestra utilizada para realizar el trabajo. Se describen los instrumentos de recolección de datos, las fuentes utilizadas, las bases de datos generadas a partir de las fuentes disponibles. Además, se explicará el procesamiento de los datos en lenguaje de programación R así como los métodos de análisis estadístico aplicados para alcanzar los objetivos del trabajo.

2.1. Tipo de estudio, población y muestra.

Se realizará un estudio cuantitativo, descriptivo y correlacional. El diseño es no experimental, transversal y retrospectivo. La unidad de análisis es la información procedente de la Encuesta Permanente de Hogares EPH correspondiente al primer semestre de 2019. Las principales variables son: Características de la vivienda. Características habitacionales del hogar. Estrategias del hogar. Resumen del hogar. Ingreso total familiar. Ingreso per cápita familiar. Organización del hogar. Características de los miembros del hogar. Nivel educativo. Condición de actividad. Condición de inactividad. Ingresos. Los indicadores son: Línea de pobreza, Ingresos de los hogares.

La población son los hogares y los individuos que componen los hogares de los aglomerados urbanos de la República Argentina. Se utilizó un muestreo probabilístico por conglomerados. La cobertura geográfica es de 31 aglomerados agrupados de la siguiente manera:

Región Gran Buenos Aires: Ciudad Autónoma de Buenos Aires; Partidos del Gran Buenos Aires.

Región Cuyo: Gran Mendoza; Gran San Juan; Gran San Luis.

Región Nordeste (NEA): Corrientes; Formosa; Gran Resistencia; Posadas.

Región Noroeste (NOA): Gran Catamarca; Gran Tucumán-Tafí Viejo; Jujuy-Palpalá; La Rioja. Salta; Santiago del Estero-La Banda.



Región Pampeana: Bahía Blanca-Cerri; Concordia; Gran Córdoba; Gran La Plata; Gran Rosario. Gran Paraná; Gran Santa Fe; Mar del Plata; Río Cuarto; Santa Rosa-Toay; San Nicolás-Villa Constitución.

Región Patagónica: Comodoro Rivadavia-Rada Tilly; Neuquén-Plottier; Río Gallegos; Ushuaia-Río Grande; Rawson-Trelew; Viedma-Carmen de Patagones.

2.2. Instrumento de recolección de datos

En cuanto a las unidades de respuesta, son los individuos de cada uno de los hogares seleccionados para responder a la encuesta. Para recolectar los datos se utilizaron los microdatos de la encuesta EPH, esta encuesta utiliza un cuestionario estructurado que consta de 88 preguntas para nivel hogar y 177 preguntas para nivel individual.

Se trabajó con datos secundarios, los microdatos correspondientes al primer y segundo trimestre de 2019 de la EPH, disponibles en formato de archivo .txt y .xls.

Las bases de datos poseen una significativa cantidad de variables de hogar y personas para posibilitar el análisis de las principales características demográficas y socioeconómicas de la población. Cada registro tiene un número de identificación (CODUSU), que permite relacionar una vivienda con los hogares y personas que la componen. Todos los hogares que pertenecen a una misma vivienda poseen el mismo CODUSU. Para identificar los hogares se debe utilizar CODUSU y NRO_HOGAR. En la base de datos de personas todos los miembros del hogar tienen el mismo CODUSU y NRO_HOGAR pero se diferencian por el número de COMPONENTE.

Con los microdatos de primer y segundo trimestre de 2019 se elaboraron las bases de datos de hogares y de individuos, correspondientes al primer semestre de 2019 uniendo los datos de primer y segundo trimestre de 2019. La base de datos resultante de hogares posee 38589 observaciones de 88 variables y la de individuos posee 118627 observaciones de 177 variables. Los hogares participan de la EPH durante los 4 trimestres. En este estudio se utilizaron los datos del primer y segundo trimestre, es decir cada hogar se encuestó dos veces. De este modo se puede observar cómo se modifica la situación de cada hogar en cuanto a su condición de pobreza.

Finalmente, luego de realizar cálculos y clasificación de individuos, se unieron las bases de datos de individuos y hogar del primer trimestre con la de individuos y hogar del



segundo trimestre generando una nueva base de datos con 118627 observaciones. Se incluyeron solamente las variables consideradas para el análisis tanto de individuos como de hogar resultando 65 variables entre las originales y las creadas para este trabajo.

La fuente de dato primario es el INDEC que pone a disposición de los usuarios las bases de microdatos de la EPH para ser utilizados con otros propósitos de análisis. En este caso se aplica el concepto de reutilización donde los datos se han recolectado en un contexto de privacidad y se emplearán en un ámbito de privacidad diferente del que fueron originados. (Steinmann, Matei, Collmann, 2016)

2.3. Base de datos y programación en R

Para procesar y analizar los datos se utilizó el programa R Studio versión 3.5.3 (2019) y R Commander versión 2.6-0 (2019)

Se instalaron todos los paquetes necesarios para trabajar y se cargaron las librerías relacionadas. Se actualizó la información de las canastas básicas por región CBA y CBT con los datos de los primeros seis meses de 2019 y se creó un archivo en formato .xls con estos datos. También se creó un archivo .xls con los datos de requerimientos nutricionales para adulto equivalente según sexo y edad de los individuos. Se crearon archivos de códigos con los nombres de regiones y aglomerados.

En primer lugar, se cargaron todos los archivos necesarios para realizar los cálculos: bases de datos de personas y hogar para cada trimestre, canastas, requerimientos nutricionales, regiones y aglomerados.

En segundo lugar, se relacionaron las bases de datos de individuos y hogar de primer y segundo trimestre. Se generaron nuevos archivos con las bases de datos.

Para el cálculo de pobreza e indigencia, se realizaron varios procedimientos utilizando las bases de datos de personas y las canastas por regiones calculando para cada individuo su índice de requerimientos nutricionales según adulto equivalente. Se calculó la canasta básica alimentaria y canasta básica total para cada hogar. Comparando los ingresos totales de los hogares (ITF) con las canastas CBA y CBT se determinó la situación de pobreza de cada individuo (indigente – pobre – no pobre).

Con la clasificación de cada individuo en relación con la situación de pobreza, se calcularon las tasas de pobreza (incluyendo a la indigencia) e indigencia por región para cada



periodo. También se calcularon tasas de pobreza e indigencia por región según sexo, nivel educativo, estado ocupacional, nivel de actividad, nivel de inactividad y grupos de edad. Tal como se presenta el código en R (Apéndice 1).

Obtenida la base final con la clasificación de cada individuo y las características del hogar donde vive, se realizó un análisis de componentes principales (Apéndice 2) para reducir las variables, análisis discriminante para predecir situación de pobreza (Apéndice 3) y análisis factorial para determinar los factores que influyen en la clasificación de los individuos (Apéndice 4).

3. Pobreza e indigencia primer semestre 2019

Luego de realizar los procedimientos de análisis de datos se presentarán los resultados obtenidos. En primer lugar, se caracterizará a la población según su situación en relación con la línea de pobreza. Para realizar dicha caracterización, se procederá a realizar diferentes cálculos, teniendo en cuenta los requerimientos nutricionales según adulto equivalente para cada individuo de la muestra y comparando los ingresos totales con las canastas básicas alimentaria y total. De este modo se obtiene tres categorías: indigente, pobre (no indigente) y no pobre. A partir de esta categorización, se determinarán las tasas de pobreza (incluye a los indigentes) e indigencia. Considerando las variables seleccionadas tanto de individuos como del hogar donde viven, se presentarán los perfiles para categorizar a los individuos bajo la línea de pobreza utilizando el método de componentes principales, posteriormente se realizará la clasificación y predicción mediante análisis discriminante y por último se presentarán los factores que influyen en la pobreza a través de análisis factorial.

3.1. Caracterización de la pobreza e indigencia

Se realizó un análisis descriptivo de los datos de la base individuos para el primer y segundo trimestre de 2019, determinando las tasas de pobreza e indigencia expresadas en porcentajes agrupadas por región, sexo, grupos de edad, nivel educativo, nivel de ocupación, categoría de actividad y categoría de inactividad.

Para el primer semestre de 2019 el porcentaje de personas por debajo de la línea de pobreza es de 35,3%, dentro de este conjunto se distinguen un 7,61% de personas indigentes.



Los datos por trimestre son: para el primer trimestre el porcentaje de personas por debajo de la línea de pobreza es del 34,44%, dentro de este conjunto se distinguen un 7,1% de personas indigentes; para el segundo trimestre el porcentaje por debajo de la línea de pobreza es del 36,44% de las personas. Dentro de este conjunto se incluyen el 8,12% de personas indigentes.

La tasa de pobreza e indigencia por región para el primer y segundo trimestre de 2019, se distingue la región Noreste del país en el primer trimestre un 42,49% de personas bajo la línea de pobreza incluyendo un 10,66% de personas indigentes, para el segundo trimestre un 42,75% de personas bajo la línea de pobreza incluyendo un 11,30% de personas indigentes. Luego sigue la región Noroeste con 37,45% de personas bajo la línea de pobreza incluyendo a 6,33% de personas indigentes para el primer trimestre, aumentando en el segundo trimestre a 42,57% y 4,65% respectivamente. Las tasas más bajas se observan en región Patagonia con 26,61% de personas bajo la línea de pobreza incluyendo a 3,77% de personas indigentes durante el primer trimestre, aumentando a 31,69% y 4,01% respectivamente.

El aglomerado con mayor tasa de pobreza es Concordia con 53,06% de pobres incluyendo a 14,96% de indigentes para el primer trimestre y 51,82% de pobres incluyendo a 15,2% de indigentes para el segundo trimestre de 2019.

La menor tasa de pobreza se encuentra en Ciudad de Buenos Aires, para el primer trimestre se observa 13,44% de pobres incluyendo 3,2% de indigentes, siendo para el segundo trimestre 14,1% de pobres incluyendo 6,31% de indigentes.

También se observan diferencias de un punto porcentual entre las tasas de pobreza e indigencia según sexo, donde resulta más alta para varón.

Respecto del nivel educativo por región se observa la mayor tasa de pobreza e indigencia en el nivel educativo 3 (secundaria incompleta) y 7(sin instrucción). Los valores más bajos se encuentran durante el primer trimestre en Patagonia con 36,77% de pobres con secundaria incompleta (4,8% de indigentes) y 36,94% de pobres sin instrucción (4,2% de indigentes). Siendo los porcentajes más altos en el segundo trimestre en Noroeste con secundaria incompleta 58,16% de pobres incluyendo a 18,1% de indigentes y sin instrucción 55,04% de pobres incluyendo 16,1% de indigentes.



En cuanto a la tasa de pobreza e indigencia se observa que el estado ocupacional 2 (desocupado) la tasa más alta se encuentra en el segundo trimestre en la región Noroeste alcanzado al 63,21% de pobres incluyendo a 14,30% indigentes, seguido por 4 (menor de 10 años) para la región Noreste con un 56,47% de pobres incluyendo 18,02% de indigentes.

La tasa de pobreza e indigencia por región según categoría ocupacional tiene los valores más altos en la región de Cuyo durante el primer trimestre con 89,68% de pobres incluyendo a 37,55% de indigentes, para la categoría 4 (trabajador familiar sin remuneración).

Considerando la categoría de inactividad, la pobreza e indigencia por región, muestra que los mayores valores son para la región Noroeste en la categoría 6 (Discapacitado) alcanzando al 74,26% de pobreza incluyendo un 10,55% de indigentes. Siendo el menor porcentaje de pobreza e indigencia para todas las regiones en la categoría jubilado o pensionado con valores entre 8% y 25% para pobreza y menores al 4% para indigencia.

Observando la tasa de pobreza e indigencia por grupos de edad y por regiones, en la región Noreste hay 56,59 personas pobres menores de 14 años incluyendo a un 15,29% de personas indigentes. Mientras que la menor pobreza se observa en el grupo de 65 años o más en la región de Cuyo con 8,3% de personas pobres incluyendo a 1,34% de personas indigentes.

Una vez realizado este análisis resulta conveniente relacionar la base de individuos con la base de hogares con el objetivo de establecer que características del individuo y del hogar resultan importantes para explicar la pobreza e indigencia. Para cumplir con este objetivo, se realiza el análisis de componentes principales que se presenta en el siguiente apartado.

3.2. Componentes principales. Perfiles para categorizar a los individuos

Con base al análisis descriptivo de los datos relacionados con la pobreza e indigencia, se relacionó la base de individuos con su clasificación de Indigente – Pobre - No pobre, con la base de hogares a las que pertenecen los individuos. Con el objetivo de analizar las variables del individuo y del hogar que son significativas como factores causales de la condición de pobreza, se realiza el análisis de componentes principales. Esto permitirá conocer si existen perfiles para categorizar a los individuos, en función de sus características



individuales y del hogar donde viven. Para realizar el estudio se generó una nueva base a partir del proceso de clasificación de los individuos según su situación de pobreza (Indigente – Pobre – No Pobre) y se relacionó con la base de hogares. Este procesamiento se realizó en lenguaje de programación en R. Se creó un nuevo archivo de datos con formato .csv que se procesó y analizó para alcanzar el segundo objetivo de este trabajo.

A partir de ese archivo de datos se realizó el análisis de las componentes principales para reducir la dimensión de las variables. Para realizar este análisis, se utilizó R Commander desde R Studio.

Las variables que se incluyeron para este procedimiento para individuos fueron: categoría de inactividad (CAT_INAC), categoría ocupacional (CAT_OCUP), edad (CH06), sabe leer y escribir (CH09), condición de actividad (ESTADO), nivel educativo (NIVEL_ED), situación de pobreza codificada (indigente – pobre – no pobre) (Situación_cod). características de la vivienda: tiene cuarto de cocina (II4_1), usa algún cuarto para dormir (II5), usa algún cuarto para trabajar (II6), baño (tenencia y uso) (II9), tiene agua (IV6), tiene baño / letrina (IV8), ubicación del baño o letrina (IV9), artefactos del baño (IV10), vive de jubilación o pensión (V2), vive de indemnización por despido (V3), vive de seguro de desempleo (V4), vive de ganancias de algún negocio en el que no trabajan (V9), vive de interés o rentas por plazos fijos (V10), vive beca de estudio (V11)

De las variables seleccionadas se consideraron 7 componentes que determinaron el 60,5% de la varianza total. En el Apéndice 2 se muestra el código en R.

Dado que ya se ha realizado una clasificación de los individuos en relación con su condición de pobreza, en el siguiente apartado se realizará el análisis discriminante. Este procedimiento se utiliza para explicar la pertenencia de distintos individuos a grupos, a partir de los valores de un conjunto de variables que describen a los individuos, a los que se pretende clasificar. Este análisis se aplica con fines explicativos y predictivos. (Manzano y Jiménez, 2017)

3.3. Análisis discriminante. Clasificación y predicción

El análisis discriminante aplicado a fines explicativos trata de determinar la contribución de cada variable clasificadora a la clasificación correcta de los individuos. En la aplicación predictiva, se trata de determinar el grupo al que es más probable que pertenezca un individuo, dadas las variables clasificadoras. (Manzano y Jiménez, 2017)



En este trabajo, la variable clasificadora es la situación en relación con la pobreza (indigente – pobre – no pobre). Se aplicó el análisis discriminante a la base de datos que contiene los 118627 registros correspondientes a la Encuesta Permanente de Hogares del primer y segundo trimestre de 2019. La base de datos se generó relacionando a los individuos con los hogares donde viven. La encuesta revela características sociodemográficas de la población de referencia. Usando solamente como variable clasificadora la situación de pobreza (indigente – pobre (no indigente) – no pobre) y en primer lugar las variables Ingreso total familiar (ITF) y edad (CH06).

Las probabilidades a priori según situación de pobreza son: 0.2333617 para indigente, 0.2511823 para pobre (no indigente), 0.5154560 no pobre. Los promedios para cada uno de los grupos de la variable ingreso total familiar (ITF) son 2775, 98 pesos para indigente, 24854,36 pesos para pobre y 51725,71 pesos para no pobre. La edad promedio de indigente 33,44 años, de pobre es 26,69 años y de no pobre 39,15 años.

Los coeficientes de las funciones discriminantes son:

$$LD1 = 3.293255e - 05 x_1 (ITF) + 1.549156e - 02 x_2 (Edad)$$

$$LD1 = - 7.961180e - 06 x_1 (ITF) + 4.361813e - 02 x_2 (Edad)$$

Dados los autovalores, la primera función discriminante es responsable de la mayor parte de separación entre los grupos (94,47%) y la segunda tiene un papel secundario (5,34%)

En cuanto a la calidad de la clasificación, la matriz de confusión presenta los aciertos y errores de clasificación. Observando dicha matriz, el porcentaje más alto es para la clasificación correcta de no pobre (89,16%) y el menor en el grupo de pobre (47,64%). El resultado muestra que el 74,39% de los casos (88250/118627) han sido correctamente clasificados.

En una segunda función discriminante, se incluyeron las variables situación, ingreso total familiar, ingresos personales, sexo, edad, procedencia del agua de la vivienda, tenencia de baño, régimen de tenencia de la vivienda.

Los coeficientes de la función discriminante son:

$$LD1 = 3.178122e - 05 x_1 (ITF) + 7.659481e - 06 x_2 (P21) + 8.984971e - 02 x_3 (Sexo) + 1.470445e - 02 x_4 (Edad) - 4.534070e - 01 x_5 (IV7) - 2.469766e - 01 x_6 (IV8) + 3.418961e - 02 x_7 (II7)$$



$$LD2 = 1.131272e - 05x_1(\text{ITF}) - 2.503453e - 05 x_2 (\text{P21}) - 6.611285e - 03 x_3(\text{Sexo}) + -3.635381e - 02 x_4 (\text{Edad}) - 5.494028e - 01x_5 (\text{IV7}) x_6 + 1.291566e+00 (\text{IV8}) + 1.279389e - 01x_7 (\text{II7})$$

Dados los autovalores, la primera función discriminante es responsable de la mayor parte de separación entre los grupos (93,87%) y la segunda tiene un papel secundario (6,13%)

La matriz de confusión indica que han sido bien clasificados el 70,57% de los indigentes, el 89,16% de los no pobres y el 47,64% de los pobres. En síntesis, el 74,39% (88250/118627) de los casos han sido correctamente clasificados.

Para determinar que variables se consideran influyentes para la situación de pobreza se realiza un análisis factorial para agrupar las variables en diferentes factores que puedan explicar esta relación.

3.4. Análisis factorial. Factores que influyen en la pobreza.

Una vez realizado el análisis de componentes principales, que determinó las variables que resultaron con mayor importancia, se realizó la técnica de análisis factorial para agrupar las variables en distintos subgrupos o factores. Para realizar este análisis, se utilizó R Commander desde R Studio. En el Apéndice 4 se puede ver el código en R. No se incluyeron en el análisis las variables ingreso total familiar ni ingreso personal, para observar que otras variables están relacionadas con la condición de pobreza.

Las variables incluidas en el análisis factorial correspondientes a características de los individuos son: categoría de inactividad (CAT_INAC), categoría ocupacional (CAT_OCUP), Edad (CH06), sabe leer y escribir (CH09), asiste o asistió a algún establecimiento educativo (CH10), condición de actividad (ESTADO), nivel educativo (NIVEL_ED), situación de pobreza codificada (indigente – pobre – no pobre) (Situacion_cod). Las variables incluidas correspondientes a características del hogar son: tiene cuarto de cocina (II4_1), usa algún cuarto para dormir (II5), usa algún cuarto para trabajar (II6), baño (tenencia y uso) (II9), tiene agua (IV6), tiene baño / letrina (IV8), ubicación del baño o letrina (IV8), artefactos del baño (IV10), tipo de desagüe del baño (IV11), la vivienda está ubicada en zona de basural (IV12_1), la vivienda está ubicada en



zona inundable (IV12_2), la vivienda está ubicada en villa de emergencia (IV12_3), viven de lo que ganan en el trabajo (V1), viven de alguna jubilación o pensión (V2), viven de indemnización por despido (V3), viven de seguro de desempleo (V4), viven de ganancias de algún negocio en el que no trabajan (V9), viven de interés o rentas por plazos fijos (V10), Viven de beca de estudio (V11).

Luego de realizar este análisis se determinó que el número óptimo de factores es 7. Se agruparon las variables para formar cada factor como se presenta a continuación:

Factor 1 (Actividad ocupacional): Categoría de inactividad, categoría ocupacional y condición de actividad.

Factor 2 (Asignaciones): Indemnización por despido, seguro de desempleo, ganancias, intereses y rentas, beca de estudio.

Factor 3 (Ambientes del hogar): Cuarto de cocina, cuarto para dormir, cuarto para trabajar.

Factor 4 (Características de la vivienda): Agua, baño, artefactos del baño.

Factor 5 (Jubilación o pensión): Edad, jubilación o pensión

Factor 6 (Baño) Tiene baño, baño tenencia y uso

Factor 7 (Alfabetización) Sabe leer y escribir, nivel educativo

Los resultados presentan que los factores con mayor importancia son la actividad ocupacional, luego los ingresos por asignación, seguido del ambiente del hogar, las características de la vivienda, la edad junto con los ingresos por jubilación o pensión son determinantes. También resultan relevantes, las características de la vivienda como ambientes del hogar, agua, baño y artefactos del baño, por último, la alfabetización.



Conclusión

Del análisis de los resultados se desprende que para el segundo trimestre de 2019 el 36,44% de las personas está por debajo de la línea de pobreza incluyendo a 8,12% de personas indigentes. La zona Noreste posee el mayor porcentaje de personas pobres (42,75%) incluyendo a indigentes (11,3%). Las mayores tasas de pobreza e indigencia se observan en personas de nivel educativo secundaria incompleta y sin instrucción.

Considerando los individuos y el hogar donde viven, los factores que influyen en la caracterización de la situación de pobreza son actividad ocupacional, las asignaciones (indemnización por despido, seguro de desempleo), subsidio o ayuda social (en dinero) del gobierno o iglesias, ganancias, intereses y rentas, beca de estudio. Una lectura cruzada de las variables permite identificar cada opción de respuesta con la situación de pobreza de los individuos. En cuanto a los ambientes del hogar, las características de la vivienda, si posee agua, baño y sus características, si posee cuarto de cocina, cuartos para dormir, cuartos para trabajar. Sobre el individuo, otros factores importantes son la edad y la percepción de jubilación o pensión como también la alfabetización.

En cuanto a la transferencia hacia el medio, a partir de los diferentes códigos de programación en R generados en este trabajo, actualizando la información de canastas básica alimentaria y canasta básica total, es posible aplicar estos procedimientos a otras bases de datos de la EPH a medida que se van publicando, para calcular la tasa de pobreza e indigencia por región, determinar la situación de pobreza de cada individuo, caracterizar a la muestra y realizar los mismos procedimientos de análisis estadísticos que se han llevado a cabo en este trabajo.

Esta información puede resultar de utilidad para cualquier persona, organización o institución pública o privada que quiera trabajar con estos resultados con el objetivo de caracterizar a la población bajo la línea de pobreza, trabajar sobre las condiciones sociodemográficas de la población, implementar políticas públicas con el objetivo de mejorar las condiciones de la población bajo estudio. Esto permite, actualizar los indicadores a medida que la información de la EPH se encuentre disponible.

El método de análisis discriminante permite realizar una clasificación para nuevos casos y una predicción de clasificación correcta de casos, Se observa que cuando se



considera la variable ingresos totales del hogar la clasificación es correcta, pero cuando no se incluye esta variable y se consideran las características individuales y habitacionales resulta otro tipo de clasificación. Esto es, un individuo categorizado como pobre según sus ingresos puede ser categorizado como no pobre según las condiciones del hogar donde vive y sus características individuales. Por lo tanto, resulta interesante plantear otro tipo de preguntas para futuros estudios. Por ejemplo: qué otro tipo de carencias tienen los individuos además de los ingresos, cuáles son las condiciones de la vivienda, qué tipo de servicios poseen, si hay hacinamiento en el hogar, desnutrición y escolaridad en los niños, de donde provienen la mayoría de los ingresos del hogar, el nivel educativo en relación con la situación ocupacional, profesiones y oficios en relación con el ámbito laboral.

Como propuesta de trabajo a futuro se puede mencionar el interés por generar un modelo predictivo que explique la situación de pobreza de los individuos a través de las variables significativas para este fenómeno. Así como elaborar indicadores para medir la pobreza desde un enfoque multidimensional, considerando otras variables que no fueron estudiadas o que no fueron incluidas en esta encuesta. Con los datos de varios trimestres se puede generar, a partir del análisis de cada trimestre, una matriz de transición para la situación de pobreza, determinando la probabilidad para cada estado: Indigente – Pobre – No pobre.



Referencias bibliográficas

1. Alkire, S., Roche, J. M., Ballon, P., Foster, J., Santos, M. E., & Seth, S. (2015). *Multidimensional poverty measurement and analysis*. Oxford University Press, USA.
2. Conconi, A., & Ham González, A. (2007). *Pobreza multidimensional relativa: Una aplicación a la Argentina*. Documentos de Trabajo del CEDLAS.
3. Gammage, S. (2009). *Género, pobreza de tiempo y capacidades en Guatemala: un análisis multifactorial desde una perspectiva económica*.
4. González, J. I. (2011). *Utilitarismo y Mediciones de Pobreza (Utilitarianism and Measurements of Poverty)*. *Revista de Economía Institucional*, 13(25), 89.
5. Instituto Nacional de Estadística y Censos. INDEC. (2016) *La medición de la pobreza y la indigencia en la Argentina. - 1a ed . - Ciudad Autónoma de Buenos Aires. Metodología INDEC*, 22.
6. Instituto Nacional de Estadística y Censos. INDEC. (2019) *Incidencia de la pobreza y la indigencia en 31 aglomerados urbanos. Segundo semestre de 2018. Informe técnico (3)*, 59.
7. Manzano, J. A., & Jiménez, E. U. (2017). *Análisis multivariante aplicado con R*. Ediciones Paraninfo, SA.
8. Nájera, H. E., & Gordon, D. (2019). *The Importance of Reliability and Construct Validity in Multidimensional Poverty Measurement: An Illustration Using the Multidimensional Poverty Index for Latin America (MPILA)*. *The Journal of Development Studies*, 0(0), 1–21.
9. Poza Lara, C., & Fernández Cornejo, J. A. (2010). *Una aproximación a la construcción de un indicador de pobreza multidimensional: Cuáles son los focos de riesgo en España?*. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 10, 43-72.
10. Ruano, M. M. (2015). *Aplicación de la técnica de componentes principales para el análisis de la pobreza en Honduras*. *Revista Ciencia y Tecnología*, (16), 82-96.



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



11. Steinmann, M., Matei, S. A., & Collmann, J. (2016). A theoretical framework for ethical reflection in big data research. In *Ethical Reasoning in Big Data* (pp. 11-27). Springer, Cham.
12. Torres, R. M. S., Cifuentes, L. A. M., & Murillo, L. D. M. (2019). Indicador de Bourguignon y Chakravarty de pobreza multidimensional en Colombia (No. 017271). Universidad Nacional de Colombia-FCE-CID.



Apéndice 1. Código en R Caracterización de la pobreza e indigencia

```
#Carga de librerías
```

```
library(tidyverse, warn = FALSE)
```

```
library(openxlsx, warn = FALSE)
```

```
# Selección de variables individual
```

```
var.ind <- c('CODUSU','NRO_HOGAR'
```

```
,'COMPONENTE','ANO4','TRIMESTRE','REGION',
```

```
  'AGLOMERADO', 'PONDERA', 'CH04',
```

```
'CH06','CH07','CH08','CH09','CH10','CH15','NIVEL_ED',
```

```
  'ESTADO', 'CAT_OCUP', 'CAT_INAC', 'ITF', 'PONDIIH','P21')
```

```
# Carga de Bases de datos y otros archivos necesarios
```

```
individual.119 <- read.table("usu_individual_t119.txt", sep=";", dec="," , header = TRUE,  
fill = TRUE) %>%
```

```
  select(var.ind)
```

```
individual.219 <- read.table("usu_individual_t219.txt", sep=";", dec="," , header = TRUE,  
fill = TRUE) %>%
```

```
  select(var.ind)
```

```
# Selección de variables hogar
```

```
var.hog<- c('CODUSU', 'NRO_HOGAR',
```

```
'IV1','IV2','IV4','IV5','IV6','IV7','IV8','IV9','IV10',
```

```
  'IV11','IV12_1','IV12_2','IV12_3','II3','II4_1','II4_2','II4_3','II5','II6','II7',
```

```
  'V1','V2','V21','V22','V3','V4','V5','V6','V7','V8','V9','V10','V11','V12','II8','II9')
```

```
hogar.119 <- read.table("usu_hogar_t119.txt", sep=";",dec="," , header= TRUE, fill=  
TRUE) %>%
```

```
  select(var.hog)
```



```
hogar.219 <- read.table("usu_hogar_t219.txt", sep=";",dec=",", header= TRUE, fill=
TRUE) %>%
  select(var.hog)
```

```
#Fusión de la base individual con hogar primer trimestre 2019
BASEHOGAR1t19<-merge(individual.119,hogar.119)
write.csv2(BASEHOGAR1t19,"Basehogar1t19.csv")
```

```
#Fusión de la base individual con hogar segundo trimestre 2019
BASEHOGAR2t19<-merge(individual.219,hogar.219)
write.csv2(BASEHOGAR2t19,"Basehogar2t19.csv")
```

```
#Fusión de las bases completas para primer y segundo trimestre 2019
Base<-rbind(BASEHOGAR1t19,BASEHOGAR2t19)
write.csv2(Base,"Basehogar19.csv")
```

```
#Se cargan las bases de calorías requeridas según sexo y edad, canastas, regiones y
operacionalización de variables
```

```
Adequi <- read.xlsx("ADEQUI.xlsx")
CBA <- read.xlsx("canastas.xlsx",sheet = "CBA")
CBT <- read.xlsx("canastas.xlsx",sheet = "CBT")
dic.regiones <- read.xlsx("Regiones.xlsx")
CH04 <- read.xlsx("Codigovar.xlsx",sheet = "Sexo")
CAT_INAC <- read.xlsx("Codigovar.xlsx",sheet = "Cat_inac")
ESTADO <- read.xlsx("Codigovar.xlsx",sheet = "Estado")
CAT_OCUP <- read.xlsx("Codigovar.xlsx",sheet = "Cat_ocup")
NIVEL_ED <- read.xlsx("Codigovar.xlsx",sheet = "Nivel_ED")
```

```
#Visualización de los archivos cargados
```

```
Adequi
CBA
```



CBT
dic.regiones
CH04
CAT_INAC
CAT_OCUP
ESTADO
NIVEL_ED

Cálculo de Canastas Trimestrales por Región

```
CBA <- CBA %>%  
  mutate(Canasta = 'CBA')  
CBT <- CBT %>%  
  mutate(Canasta = 'CBT')  
  
Canastas_Reg <- bind_rows(CBA,CBT) %>%  
  gather(.,Region, Valor, c(3:(ncol(.)-1) )) %>%  
  mutate(Trimestre = case_when(Mes %in% c(1:3) ~1,  
    Mes %in% c(4:6) ~2,  
    Mes %in% c(7:9) ~3,  
    Mes %in% c(10:12) ~4),  
    Periodo = paste(Año, Trimestre, sep='.'))  
Canastas_Reg  
  
Canastas_Reg_2 <- Canastas_Reg %>%  
  group_by(Canasta, Region, Periodo) %>%  
  summarise(Valor = mean(Valor)) %>%  
  spread(., Canasta,Valor) %>%  
  left_join(., dic.regiones, by = "Region") %>%  
  ungroup()  
Canastas_Reg_2
```



Aplicación de sucesivos pasos para el cálculo de Pobreza e Indigencia por Periodo

```
Pobreza_Individual <- Base %>%  
  mutate(Periodo = paste(ANO4, TRIMESTRE, sep='.')) %>%  
  left_join(., Adequi, by = c("CH04", "CH06")) %>%  
  left_join(., Canastas_Reg_2, by = c("REGION", "Periodo"))  
Pobreza_Individual
```

```
Pobreza_Individual_paso2 <- Pobreza_Individual %>%  
  group_by(CODUSU, NRO_HOGAR, Periodo) %>%  
  mutate(Adequi_hogar = sum(adequi)) %>%  
  ungroup()  
Pobreza_Individual_paso2
```

#Se clasifica según situación de pobreza a los individuos

```
Pobreza_Individual_paso3 <- Pobreza_Individual_paso2 %>%  
  mutate(CBA = CBA*Adequi_hogar,  
         CBT = CBT*Adequi_hogar,  
         Situacion = case_when(ITF<CBA ~ 'Indigente',  
                               ITF>=CBA & ITF<CBT ~ 'Pobre',  
                               ITF>=CBT ~ 'No Pobre'))  
Pobreza_Individual_paso3
```

Se genera una nueva variable con los resultados

```
Pobreza3<-cbind(Pobreza_Individual_paso3)
```

Se genera una base de datos con los resultados en un archivo .csv

```
write.csv2(Pobreza_Individual_paso3,"Pobreza3.csv")  
write.csv2(Basehogar,"Hogar.csv")
```

Cálculo de Tasas de Pobreza e Indigencia

```
Pobreza_resumen <- Pobreza_Individual_paso3 %>%
```



```
group_by(Periodo) %>%
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
            sum(PONDIH,na.rm = TRUE),

            Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
            sum(PONDIH,na.rm = TRUE))

Pobreza_resumen
Pobreza<-cbind(Pobreza_resumen)
write.csv2(Pobreza_resumen,"Pobreza.csv")

# Cálculo de Tasas de Pobreza e Indigencia
Pobreza1s19_resumen <- Pobreza_Individual_paso3 %>%
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
            sum(PONDIH,na.rm = TRUE),

            Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
            sum(PONDIH,na.rm = TRUE))

Pobreza1s19_resumen
Pobreza1s19<-cbind(Pobreza1s19_resumen)
write.csv2(Pobreza1s19_resumen,"Pobreza1s19.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones
Pobreza_resumen_region <- Pobreza_Individual_paso3 %>%
  group_by(Periodo,Region) %>%
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
            sum(PONDIH,na.rm = TRUE),

            Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
```



```
sum(PONDIH,na.rm = TRUE))

Pobreza_resumen_region
Pobrezaregion<-cbind(Pobreza_resumen_region)
write.csv2(Pobreza_resumen_region,"Pobreza_region.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por sexo
Pobreza_resumen_region_sexo <- Pobreza_Individual_paso3 %>%
  group_by(Periodo,Region,CH04) %>%
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
  sum(PONDIH,na.rm = TRUE),
  Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
  sum(PONDIH,na.rm = TRUE))
Pobreza_resumen_region_sexo
Pobrezasexo<-cbind(Pobreza_resumen_region_sexo)
write.csv2(Pobreza_resumen_region_sexo,"Pobreza_region_sexo.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por sexo primer semestre
Pobreza_resumen_region_sexo2 <- Pobreza_Individual_paso3 %>%
  group_by(Region,CH04) %>%
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
  sum(PONDIH,na.rm = TRUE),
  Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
  sum(PONDIH,na.rm = TRUE))
Pobreza_resumen_region_sexo2
Pobrezasexo2<-cbind(Pobreza_resumen_region_sexo2)
write.csv2(Pobreza_resumen_region_sexo2,"Pobreza_region_sexo2.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por Nivel Educativo
Pobreza_resumen_region_nivel_ed <- Pobreza_Individual_paso3 %>%
```



```
group_by(Periodo,Region,NIVEL_ED) %>%
summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
sum(PONDIH,na.rm = TRUE),

Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
sum(PONDIH,na.rm = TRUE))
Pobreza_resumen_region_nivel_ed
Pobreza_niveled<-cbind(Pobreza_resumen_region_nivel_ed)
write.csv2(Pobreza_resumen_region_nivel_ed,"Pobreza_region_nivel_ed.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por Estado
Pobreza_resumen_region_ESTADO <- Pobreza_Individual_paso3 %>%
group_by(Periodo,Region,ESTADO) %>%
summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
sum(PONDIH,na.rm = TRUE),
Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
sum(PONDIH,na.rm = TRUE))
Pobreza_resumen_region_ESTADO
Pobreza_ESTADO<-cbind(Pobreza_resumen_region_ESTADO)
write.csv2(Pobreza_resumen_region_ESTADO,"Pobreza_region_ESTADO.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por CAT_OCUP
Pobreza_resumen_region_CAT_OCUP <- Pobreza_Individual_paso3 %>%
group_by(Periodo,Region,CAT_OCUP) %>%
summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
sum(PONDIH,na.rm = TRUE),
Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
sum(PONDIH,na.rm = TRUE))
```



```
Pobreza_resumen_region_CAT_OCUP
PobrezaCATOCUP<-cbind(Pobreza_resumen_region_CAT_OCUP)
write.csv2(Pobreza_resumen_region_CAT_OCUP,"Pobreza_region_CAT_OCUP.csv")

# Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por CAT_INAC
Pobreza_resumen_region_CAT_INAC <- Pobreza_Individual_paso3 %>%
  group_by(Periodo,Region,CAT_INAC) %>%
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
'Indigente')],na.rm = TRUE)/
  sum(PONDIH,na.rm = TRUE),

  Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
  sum(PONDIH,na.rm = TRUE))
Pobreza_resumen_region_CAT_INAC
PobrezaCATINAC<-cbind(Pobreza_resumen_region_CAT_INAC)
write.csv2(Pobreza_resumen_region_CAT_INAC,"Pobreza_region_CAT_INAC.csv")

#Se clasifican los individuos según sexo
Pobreza_Individual_paso4 <- Pobreza_Individual_paso3 %>%
mutate(CH04=case_when(CH04==1~'Varon',
  CH04==2~'Mujer'))%>%
group_by(CH04)
Pobreza_Individual_paso4

#Se clasifican los individuos según grupos de edad
Pobreza_Individual_paso5 <- Pobreza_Individual_paso3 %>%
mutate(EDAD=case_when(CH06 %in% c(-1:14)~'0 a 14',
  CH06 %in% c(15:29)~'15 a 29',
  CH06 %in% c(30:64)~'30 a 64',
  CH06 %in% c(65:107)~'65 y mas')) %>%
group_by(EDAD)
```



Pobreza_Individual_paso5

Cálculo de Tasas de Pobreza e Indigencia abierto por Regiones por grupos de edad

```
Pobreza_resumen_region_EDAD <- Pobreza_Individual_paso5 %>%
```

```
  group_by(Periodo,Region,EDAD) %>%
```

```
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
```

```
'Indigente']),na.rm = TRUE)/
```

```
    sum(PONDIH,na.rm = TRUE),
```

```
    Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
```

```
    sum(PONDIH,na.rm = TRUE))
```

```
Pobreza_resumen_region_EDAD
```

```
PobrezaRE<-cbind(Pobreza_resumen_region_EDAD)
```

```
write.csv2(Pobreza_resumen_region_EDAD,"Pobreza_region_edad.csv")
```

Cálculo de Tasas de Pobreza e Indigencia abierto por aglomerados

```
Pobreza_resumen_AGLOMERADOS<- Pobreza_Individual_paso5 %>%
```

```
  group_by(Periodo,AGLOMERADO) %>%
```

```
  summarise(Tasa_pobreza = sum(PONDIH[Situacion %in% c('Pobre',
```

```
'Indigente']),na.rm = TRUE)/
```

```
    sum(PONDIH,na.rm = TRUE),
```

```
    Tasa_indigencia = sum(PONDIH[Situacion == 'Indigente'],na.rm = TRUE)/
```

```
    sum(PONDIH,na.rm = TRUE))
```

```
Pobreza_resumen_AGLOMERADOS
```

```
PobrezaAGLO<-cbind(Pobreza_resumen_AGLOMERADOS)
```

```
write.csv2(Pobreza_resumen_AGLOMERADOS,"Pobreza_AGLOMERADOS.csv")
```

#Se clasifica según situación de pobreza a los individuos

```
Pobreza_Individual_paso3cod <- Pobreza_Individual_paso2 %>%
```

```
  mutate(CBA = CBA*Adequi_hogar,
```



$CBT = CBT * Adequi_hogar,$

$Situacion2 = case_when(ITF < CBA \quad \sim 1,$

$ITF >= CBA \ \& \ ITF < CBT \sim 2,$

$ITF >= CBT \quad \sim 3))$

Pobreza_Individual_paso3cod



Apéndice 2. Código en R Componentes Principales

```
local({  
  .PC <- princomp(~CAT_INAC + CAT_OCUP + CH06 + CH09 + ESTADO + II4_1 +  
II5 + II6 + II9 + IV6 + IV8 + IV9 + IV10 +  
  NIVEL_ED + Situacion_cod + V2 + V3 + V4 + V9 + V10 + V11, cor =  
TRUE, data = Dataset)  
  cat("\nComponent loadings:\n")  
  print(unclass(loadings(.PC)))  
  cat("\nComponent variances:\n")  
  print(.PC$sd^2)  
  cat("\n")  
  print(summary(.PC))  
  screeplot(.PC)  
  Dataset <<- within(Dataset, {  
    PC7 <- .PC$scores[, 7]  
    PC6 <- .PC$scores[, 6]  
    PC5 <- .PC$scores[, 5]  
    PC4 <- .PC$scores[, 4]  
    PC3 <- .PC$scores[, 3]  
    PC2 <- .PC$scores[, 2]  
    PC1 <- .PC$scores[, 1]  
  })  
})
```



Apéndice 3. Código en R Análisis Discriminante

```
#Cargamos librerías necesarias para trabajar
library(ggplot2)
library(MASS)
library(readxl)
library(descr)
library(biotools)
library(mvnormtest)

#Cargamos la base de datos desde un archivo Excel
Datos <- read.xlsx("Pobreza3h.xlsx")
#Visualizamos los datos
View(Datos)
#Aplicamos la función lda para la estimación
fit<-lda(data=Datos, Situacion~ITF+CH06)
#Coeficientes de la función discriminante de Fisher
fit
#Obtenemos la pertenencia predicha y la anexamos al fichero de datos
fit.p<-predict(fit)$class
Datos<-data.frame(Datos,fit.p)
Datos
#Tabla cruzada de aciertos y errores de clasificación
#Matriz de confusión
CrossTable(Datos[,65],Datos$fit.p,digits=2,format="SPSS",
prop.c=FALSE, prop.chisq=FALSE, prop.t=FALSE,
dnn=c("Grupo real","Grupo pronosticado"))

#Aplicamos la función lda para la estimación tomando otras variables
fit2<-lda(data=Datos, Situacion~ITF+P21+CH04+CH06+IV7+IV8+II7)
#Coeficientes de la función discriminante de Fisher
fit2
```



#Obtenemos la pertenencia predicha y la anexamos al fichero de datos

```
fit2.p<-predict(fit2)$class
```

```
Datos<-data.frame(Datos,fit2.p)
```

Datos

#Tabla cruzada de aciertos y errores de clasificación

#Matriz de confusión

```
CrossTable(Datos[,65],Datos$fit2.p,digits=2,format="SPSS",  
prop.c=FALSE, prop.chisq=FALSE, prop.t=FALSE,  
dnn=c("Grupo real","Grupo pronosticado"))
```



Apéndice 4. Código en R Análisis Factorial

```
local({  
  .FA <- factanal(~Situacion_cod + CAT_INAC + CAT_OCUP + CH06 + CH09 + CH10 +  
  ESTADO + II4_1 + II5 + II6 + II9 + IV1 + IV2 + IV4 + IV5 + IV6 + IV8+ IV9 + IV10 +  
  IV11 + IV12_1 + IV12_2 + IV12_3 + NIVEL_ED + V1 + V2 + V3 + V4 + V9 + V10 +  
  V11, factors = 7, rotation = "varimax", scores = "regression", data = Dataset)  
  print(.FA)  
  Dataset <<- within(Dataset, {  
    F7 <- .FA$scores[, 7]  
    F6 <- .FA$scores[, 6]  
    F5 <- .FA$scores[, 5]  
    F4 <- .FA$scores[, 4]  
    F3 <- .FA$scores[, 3]  
    F2 <- .FA$scores[, 2]  
    F1 <- .FA$scores[, 1]  
  })  
})
```