UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Aplicación de Inferencia estadística para la identificación del comportamiento de la pobreza y las carencias sociales en México.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A :

BRENDA ZURAZY RODRÍGUEZ PÉREZ

DIRECTOR DE TESIS:
DRA. SILVIA RUÍZ VELASCO ACOSTA
CIUDAD DE MÉXICO (2019)

Índice general

Introducción

La pobreza en un principio fue definida únicamente como la falta de ingresos de una persona para tener la capacidad de cubrir la canasta básica, sin embargo, también se tiene el concepto de pobreza multidimensional la cual, además de considerar la falta de ingresos de los individuos considera la falta de acceso a los derechos sociales como lo son: el rezago educativo, servicios de salud, seguridad social, la calidad de espacios de la vivienda, los servicios básicos de la vivienda y la alimentación, que a su vez son factores que intensifican la condición de desigualdad e incrementan la vulnerabilidad de la población.

Actualmente en nuestro país, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), es el organismo responsable de emitir los lineamientos y metodologías para la medición de la pobreza multidimensional.

A grandes rasgos, el resultado de la clasificación de la pobreza, según la metodología del CONEVAL es:

- Pobres. Población con ingreso inferior al valor de la línea de bienestar y que padece al menos una carencia social.
- Pobres extremos. Población con ingreso inferior al valor de la línea de bienestar y que padece tres o más carencias sociales.
- Vulnerables por carencias sociales. Población que presenta una o más carencias sociales, pero cuyo ingreso es superior a la línea de bienestar.
- Vulnerables por ingresos. Población que no presenta carencias sociales y cuyo ingreso es inferior o igual a la línea de bienestar.
- No pobre multidimensional y no vulnerable. Población cuyo ingreso es superior a la línea de bienestar y no tiene carencia social alguna.

El objetivo de esta tesis principalmente consiste en analizar los resultados de la Pobreza del Consejo Nacional de Evaluación de la Política de Desarrollo Social,

CONEVAL, para ello el desarrollo del documento se concentra en los siguientes capítulos:

1. Se expone el marco jurídico para la definición, identificación y medición de la pobreza, así como una introducción del concepto de pobreza y una breve explicación de la composición de las carencias sociales.

2. Se presenta el concepto del Análisis de Componentes Principales, mediante la descripción del proceso analítico utilizado para obtener las componentes a partir de la matriz de covarianzas donde los renglones representan individuos y las columnas representan los valores de las variables de estudio.

3. En éste capítulo se encuentra la aplicación del Análisis de Componentes Principales con la matriz de covarianzas y correlación con las carencias sociales como variables iniciales, esto permitió la simplificación de la información la cual se aprovechó para generar una representación de la población en dos y tres dimensiones, identificando a su vez mediante colores a la población en condiciones de pobreza extrema, pobreza, vulnerabilidad y a la población no pobre. Finalmente a través de éste análisis se identificó la estructura de los datos detectando las tendencias para aquellas componentes de mayor variabilidad.

4. Se muestra una introducción teórica de los modelos de regresión logística tanto simple como múltiple, así como la formulación del modelo e interpretación de los parámetros estimados.

5. Aplicación del análisis de regresión logística para identificar la relación y dependencia entre las variables de las carencias sociales y por lo tanto la situación de pobreza, en éste capítulo se hicieron dos ejercicios, uno tomando cada una de las variables sociales como variables independientes de manera individual y el segundo ejercicio de manera conjunta.

6. Se exponen las conclusiones del resultado del análisis de componentes principales y la predicción obtenida en el análisis de regresión logística.

Al final de éste trabajo se encuentran los apéndices con la descripción de la base de datos, la construcción analítica de las variables de carencias sociales así como los códigos en R project para la aplicación del análisis de componentes principales y la regresión logística.

Capítulo 1

Concepto de Pobreza

1.1. Marco jurídico

La Ley General de Desarrollo Social (LGDS) promulgada el 20 de enero de 2004 establece en su capítulo II artículo 14, que la Política Nacional de Desarrollo Social debe incluir al menos las siguientes vertientes:

1. Superación de la Pobreza a través de la educación, salud, alimentación nutritiva y de calidad, generación de empleo e ingreso, auto empleo y capacitación;
2. Seguridad Social y programas asistenciales;
3. Desarrollo Regional;
4. Infraestructura social básica y
5. Fomento al sector social de la economía.

Es decir que la LGDS establece el aseguramiento de los derechos sociales tanto individuales como colectivos, así como el desarrollo económico con sentido social que eleve el ingreso de la población y contribuya a reducir la desigualdad y por tanto la pobreza.

El concepto de pobreza por mucho tiempo fue definido únicamente como la falta de ingresos de una persona para tener la capacidad de cubrir la canasta básica, sin embargo, hoy en día se tiene el concepto de pobreza multidimensional, la cual, además de considerar la falta de ingresos de los individuos, también considera la falta de acceso a los derechos sociales, esto derivado de que son factores que intensifican la condición de desigualdad e incrementan la vulnerabilidad de la población.

En nuestro país, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL)[1], un organismo con autonomía técnica y de gestión, es el responsable de emitir los lineamientos y criterios para la definición, identificación y medición de la pobreza multidimensional, para la cual por ley, debe considerar los Derechos Sociales y Bienestar económico con una periodicidad de dos años a nivel estatal y cinco a nivel municipal con información generada por el Instituto Nacional de Estadística y Geografía (INEGI).

Con el propósito de brindar una respuesta metodológicamente alineada con los mandatos de la LGDS, el CONEVAL desarrolló dos líneas de investigación entre el 2006 y 2009. La primera, se enfocó en definir el marco teórico-metodológico de la medición multidimensional de la pobreza, en tanto que la segunda se orientó a la generación de la información necesaria para realizarla.[2]

La primera línea de investigación consistió en la realización de varios estudios y seminarios con especialistas nacionales e internacionales. En la primera etapa, se consultó a un grupo de expertos en medición de pobreza, a fin de identificar los principales retos para definir y medir la pobreza multidimensional. A partir de los resultados de esas primeras sesiones, en 2007 se decidió emprender la segunda etapa, en la cual el CONEVAL solicitó a un conjunto de expertos la elaboración de cinco propuestas metodológicas que permitieran resolver el problema de la medición multidimensional de la pobreza, de acuerdo con los requerimientos de la LGDS. Las propuestas elaboradas fueron presentadas en dos talleres de trabajo y un seminario académico internacional, durante los cuales se discutieron y analizaron sus principales características, propiedades y alcances.

Como resultado, el CONEVAL emprendió durante una tercera etapa, la elaboración de una metodología para la medición de la pobreza que cumpliera con las disposiciones legales, que fuera sensible a la problemática social mexicana y que estuviera fundamentada en sólidos criterios metodológicos. Finalmente, bajo estas premisas, se elaboró la Metodología para la medición multidimensional de la pobreza en México.

Así, la metodología para la medición multidimensional de la pobreza en México fue elaborada por el CONEVAL con base en la LGDS, en las propuestas presentadas por los especialistas y en el conocimiento científico y técnico acumulado sobre la medición de pobreza, manteniendo separados los ámbitos de bienestar económico y los derechos, por ser de naturaleza distinta.

[1] Organismo Estatal Mexicano que la Ley General de Desarrollo Social (LGDS) estableció como la institución encargada de evaluar la política social del Estado Mexicano.

[2] Metodología para la medición multidimensional de la pobreza en México.

Esto llevó a definir una medida bidimensional: una de las dimensiones hace referencia a las carencias en derechos sociales y la otra al bienestar económico.

1.2. Derechos sociales

Los indicadores de los derechos sociales que la metodología de la pobreza multidimensional considera como fundamentales para que las personas estén en posibilidad de desenvolverse de manera adecuada en su entorno social, son:

- Acceso a la educación
- Acceso a los servicios de salud
- Acceso a la seguridad social
- Calidad y espacios de la vivienda
- Acceso a los servicios básicos en la vivienda
- Acceso a la alimentación

Cabe destacar que el estudio y medición de dichos indicadores, se han convertido en una herramienta fundamental para la orientación de políticas públicas dirigidas a superar e impulsar el bienestar social de las personas.

A continuación, se realizará una breve descripción de cada uno estos indicadores[3].

1.2.1. Acceso a la educación

El CONEVAL, utiliza los criterios de la Norma de Escolaridad Obligatoria del Estado Mexicano (NEOEM)[4] para la construcción del índice de carencia por rezago educativo, en los cuales se establece que una persona tiene dicha carencia si cumple con alguno de los siguientes puntos:

- Tener de tres a quince años, y no contar con la educación básica obligatoria o no asistir a la escuela

[3] En el apéndice B. Construcción de los indicadores se describen de manera analítica los indicadores de los derechos sociales

[4] En el artículo 3° de la Constitución Política de los Estados Unidos Mexicanos (CPEUM) y los artículos 2°, 3° y 4° de la Ley General de Educación establecen que toda la población debe cursar la educación básica obligatoria, incorporando como derecho fundamental la enseñanza obligatoria a nivel primaria y en 1993 se amplió la enseñanza obligatoria para incluir la educación secundaria.

- Si nació antes de 1982 y no cuenta con la primaria completa
- Si nació a partir de 1982 y no cuenta con la secundaria completa

1.2.2. Acceso a los servicios de salud

El acceso a los servicios de salud es un derecho constitucional[5], el cual refiere que todo mexicano tiene derecho a ser incorporado al Sistema de Protección Social en Salud. De esta forma, las familias y personas que no sean derecho habientes de las instituciones de seguridad social, o no cuenten con algún otro mecanismo de previsión social en salud, deben ser inscritas en dicho sistema.

A partir de estos criterios, se considera que una persona se encuentra en situación de carencia por acceso a los servicios de salud cuando:

- No cuenta con adscripción o derecho a recibir servicios médicos de alguna institución que los presta, incluyendo el Seguro Popular, las instituciones públicas de seguridad social (IMSS, ISSSTE federal o estatal, Pemex, Ejército o Marina) o los servicios médicos privados.

1.2.3. Acceso a la seguridad social

Se refiere a los medios que garantizan a las personas tanto económicamente activas y no activas, así como sus familiares, para enfrentar situaciones tales como enfermedades, accidentes, jubilación, desempleo, etc. Por lo tanto, las personas se encuentran en situación de carencia por acceso a la seguridad social cuando:

- No tienen acceso a una pensión o jubilación como prestación de su empleo o por medio de algún programa social.
- No cuenta con una pensión o jubilación.
- Personas asalariadas que no cuentan con las prestaciones establecidas por la Ley.

1.2.4. Calidad y espacios de la vivienda

En conjunto con la Comisión Nacional de Vivienda (CONAVI), el CONEVAL estableció los criterios fundamentales para el cálculo del indicador de esta carencia: el material de construcción y los espacios.

[5] El artículo 4° de la Constitución establece que toda la población mexicana tiene derecho a la protección de la salud

De acuerdo con estos criterios, se considera que la población que se encuentra en situación de carencia por calidad y espacios de la vivienda, son las que habitan en viviendas con las siguientes características:

- Pisos cuyo material es únicamente tierra.
- Techos de lámina de cartón o desechos.
- Muros de carrizo, bambú o palma; de lámina de cartón, metálica o asbesto; o material de desecho.
- Cuando la razón de personas por cuarto (hacinamiento) es mayor que 2.5.

1.2.5. Acceso a servicios básicos de la vivienda

Al igual que en el indicador de Calidad y Espacios de la Vivienda, el CONEVAL en colaboración con el CONAVI, definieron las características de la vivienda con disposición de servicios básicos indispensables a partir de los cuales se puede identificar a las viviendas que tienen condiciones de habitabilidad inadecuadas.

De acuerdo con lo anterior, se considera como población en situación de carencia por servicios básicos en la vivienda a las personas que residen en viviendas que presenten, al menos, una de las siguientes características:

- No tiene acceso al agua potable en la vivienda, es decir, el agua se obtiene de un pozo, río, lago, arroyo, pipa; o bien, el agua entubada la adquieren por acarreo de otra vivienda, o de la llave pública o hidrante.
- No cuentan con servicio de drenaje o el desagüe tiene conexión a una tubería que va a dar a un río, lago, mar, barranca o grieta.
- No disponen de energía eléctrica.
- El combustible que se usa para cocinar o calentar los alimentos es leña o carbón sin chimenea.

1.2.6. Acceso a la alimentación

El estado mexicano ha ratificado diversos acuerdos internacionales en materia de derecho a la alimentación, entre los cuales se encuentran: el Pacto Internacional de Derechos Económicos, Sociales y Culturales y la Declaración de Roma de 1996 sobre la Seguridad Alimentaria Mundial los cuales establecen el derecho a no padecer hambre y el derecho a gozar de acceso a una alimentación sana y nutritiva. Por ello es relevante que el acceso a la alimentación sea considerado como uno de los derechos sociales para el desarrollo social en el país.

Para el cálculo de este indicador, el CONEVAL tomó como referencia la definición de seguridad alimentaria propuesta por la FAO (2006), que dice: *"la seguridad alimentaria, comprende el acceso en todo momento a comida suficiente para llevar una vida activa y sana, lo cual está asociado a los conceptos de estabilidad, suficiencia y variedad de los alimentos".*

Para evaluar el derecho a la alimentación, se establece que la seguridad alimentaria esté en función de la falta de dinero o recursos que tiene como consecuencia el difícil acceso a una alimentación variada y suficiente, la cual comprende la siguiente escala:

- Inseguridad alimentaria severa.
- Inseguridad alimentaria moderada.
- Inseguridad alimentaria leve y
- Seguridad alimentaria.

Se considera en situación de carencia por acceso a la alimentación a los hogares que presenten un grado de inseguridad alimentaria moderado o severo.

1.2.7. Bienestar económico

El bienestar económico, se define a partir de una "Línea de Bienestar" la cual marca el ingreso corriente per cápita necesario para que un individuo pueda adquirir los bienes y servicios básicos para satisfacer sus necesidades, para ello, se establecieron líneas de bienestar partir del gasto y consumo de la población.

El ingreso corriente total, se compone de la suma de percepciones de todos los miembros del hogar, monetarios y no monetarios, e incluye las remuneraciones al trabajo, el ingreso por la explotación de negocios propios, la renta del capital, las transferencias, los ingresos por cooperativas, el valor imputado por autoconsumo, el pago en especie, los regalos recibidos en especie y la estimación de la renta por el uso de la vivienda propia.

La estimación de los recursos tiene el propósito de reflejar de manera adecuada los recursos de los que disponen los hogares para la satisfacción de sus necesidades. Se definieron dos canastas básicas, una alimentaria y otra no alimentaria, las cuales reflejan los patrones de consumo de la población y que permiten efectuar estimaciones para el conjunto de las localidades rurales y urbanas.

A partir de estas canastas básicas se determina la línea del bienestar, la cual es la suma de los costos de la canasta alimentaria y no alimentaria, y la línea de bienestar mínimo equivalente al costo únicamente de la canasta alimentaria.

1.3. Clasificación de Pobreza

Con base en lo anteriormente establecido, se define que una persona se encuentra en situación de pobreza multidimensional cuando no tiene garantizado el ejercicio de al menos uno de sus derechos para el desarrollo social, y además si sus ingresos son insuficientes para adquirir los bienes y servicios que requiere para satisfacer sus necesidades.

La clasificación del nivel de pobreza o vulnerabilidad es la siguiente:

- **Pobres**. Población con ingreso inferior al valor de la Línea de Bienestar y que padece al menos una carencia social.
- **Pobres extremos** Población con ingreso inferior al valor de la Línea de Bienestar y que padece tres o más carencias sociales.
- **Vulnerables por carencias sociales**. Población que presenta una o más carencias sociales, pero cuyo ingreso es superior a la Línea de Bienestar.
- **Vulnerables por ingresos**. Población que no presenta carencias sociales y cuyo ingreso es inferior o igual a la Línea de Bienestar.
- **No pobre multidimensional y no vulnerable**. Población cuyo ingreso es superior a la Línea de Bienestar y no tiene carencia social alguna.

Capítulo 2

Análisis Exploratorio

2.1. Análisis de Componentes Principales

Esta técnica fue inicialmente desarrollada por Pearson (1901) a finales del siglo XIX, el cual buscaba ajustar un conjunto de datos a líneas y planos por medio de mínimos cuadrados con el fin de estudiar la relación entre las variables. Posteriormente Hotelling (1933) en el siglo XX mediante combinaciones lineales de variables, buscaba un conjunto más pequeño donde las variables fueran independientes.

De esta manera, el objetivo principal del análisis de componentes principales es la reducción de un conjunto original de variables correlacionadas en un conjunto más pequeño de variables no correlacionados que representen la mayor parte de la información encontrada en las variables originales (cuanto mayor sea la variabilidad de los datos, es decir la varianza, se considera que existe mayor información) y que ayudan a entender mediante la construcción de índices la estructura inherente a los datos.

La transformación del nuevo conjunto de variables son llamadas Componentes Principales, las cuales son ordenadas de forma ascendente de acuerdo con la cantidad de varianza que contengan.

2.1.1. Cálculo de Componentes Principales Poblacionales

Definición: La primera componente principal es una función que es una combinación lineal $\alpha_{\mathbf{1}}{}'\mathbf{X}$ donde $\alpha_{\mathbf{1}} = (\alpha_{11}, \alpha_{12}, ...\alpha_{1p})$ es un vector de p constantes y $\mathbf{X} = (x_1, x_2, ...x_p)$ un vector aleatorio p dimensional, además se denota $'$ al

vector transpuesto, de tal manera que:

$$Z_1 = \alpha'_{\mathbf{1}}\mathbf{X} = \alpha_{11}x_1 + \alpha_{21}x_2 + ... + \alpha_{p1}x_p = \sum_{j=1}^{p} \alpha_{j1}x_j \tag{2.1.1}$$

y

$$Var\,(Z_1) = Var\,(\alpha'_1\mathbf{X}) = máx_{\alpha_1}\ \{Var\,(\alpha'_{\mathbf{1}}\mathbf{X}) :\ \alpha'_1\alpha_1 = 1\} \tag{2.1.2}$$

de tal manera que la primera componente principal es una combinación lineal normalizada de $\mathbf{X}$, que además es la que tiene mayor varianza.

Por propiedades de la varianza y covarianza, podemos reescribir la varianza como:

$$Var\,[\alpha'_{\mathbf{1}}\mathbf{X}] = Cov\,[\alpha'_{\mathbf{1}}\mathbf{X},\, \alpha'_{\mathbf{1}}\mathbf{X}] = \alpha'_1 Cov\,[\mathbf{X},\, \mathbf{X}]\,\alpha_1 = \alpha'_1\Sigma\alpha_1 \tag{2.1.3}$$

Con base en lo anterior, el problema para obtener la primer componente principal consiste en encontrar el vector de constantes α_1 que maximicen la varianza de la función $\alpha'_1 X$, es decir:

$$máx_{\alpha_1}\ \{Var(\alpha'_{\mathbf{1}}\mathbf{X}) :\ \alpha'_1\alpha_1 = 1\} \tag{2.1.4}$$

Usando el método de multiplicadores de Lagrange, y reescribiendo la varianza en términos de la covarianza, se tiene que:

$$L = \alpha'_1\Sigma\alpha_1 - \lambda_1(\alpha'_1\alpha_1 - 1) \tag{2.1.5}$$

donde λ es el multiplicador de Lagrange, ahora se deriva la ec. 2.1.5 con respecto con $\alpha_{\mathbf{1}}$ e igualamos a 0 ya que se busca el máximo:

$$\frac{dL}{d\alpha_1} = \alpha'_1\Sigma d\alpha_1 - \lambda_1(\alpha'_1 d\alpha_1 - 1) = 0 \tag{2.1.6}$$

$$\Leftrightarrow\ \alpha'_1\Sigma - \lambda_1\alpha'_1 = 0$$

además, como sabemos que $\alpha'_1\alpha_1 = 1$

$$\Sigma\alpha_1 - \lambda_1\alpha_1 = 0$$

$$\Leftrightarrow (\Sigma - \lambda_1 I_p)\alpha_1 = 0$$

donde I_p es una matriz identidad de dimensión $p \times p$, ahora bien, la expresión anterior es un sistema homogéneo que por propiedades de álgebra lineal se sabe que tiene solución si el determinante de la matriz es nulo, es decir:

$$Det\,|\Sigma - \lambda_1 I_p| = 0$$

Lo cual sucede si y sólo si el escalar λ_1 es valor propio de la matriz de covarianzas Σ de tamaño $p \times p$, cabe destacar que el $Det\,|\Sigma - \lambda_1 I_p| = 0$ es el polinomio característico de Σ lo cual indica que tiene p valores propios diferentes que son solución y dado que Σ es definida positiva, entonces todos sus valores propios son positivos, esta expresión se maximiza cuando α_1 es su vector propio.

Por lo tanto,

$$Var\,(Z_1) = Var\,(\alpha_{\mathbf{1}}'\mathbf{X}) = \lambda_1 \tag{2.1.7}$$

Posteriormente, se construye la función lineal $\alpha'_{\mathbf{2}}\mathbf{X}$ de tal manera que no está correlacionada con $\alpha'_{\mathbf{1}}\mathbf{X}$.

Definición: Se define la segunda componente principal como la variable aleatoria Z_2 tal que:

$$Z_2 = \alpha_{\mathbf{2}}'\mathbf{X} = \alpha_{12}x_1 + \alpha_{22}x_2 + ... + \alpha_{p2}x_p = \sum \alpha_{j2}x_j \tag{2.1.8}$$

con $\alpha_2' = (\alpha_{12}, \alpha_{22}, ...\alpha_{p2})' \in \mathbb{R}^{\mathbb{P}}$ y

$$Var\,(Z_2) = Var\,(\alpha'_{\mathbf{2}}\mathbf{X}) = máx_{\alpha_2}\ \{Var\,(\alpha_{\mathbf{2}}'\mathbf{X}) :\ \alpha_1'\alpha_1 = 1,\ \alpha_1'\alpha_2 = 0\} \tag{2.1.9}$$

La segunda componente principal, es otra combinación lineal de las variables de X y de entre todas las combinaciones lineales formadas por vectores unitarios ortogonales con α_1 es la que tiene mayor varianza.

De manera similar que con la primera componente, el problema consiste en buscar:

$$máx_{\alpha_2}\ \{Var(\alpha_{\mathbf{2}}'\mathbf{X}) :\ \alpha_2'\alpha_2 = 1,\ \alpha_1'\alpha_2 = 0\} \tag{2.1.10}$$

Resolviendo el problema con multiplicadores de Lagrange, y reescribiendo la varianza en términos de la covarianza, se tiene que:

$$L = \alpha_2'\Sigma\alpha_2 - \lambda_2(\alpha_2'\alpha_2 - 1) - \mu_2\alpha_2'\alpha_1 \tag{2.1.11}$$

Derivando respecto a α_2 e igualando a cero, se tiene que:

$$\frac{dL}{d\alpha_2} = \Sigma\alpha_2 - \lambda_2\alpha_2 - \mu_2\alpha_1 = 0 \tag{2.1.12}$$

multiplicando la ec. 2.1.12 por α_1',

$$\alpha_1'\Sigma\alpha_2 - \lambda_2\alpha_1'\alpha_2 - \mu_2\alpha_1'\alpha_1 = 0$$

Y dado que $\alpha_1'\alpha_2 = 0$ y $\alpha_1'\alpha_1 = 1$ se tiene que:

$$\alpha_1'\Sigma\alpha_2 - \mu_2 = 0 \tag{2.1.13}$$

además, notando que $\alpha_1'\Sigma\alpha_2 > 0$, $\Rightarrow$para cumplir con la ec. 2.1.13 se tendría que $\alpha_2 = 0$ y por lo tanto:

$$\alpha_1'\Sigma\alpha_2 = 0$$

por lo que se tiene que:

$$\mu_2 = 0$$

Sustituyendo el hecho de que $\mu = 0$ en la ec. 2.1.12 se tiene que:

$$\Sigma\alpha_2 - \lambda_2\alpha_2 = 0$$

buscando α_2 que maximice la función sujeto a que es un valor propio de Σ y su valor asociado λ_2. Así como $\alpha_1'\alpha_2$=0, es decir que son ortogonales entonces:

$$Var\,(Z_2) = Var(\alpha_{\mathbf{2}}'\mathbf{X}) = \alpha_2'\Sigma\alpha_2 = \lambda_2 \tag{2.1.14}$$

Definición: Sea $\mathbf{X} = (x_1, x_2, ...x_p)$ un vector de p variables aleatorias y $\alpha_{\mathbf{1}} = (\alpha_{11}, \alpha_{12}, ...\alpha_{1p})$ un vector de p constantes, se definen las p componentes principales como las variables aleatorias ($\alpha_{\mathbf{1}}'\mathbf{X}$, $\alpha_{\mathbf{2}}'\mathbf{X}$, $\alpha_{\mathbf{3}}'\mathbf{X}$ ... $\alpha_{\mathbf{p}}'\mathbf{X}$), tales que:

$$Var\,(Z_1) = Var\left(\alpha'_{\mathbf{1}}\mathbf{X}\right) = máx_{\alpha_1}\left\{Var\left(\alpha_{\mathbf{1}}'\mathbf{X}\right) :\ \alpha_1 \in \mathbb{R}^{\mathbb{P}},\ \alpha_1'\alpha_1 = 1\right\}$$

$$Var\left(Z_2\right) = Var\left(\alpha'_{\mathbf{2}}\mathbf{X}\right) = máx_{\alpha_2}\ \left\{Var\left(\alpha'_{\mathbf{2}}\mathbf{X}\right):\ \alpha_2 \in \mathbb{R}^{\mathbb{P}},\ \alpha'_2\alpha_1 = 0\right\}$$

...

...

$$Var\left(Z_p\right) = máx_{\alpha_p}\ \left\{Var\left(\boldsymbol{\alpha}'_{\boldsymbol{p}}\boldsymbol{X}\right):\ \alpha_p \in \mathbb{R}^{\mathbb{P}},\ \alpha'_2\alpha_1 = 0, ..., \alpha'_p\alpha_1 = 0\right\}$$

Teorema: Las p componentes principales de X adoptan la forma:

$$Z_j = \alpha'_{\mathbf{j}}\mathbf{X}\ j \in \{1, ...p\} \tag{2.1.15}$$

siendo $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$ los p valores propios ordenados de $\Sigma = D\left(X\right)$ y $\alpha_1, \alpha_2, ..., \alpha_p$ sus vectores propios asociados normalizados, es decir, es una base ortonormal de valores propios. Además las componentes no están correlacionadas ya que $Cov\left(Z_j, Z_k\right) = 0$ si $k \neq j$ y

$$Var\left(Z_j\right) = \lambda_j\ j \in \{1, 2, ..., p\} \tag{2.1.16}$$

generalizando se tiene que la función $\alpha'_{\mathbf{p}}\mathbf{X}$ no está correlacionada con $\alpha_{\mathbf{1}}\mathbf{X}, \alpha_{\mathbf{2}}\mathbf{X}, ..., \alpha_{\mathbf{p-1}}\mathbf{X}$.

Es importante mencionar que en ocasiones las componentes se calculan a partir de la matriz de correlación, esto se utiliza cuando una de las variables tiene una varianza mucho mayor o cuando las escalas son diferentes lo cual es equivalente a trabajar con variables estandarizadas.

2.2. Componentes Principales Muestrales

Sea la matriz de datos $X_{n\times p}$ la cual tiene los valores de p variables para n individuos, entonces el cálculo de las componentes principales se realiza a partir de su matriz de varianzas y covarianzas muestrales $\hat{\Sigma}$ donde los valores propios muestrales son $\hat{\lambda}_i$ y su vector propio $\hat{\alpha}$

Con ello, la i-ésima componente principal muestral se obtiene de la siguiente manera:

$$\hat{Z}_i = \hat{a}_i X = \hat{a}_{i1}x_{i1} + \hat{a}_{i2}x_{i2} + ... + \hat{a}_{ip}x_{ip}\ i \in \{1, ...n\}$$

La varianza muestral de la í-esima componente es:

$$Var\left(\hat{Z}_i\right) = \hat{\lambda}_i\ i \in \{1, 2, ..., p\}$$

y la varianza total muestral

$$Var\left(\hat{Z}_1\right) + Var\left(\hat{Z}_2\right) + ... + Var\left(\hat{Z}_p\right) = \hat{\lambda}_1 + \hat{\lambda}_2 + ... + \hat{\lambda}_p$$

finalmente la covarianza muestral

$$\hat{\sum} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Z_{1i} - \hat{Z}_1\right)\left(Z_{2i} - \hat{Z}_2\right)$$

2.3. Selección del número de componentes

No perdiendo de vista el objetivo del Análisis de Componentes Principales, el cual consiste en reducir la dimensión de los datos originales, es importante el número de componentes que resulten suficientes para explicar los datos originales.

Para ello existen diversas reglas empíricas para seleccionar el número de componentes, las cuales se mencionan a continuación:

1. Se observa la suma de varianza de las componentes que contengan los valores más altos, hasta que éste llegue a un porcentaje acumulado que el experto considere suficiente, normalmente se llega cerca de un ochenta por ciento o mayor.

2. Se conservan las componentes a partir de la matriz de correlaciones, suponiendo de las variables observadas tienen varianza uno. Por lo tanto una componente principal con valor propio menor a uno explica menos variabilidad, es decir, se retiene las primeras m componentes principales tales que $\lambda_m \geq 1$, donde $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$ son los valores propios de la matriz de correlaciones, que también son la varianza de las componentes.

3. Se representa un gráfico de sedimentación de las componentes, el cual es un contraste gráfico, se retienen todas las componentes que tienen las pendientes más altas, dejando fuera las que tienen pendientes cercanas a cero o aproximadamente iguales, la idea es buscar un "codo" en el gráfico.

4. Se eligen las componentes que expliquen más que un cierto porcentaje acumulado de la varianza.

5. Cuando la diferencia de la variabilidad explicada con la variabilidad explicada por la componente anterior sea más grande que alguna cota.

Es importante mencionar que es posible generar pruebas estadísticas para seleccionar el número de componentes, a continuación se muestran algunas pruebas nulas para $\lambda_1 > \lambda_2 > ... > \lambda p$:

$$H_0 : \lambda_j = 0$$

$$H_0 : \lambda_j = \lambda_{j+1} = ... = \lambda p = 0$$

$$H_0 : \frac{\sum_{i=1}^{q} \lambda_i}{\sum_{i=1}^{p} \lambda_i} = cte$$

Capítulo 3

Análisis de Componentes Principales

Las estimaciones de la pobreza en México se sabe que se calculan a partir de datos que contiene el Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MCS-ENIGH) que realiza el Instituto Nacional de Estadística y Geografía (INEGI). Los datos que se utilizan en el presente trabajo son referentes al levantamiento que se realizó del día 11 de agosto al 28 de noviembre de 2014, y que fueron publicados el 16 de julio de 2015.

El objetivo del MCS es contar con información detallada del monto, la estructura y la distribución de los ingresos de los hogares; el acceso a la salud, a la seguridad social y a la educación de los integrantes del hogar; la seguridad alimentaria de los hogares; las características de las viviendas que ocupan y los servicios con que cuentan estas viviendas,[1] el cual se realizó bajo el siguiente planteamiento estadístico:

- Tamaño de muestra: 64 mil viviendas
- Unidad de muestreo: Vivienda
- Unidades de observación: Hogar, personas, viviendas
- Marco muestral: Marco Nacional de Viviendas 2012 del INEGI, construido a partir de la información demográfica y cartográfica que se obtuvo del Censo de Población y Vivienda 2010.
- Esquema muestral: Probabilístico, unietápico, estratificado y por conglomerados.

[1] Instituto Nacional de Estadística y Geografía INEGI

En cumplimiento de las disposiciones establecidas en la Ley General de Desarrollo Social (LGDS), el CONEVAL publica las estimaciones de pobreza en México a nivel nacional y para cada una de las entidades federativas.

La siguiente tabla contiene las estimaciones de la pobreza para cada entidad federativa que presentó CONEVAL correspondiente a 2014 publicadas el día 23 de julio de 2015.

Cuadro 3.1: Resultados de la medición de la Pobreza 2014

Entidad Federativa	**Pobreza**		**Pobreza Extrema**	
	Porcentaje	**Miles de personas**	**Porcentaje**	**Miles de personas**
Aguascalientes	34.8	442.9	2.1	26.7
Baja California	28.6	984.9	3.1	105.5
Baja California Sur	30.3	226.2	3.9	29.5
Campeche	43.6	391.0	11.1	99.2
Coahuila	30.2	885.8	3.7	109.7
Colima	34.3	244.9	3.4	24.4
Chiapas	76.2	3,961.0	31.8	1,654.4
Chihuahua	34.4	1,265.5	5.4	200.3
Distrito Federal	28.4	2,502.5	1.7	150.5
Durango	43.5	761.2	5.3	93.0
Guanajuato	46.6	2,683.3	5.5	317.6
Guerrero	65.2	2,315.4	24.5	868.1
Hidalgo	54.3	1,547.8	12.3	350.5
Jalisco	35.4	2,780.2	3.2	253.2
México	49.6	8,269.9	7.2	1,206.9
Michoacán	59.2	2,708.6	14.0	641.9
Morelos	52.3	993.7	7.9	149.3
Nayarit	40.5	488.8	8.5	102.1
Nuevo León	20.4	1,022.7	1.3	66.7
Oaxaca	66.8	2,662.7	28.3	1,130.3
Puebla	64.5	3,958.8	16.2	991.3
Querétaro	34.2	675.7	3.9	76.1
Quintana Roo	35.9	553.0	7.0	107.6
San Luis Potosí	49.1	1,338.1	9.5	258.5
Sinaloa	39.4	1,167.1	5.3	155.8
Sonora	29.4	852.1	3.3	95.6
Tabasco	49.6	1,169.8	11.0	260.3
Tamaulipas	37.9	1,330.7	4.3	151.6
Tlaxcala	58.9	745.1	6.5	82.6
Veracruz	58.0	4,634.2	17.2	1,370.5
Yucatán	45.9	957.9	10.7	223.2
Zacatecas	52.3	819.8	5.7	89.4
Total	**46.2**	**55,341.6**	**9.5**	**11,442.3**

Fuente: Estimaciones de CONEVAL con base en el MCS-ENIGH 2012 y 2014.

Además, se muestra la tabla de frecuencias de las carencias por entidad federativa:

Cuadro 3.2: Porcentajes por carencia social, según entidad federativa

Entidad	**Alimentos**	**Servicios de salud**	**Calidad y espacios**	**Rezago edu**	**Servicios básicos**	**Seg. social**	**Ingreso**
Aguascalientes	21.6	12.5	3.3	14.4	3.6	43.2	43.7
Baja C.	17.2	19.4	10.6	15.4	12.1	51.8	35.2
Baja C. Sur	24.6	14.2	16.5	14.9	12.4	46.6	35.9
Campeche	24.3	12.5	**19.5**	18.8	38.8	60.1	47.6
Coahuila	22.0	15.6	5.0	12.5	5.6	34.2	41.3
Colima	25.4	12.7	10.9	17.5	9.6	51.9	40.7
Chiapas	27.5	**20.7**	**26.9**	**30.7**	**57.4**	82.8	**78.7**
Chihuahua	18.7	14.6	7.9	17.3	7.9	43.4	46.4
DF	11.7	19.9	5.4	8.8	1.7	46.3	36.4
Durango	19.9	16.5	5.8	15.5	13.0	51.3	53.9
Guanajuato	22.9	15.4	9.8	21.0	14.9	57.9	55.1
Guerrero	**38.5**	19.2	**32.9**	**26.8**	**58.0**	**78.1**	**67.9**
Hidalgo	**31.7**	17.3	9.2	19.1	27.0	68.9	59.4
Jalisco	16.5	19.1	6.6	17.7	7.0	49.6	43.3
México	21.3	19.7	10.3	15.3	12.4	60.6	58.9
Michoacán	**34.7**	**26.2**	15.4	**27.6**	26.6	71.3	63.3
Morelos	26.9	16.6	13.4	16.6	24.6	66.2	58.4
Nayarit	24.1	16.3	10.1	17.4	15.3	54.4	47.0
Nuevo León	14.2	13.7	4.6	10.8	4.3	33.4	29.5
Oaxaca	**36.1**	**19.9**	**24.5**	**27.2**	**60.5**	**77.9**	**68.8**
Puebla	23.9	**21.2**	18.9	22.9	30.6	**75.2**	**69.7**
Querétaro	15.8	15.8	8.9	16.4	14.8	54.3	42.0
Quintana Roo	23.2	18.5	**18.4**	15.1	18.1	51.5	42.1
San Luis P.	21.6	10.7	11.0	18.4	28.1	59.1	56.7
Sinaloa	29.6	15.2	10.8	19.1	18.0	49.3	46.3
Sonora	24.9	14.4	10.1	12.1	8.9	41.8	36.6
Tabasco	**45.0**	16.9	13.4	17.0	**43.9**	**72.7**	51.9
Tamaulipas	19.5	15.0	8.2	16.0	11.5	45.5	49.1
Tlaxcala	24.0	17.5	9.4	14.9	12.1	**71.5**	**66.6**
Veracruz	30.0	**21.7**	16.8	**27.8**	**40.0**	68.5	63.0
Yucatán	18.4	14.5	17.5	21.8	**40.4**	54.4	52.8
Zacatecas	16.8	14.9	4.9	21.6	13.3	63.4	59.7
Total	**23.4**	**18.2**	**12.3**	**18.7**	**21.2**	**58.5**	**53.2**

Fuente: Estimaciones de CONEVAL con base en el MCS-ENIGH 2012 y 2014.

A continuación se enlistan las variables de las carencias sociales con las que se realizó el análisis, dichas variables son dicotómicas y toman el valor de 0 cuando el individuo no es carente y 1 en caso de que el individuo si presente la carencia.

1. Carencia por Rezago educativo
2. Carencia por acceso a los servicios de salud
3. Carencia por acceso a la seguridad social
4. Carencia por calidad y espacios en la vivienda
5. Carencia por acceso a los servicios básicos en la vivienda
6. Carencia por acceso a la alimentación
7. Población con ingreso inferior a la línea de bienestar

Elaborando el Análisis de Componentes Principales utilizando la información que publicó el CONEVAL, la cual corresponde a 216,250 registros de los cuales se eliminaron 41 registros debido a que no cuentan con el total de la información para identificar si cuentan con las carencias o no, por lo tanto el ejercicio se hizo con 216,209 registros.

3.1. Análisis de Componentes Principales con la Matriz de Covarianzas

De manera preliminar, en el Cuadro 3.3 se muestra la matriz de covarianzas con las variables de las carencias sociales, de esta manera vemos que las variables se relacionan de manera positiva en todos los casos.

Cuadro 3.3: Matriz de covarianzas

	Alimentos	Salud	Calidad y espacios	Rez. educativo	Servicios básicos	Seg. social	Ingreso
Alimentos	0.185	0.005	0.026	0.017	0.037	0.032	0.047
Salud	0.005	0.141	0.003	0.006	0.002	0.063	0.010
Calidad y espacios	0.026	0.003	0.111	0.015	0.043	0.028	0.031
Rez. educativo	0.017	0.006	0.015	0.152	0.034	0.011	0.027
Servicios básicos	0.037	0.002	0.043	0.034	0.173	0.055	0.048
Seg social	0.032	0.063	0.028	0.011	0.055	0.243	0.075
Ingreso	0.047	0.010	0.031	0.027	0.048	0.075	0.249

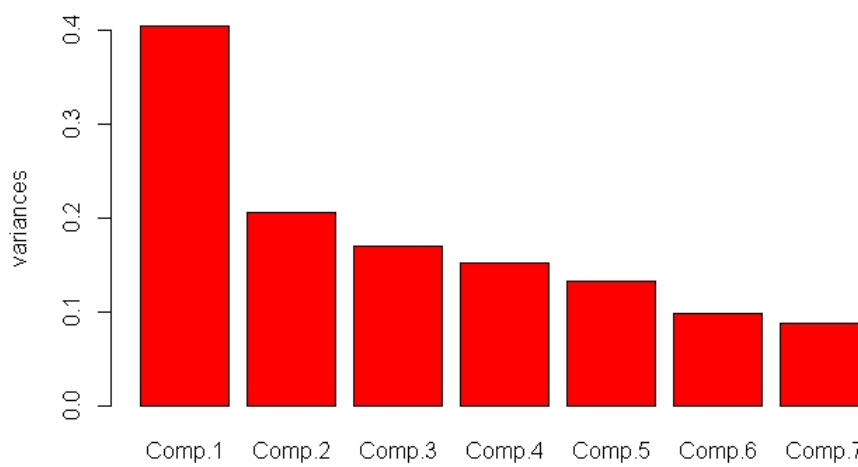
Aplicando el análisis de componentes principales con la matriz de covarianzas, se obtuvieron los siguientes resultados:

Cuadro 3.4: Resumen de resultados

	1	**2**	**3**	**4**	**5**	**6**	**7**
Desviación estándar	0.636	0.455	0.412	0.391	0.365	0.315	0.296
Proporción de varianza	0.322	0.165	0.136	0.122	0.106	0.079	0.069
Varianza acumulada	0.322	0.487	0.623	0.745	0.851	0.930	1
Varianza	0.404	0.207	0.170	0.153	0.133	0.099	0.088

Respecto a la varianza obtenida y con el apoyo del gráfico de barras de la Figura 3.1.1, es posible visualizar de una manera sencilla la importancia de cada componente, ya que las alturas representan la cantidad de información que tienen es decir, permite visualizar la varianza asociada a cada componente.

Figura 3.1.1: Varianza de las Componentes Principales



Como se esperaba, la primera componente es la que tiene mayor cantidad de varianza con un 0.404, ésto en términos de porcentaje se tiene que la primera componente tiene un 32.2 % de información, la segunda 16.5 % y la tercera 13.6 %, acumulando un total de 62.30 % de la varianza total de la información.

Posteriormente, en el Cuadro 3.5 se muestran los coeficientes de las componentes, los cuales son los vectores propios asociados a cada valor propio.

Cuadro 3.5: Coeficientes por componente principal

Carencia	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Alimentación	0.312	**0.365**	**0.307**	0.790	0.205	0.066	0.061
Servicios de salud	0.176	**-0.493**	0.082	0.065	0.418	0.718	0.159
Calidad y espacios	0.208	0.133	0.202	-0.071	-0.215	-0.340	-0.855
Rezago educativo	0.174	0.258	0.292	-0.499	0.708	0.247	-0.082
Servicios básicos	**0.372**	0.219	**0.472**	-0.329	-0.472	-0.208	0.465
Seguridad social	**0.574**	**-0.628**	0.078	0.029	-0.109	0.495	-0.112
Ingreso	**0.574**	0.311	**-0.737**	-0.088	0.039	-0.128	0.062

Los números el color color rojo son los de valor absoluto de mayor magnitud de signo negativo y los de color verde son los de valor absoluto de mayor magnitud de signo positivo

A continuación se encuentra la representación analítica de la primer componente:

$$\begin{aligned} C_1 = 0.312 * (Alimentación) + 0.176 * (Servicios\,de\,salud) \\ +0.208 * (Calidad\,y\,espacios) + 0.174 * (Rezago\,educativo) \\ +0.372 * (Sevicios\,básicos\,de\,la\,vivienda) + 0.574 * (Seguridad\,Social) \\ +0.574 * (Ingreso) \end{aligned}$$

Para interpretar cada componente principal, es importante tomar en cuenta la magnitud y dirección de los coeficientes de las variables originales, que en éste caso se trata de las carencias sociales las cuales son variables dicotómicas que toman el valor de 0 cuando el individuo no es carente y 1 en caso de que el individuo si presente la carencia.

Entonces, si se tiene que para una componente los coeficientes asociados a las carencias de mayor peso son de signo positivo, se tendría que la población que toma valores grandes en ésta componente tienen las carencias, y en el caso contrario donde los coeficientes de mayor peso sean de signo negativo, se tendría que la población que toma valores grandes no presentan éstas carencias.

El siguiente paso fue identificar a las variables que predominan por componente, esto se hizo a través de los coeficientes con los valores absolutos de mayor magnitud, las cuales se muestran en el Cuadro 3.6.

Cuadro 3.6: Variables con mayor valor absoluto por componente

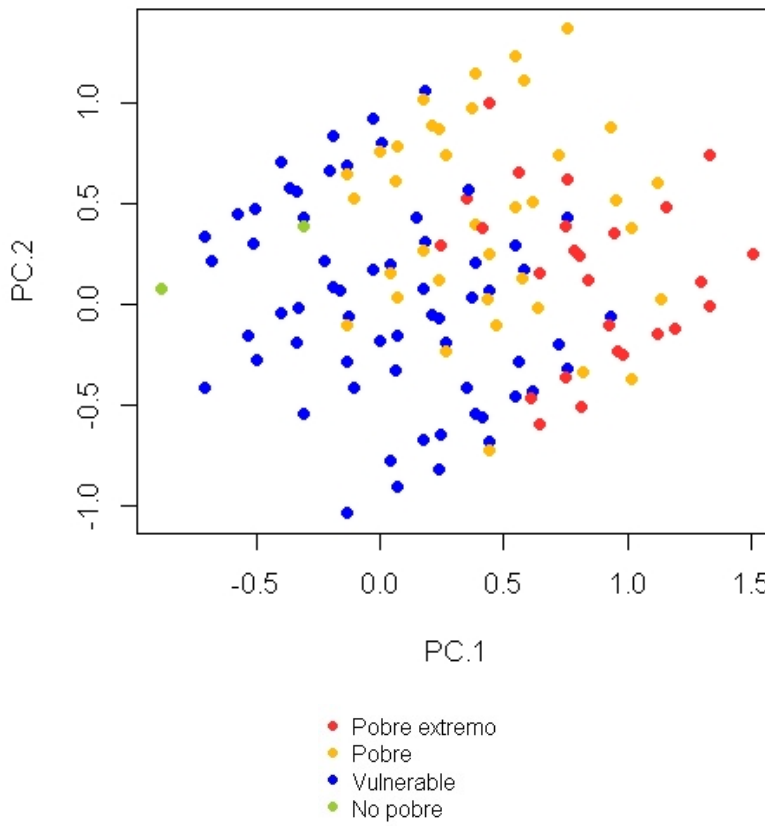
1	**2**	**3**
Seguridad social	Seguridad social	Ingreso
Ingreso	Servicios de salud	Servicios básicos de la vivienda

En resumen, para las primeras tres componentes las variables con coeficientes de mayor valor absoluto son:

1. Primera Componente: Carencia por acceso a la seguridad social e ingresos por debajo de la línea de bienestar
2. Segunda Componente: Carencia por acceso a la seguridad social y servicios de salud
3. Tercera Componentes: Ingresos por debajo de la línea de bienestar y la carencia de acceso a servicios básicos de la vivienda

Siguiendo con el análisis, en la Figura 3.1.2 se encuentra la gráfica con la representación de la población con las primeras dos componentes como ejes, donde el eje de las abscisas será la primera componente y el eje de las ordenadas la segunda componente, las cuales en conjunto explican un 48.7 %.

Figura 3.1.2: Representación de la población con las primeras dos componentes



Es posible observar que se obtuvieron tres grupos de puntos, los cuales con el fin de facilitar el análisis de la gráfica se denominaran conforme a la altura que tienen en la componente 2 y de manera descendente como grupo 1,2 y 3, por lo que el primer grupo corresponde al que se encuentra en la parte superior del gráfico, además no se omite mencionar que al interior de cada grupo se puede identificar mediante colores la composición de la población.

Respecto al grupo 1, se observa que lo compone en su mayoría población con individuos que son vulnerables y pobres y es donde se encuentra la población

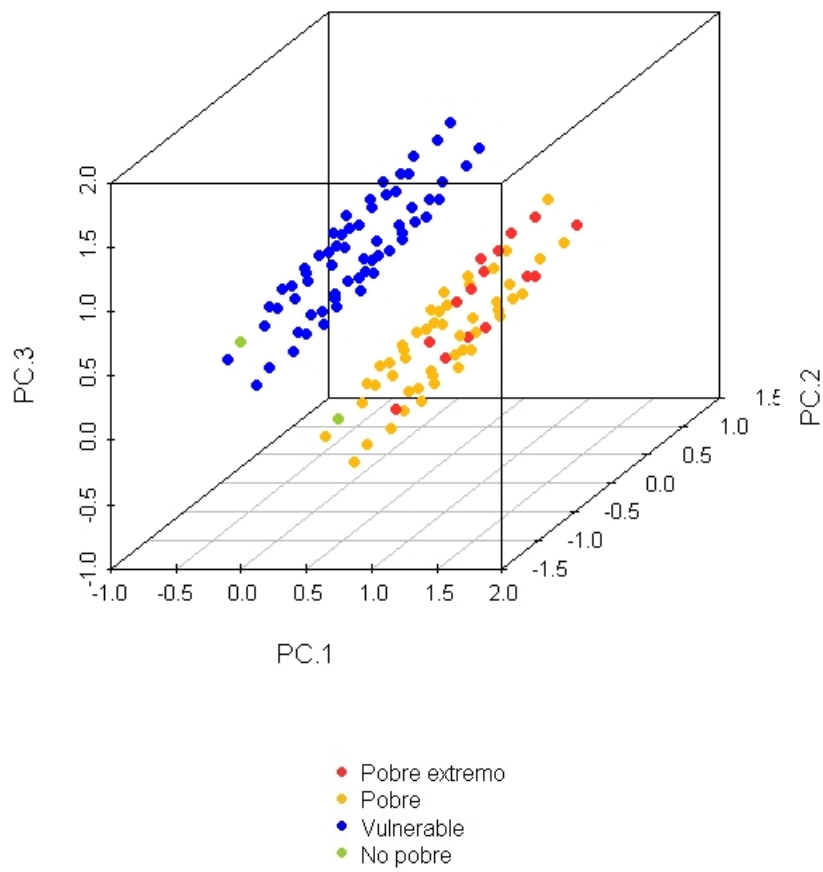
no pobre, mientras que en los grupos 2 y 3 se hace visible la población que se encuentra en condiciones de pobreza extrema.

Ahora bien si se observa la primer componente, se identifica que existe una diferencia en la composición de la población ya que, en la parte negativa tiende a encontrarse la población vulnerable mientras que en la parte positiva en su mayoría es población que se encuentra en condiciones de pobreza y pobreza extrema y dado que las carencias que predominan son la seguridad social e ingreso (ambas con signo positivo), es congruente ya que indica que la población en pobreza y pobreza extrema son las que tienden a tener éstas carencias.

Para la segunda componente donde las variables que dominan son la seguridad social y los servicios de salud (ambas de signo negativo), no hay una tendencia definida para la población vulnerable y pobre ya que este tipo de población se encuentra a lo largo de toda la componente, sin embargo la población en condición de pobreza extrema en su mayoría tienen ordenadas negativas lo cual indica que éstas son las que tienden a ser carentes de seguridad social y servicios de salud.

Incluyendo la tercera componente, se generó la gráfica de la Figura 3.1.3 donde el eje de las abscisas es la primera componente, el eje de las ordenadas la segunda componente y el eje de las cotas la tercera componente, las cuales en conjunto explican un total de 62.3 % de la varianza de los datos.

Figura 3.1.3: Representación de la población con las primeras tres componentes principales



A diferencia de la representación en dos dimensiones, se puede observar que en este caso se obtienen dos conjuntos de puntos, por un lado se encuentra la población en condiciones de pobreza y pobreza extrema que para fines prácticos será denominado grupo 1 y el otro conformado por la población vulnerables y no pobre grupo 2.

Siguiendo con el análisis de la Figura 3.1.3, es importante mencionar que es congruente con lo observado en la Figura 3.1.2 ya que se observa que en la

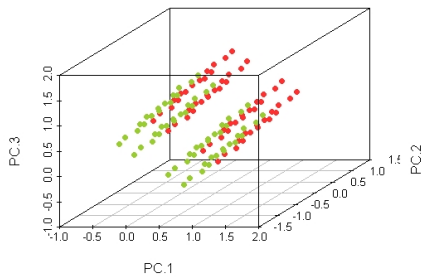
primera componente como eje de las abscisas el grupo 1 tiende a concentrarse en la parte positiva de este eje mientras que el grupo 2 tiende a concentrarse en la parte negativa de esta componente en la que los coeficientes de mayor peso son las carencias de seguridad social e ingreso (ambas con signo positivo), lo que indica que el grupo 1 compuesto por la población en condiciones de pobreza y pobreza extrema tienen mayor grado de presentar estas carencias, caso contrario para el grupo 2 compuesto por la población no pobres y vulnerables.

En la tercera componente como eje de las cotas se genera la discriminación entre estos dos grupos, se tiene que el grupo 2 tiende a tener valores de mayor magnitud en comparación al grupo 1 y dado que los coeficientes con mayor valor absoluto en esta componente son el ingreso (con signo negativo) y los servicios básicos de la vivienda (con signo positivo), se concluye que la población del grupo 2 tienen mayor grado de no presentar la carencia de ingresos pero mayor tendencia a tener la carencia de servicios básicos de la vivienda en comparación al grupo 1.

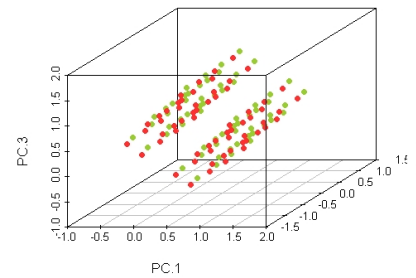
Finalmente, si en lugar de identificar mediante colores la clasificación de la pobreza, se identifica para cada una de las carencias, se obtienen los siguientes gráficos:

Figura 3.1.4: Población diferenciada por carencia social con los primeros tres componentes

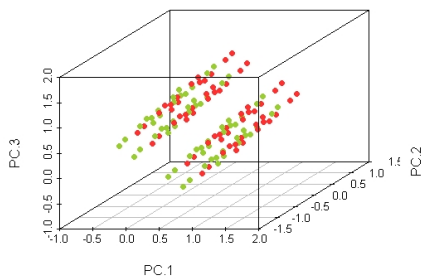
(a) Alimentación



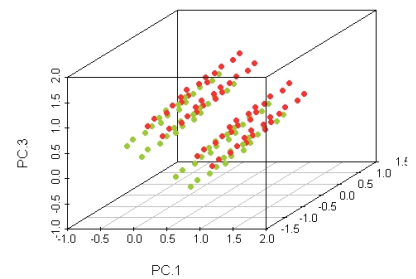
(b) Servicios de la salud



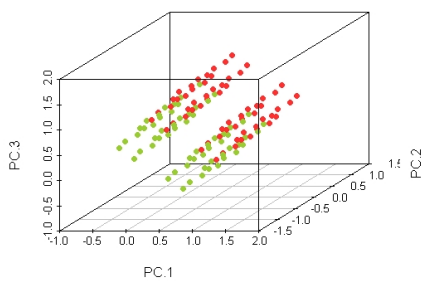
(c) Calidad y espacios de la vivienda



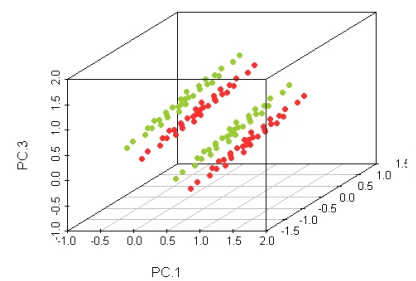
(d) Rezago educativo



(e) Servicios básicos de la vivienda



(f) Acceso a la seguridad social



(g) Ingreso

Las gráficas anteriores se componen con las primeras tres componentes, diferenciando las más relevantes son las correspondientes a:

- Carencia por acceso a la Seguridad Social ya que como se puede observar en la Figura 3.1.4f segmenta a la población en cuatro grupos, dos de ellos se componen de población que tiene la carencia mientras que los dos restantes es población que no tiene esta carencia social.
- Carencia por ingresos menor a la línea de bienestar en la Figura 3.1.4g asocia completamente a la población en dos grupos bien definidos.

Lo anterior tiene sentido teniendo en cuenta que las componentes uno y dos están predominadas por la carencias por acceso a la seguridad social, mientras que la componente tres está asociado a la variable de ingreso.

3.2. Análisis de Componentes Principales con la Matriz de Correlación

El método que se utilizó para el cálculo de la matriz de correlación fue la correlación policórica, ya que ésta es una medida de asociación para variables ordinales.

La correlación policórica, propuesta por Pearson (1901), la idea fundamental consiste en suponer que las variables tienen una distribución normal multivariada, la cual consiste en resolver la ec. 3.2.1 para $\sum$.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \mid \sum \mid^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)\sum^{-1}(x-\mu)\right] \prod_{i=1}^{p} I(x_i) \qquad (3.2.1)$$

donde $\mu \in \mathbb{R}^p$ y $\sum$ es una matriz de tamaño $p \times p$ definida positiva; por ejemplo, si $p = 2$, entonces se tendría que:

$$\sum = \begin{pmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{21} & \sigma_2 \end{pmatrix}$$

$$y\, por\, lo\, tanto,\, el\, coeficiente\, de\, correlación\, policórica\, es:\ \rho_{1,2} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

En el Cuadro 3.7 se encuentra la matriz de correlaciones para las variables de las carencias sociales,

Cuadro 3.7: Matriz de correlación policórica

	Alimentos	Salud	Calidad y espacios	Rez. educativo	Servicios básicos	Seg. social	Ingreso
Alimentación	1	0.058	0.343	0.189	0.355	0.267	0.377
Salud	0.058	1	0.064	0.083	0.030	0.721	0.101
Calidad y espacios	0.343	0.064	1	0.237	0.548	0.367	0.384
Rezago educativo	0.189	0.083	0.237	1	0.369	0.108	0.256
Servicios básicos	0.355	0.030	0.548	0.369	1	0.485	0.407
Seguridad Social	0.267	0.721	0.367	0.108	0.485	1	0.466
Ingreso	0.377	0.101	0.384	0.256	0.407	0.466	1

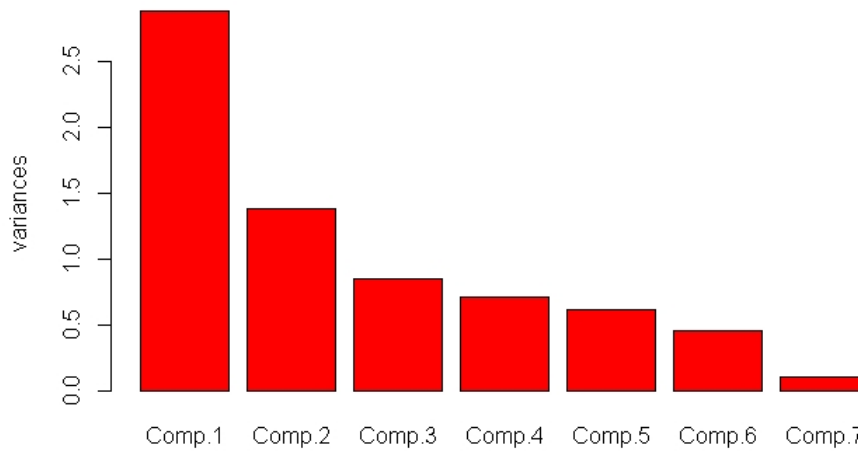
Aplicando el Análisis de Componentes Principales, con la matriz de correlación se obtienen los siguientes resultados:

Cuadro 3.8: Resumen de resultados

	1	**2**	**3**	**4**	**5**	**6**	**7**
Desviación estándar	1.696	1.176	0.924	0.846	0.782	0.676	0.318
Proporción de varianza	0.411	0.198	0.122	0.102	0.087	0.065	0.014
Varianza acumulada	0.411	0.609	0.731	0.833	0.92	0.986	1
Varianza	2.877	1.383	0.854	0.716	0.612	0.458	0.101

Haciendo uso del resultado obtenido respecto a la cantidad de varianza, en la Figura 3.2.1 se muestra la cantidad de información que tiene cada componente para así poder considerar el número de componentes necesarias para la interpretación del análisis .

Figura 3.2.1: Varianza de los Componentes Principales



En términos de porcentaje, la varianza asociada a la primera componente es de 41.1 %, para la segunda es de 19.8 % y la tercera es de 12.2 %, teniendo un acumulando de varianza de éstas tres componentes es de 73.1 % de la variación del total de los datos.

Posteriormente se calcularon los vectores propios asociados a cada valor propio, es decir, los coeficientes de las componentes los cuales pueden verse en el Cuadro 3.9.

Cuadro 3.9: Coeficientes por componente principal

Carencia	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Alimentación	-0.343	-0.221	**-0.396**	**0.662**	**0.484**	0.043	-0.011
Servicios de salud	-0.241	**0.726**	0.182	0.113	0.184	-0.166	**0.553**
Calidad y espacios	-0.410	-0.232	-0.155	**-0.508**	0.230	**-0.664**	0.003
Rezago educativo	-0.268	-0.264	**0.861**	0.264	0.039	-0.132	-0.162
Servicios básicos	**-0.449**	-0.240	0.068	-0.384	0.101	**0.676**	0.344
Seguridad Social	**-0.457**	**0.479**	-0.067	-0.102	-0.053	0.186	**-0.712**
Ingreso	**-0.414**	-0.107	-0.183	0.247	**-0.814**	-0.134	0.201

Los números el color color rojo son los de valor absoluto de mayor magnitud de signo negativo y los de color verde son los de valor absoluto de mayor magnitud de signo positivo

Cada una de las columnas contiene los coeficientes de una combinación lineal de las variables originales que se utilizaron para el cálculo de las componentes principales, como ejemplo se muestra la forma analítica de la primer componente:

$$C_1 = -0.334 * (Alimentación) - 0.241 * (Servicios\,de\,salud)$$
$$-0.410 * (Calidad\,y\,espacios\,de\,la\,vivienda) - 0.268 * (Rezago\,educativo)$$
$$-0.449 * (Sevicios\,básicos\,de\,la\,vivienda) - 0.457 * (Seguridad\,Socia)$$
$$-0.414 * (Igreso)$$

Ahora bien, en el Cuadro 3.10 se muestran las variables asociadas a los coeficientes por componente cuyos valores absolutos son los de mayor magnitud.

Cuadro 3.10: Variables con los coeficientes de mayor valor absoluto por componente

1	**2**	**3**
Seguridad social	Seguridad salud	Rezago educativo
Servicios básicos de la vivienda	Seguridad social	Alimentación
Ingreso		

En resumen:

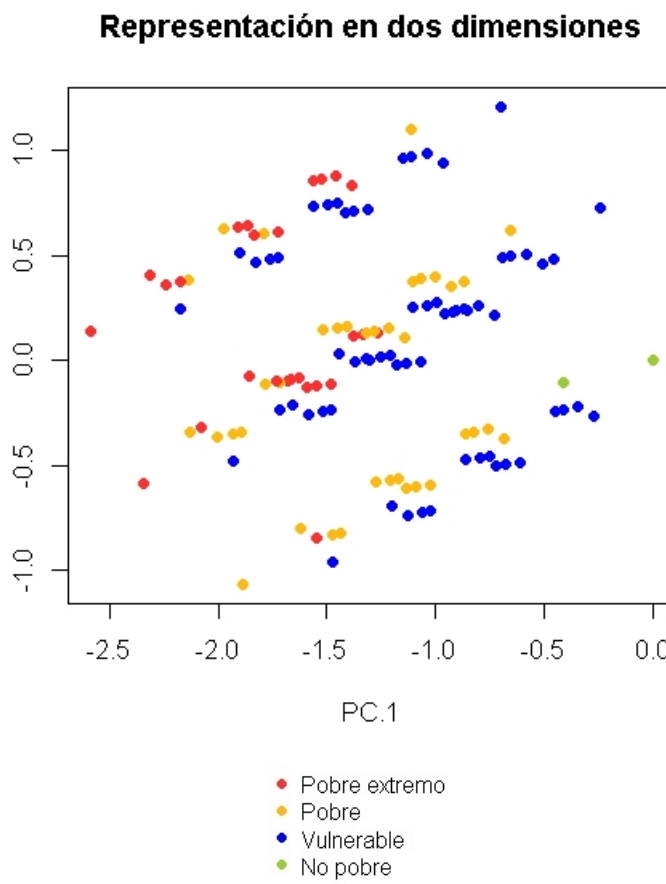
1. En la primer componente predominan las carencias de seguridad social, servicios básicos de la vivienda, e ingreso, las cuales tienen coeficientes negativos, es decir que a valores mas pequeños de la componente representan a individuos con éstas carencias

2. En la segunda componente predominan las carencias de servicios de salud y seguridad social con coeficientes positivos, por lo que a valores grandes de la componente representarán a la población que tiene éstas carencias.

3. Finalmente en la tercer componente predominan el rezago educativo y el acceso a la alimentación, en éste caso el rezago educativo tiene un coeficiente positivo, y el acceso a la educación un coeficiente negativo es decir, que para valores grandes en ésta componente se encontrará a la población con rezago educativo pero sin carencia alimentaria.

Utilizando todos los individuos a los cuales se les ha evaluado el nivel de bienestar económico y las carencias sociales, se elaboró la gráfica con las primeras dos componentes como ejes, donde el eje de las abscisas corresponde a la primer componente y el eje de las ordenadas a la segunda componente, las cuales en conjunto explican un 60.9 % , véase la Figura 3.2.2.

Figura 3.2.2: Representación de la población con las primeras dos componentes



Como se puede observar, se obtienen tres grupos en los cuales se puede identificar mediante colores la clasificación conforme a los resultados del CONEVAL es decir, a la población en condiciones de pobreza extrema, pobreza, vulnerables y no pobres; cabe destacar que más adelante se mostrará el ejercicio por carencia.

Y del mismo modo que en el análisis de componentes principales con la matriz de covarianza, se nombrarán a los grupos de forma descendente como grupo 1,2 y 3 de tal manera que el primero corresponde al grupo que se encuentra en la parte superior del gráfico.

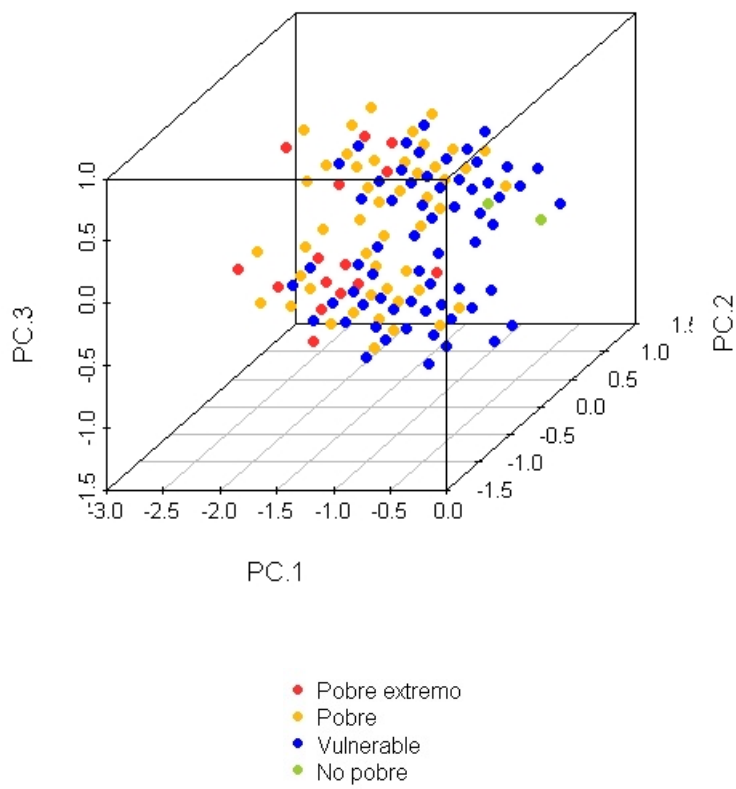
En el grupo 1, se observa que lo compone en su mayoría población con individuos que son pobres extremos y vulnerables, mientras que en el grupo 2 se observa que cuenta con los diversos tipos de población, excepto la población no pobre y finalmente para el grupo 3 se observa que se encuentra dominado por pobres y vulnerables, además es el grupo donde se encuentra la población no pobre.

En la primera componente, es posible notar que en la parte izquierda tienden a encontrarse la población en condiciones de pobreza extrema mientras que en la parte derecha del gráfica se encuentra la población no pobre, y destacando que las variables que predominan son los servicios básicos de la vivienda, seguridad social e ingreso (las tres con coeficiente de signo negativo), es congruente ya que ésto indica que la población en pobreza extrema tiende a tener más éstas carencias en comparación de la población no pobre.

Respecto a la segunda componente las carencias con mayor peso son las de acceso a los servicios de salud y seguridad social (ambas con signo positivo), además es importante mencionar que ésta componente es la que marca la diferencia de agrupamiento y de la clasificación de los individuos; tomando los extremos se puede ver que en la parte superior prevalecen en su mayoría la población en condición de pobreza extrema mientras que en la parte inferior se mantiene la población no pobre. Por lo anterior, se puede concluir que los pobres extremos tienen mayor tendencia a tener estas carencias en comparación de los no pobres.

Ahora bien, si se agrega la tercera componente, donde el eje de las abscisas será la primera componente y el eje de las ordenadas la segunda componente y finalmente el eje de las cotas la tercera componente, las cuales en conjunto explican un 73.1 %, se obtiene la gráfica de la Figura 3.2.3.

Figura 3.2.3: Representación de la población con las primeras tres componentes



El cambio notorio al agregar la tercera componente es que se obtienen 2 grupos y de la misma manera que las gráficas anteriores se identifica mediante colores la clasificación de la población.

En general, en la Figura 3.2.3, se observa que ambos grupos contienen población en condiciones de pobreza, pobreza extrema y vulnerable es decir, no se tiene una tendencia definida para ésta población sin embargo, en el primer grupo que se encuentra en la parte superior del gráfico se tiene a la población no pobre.

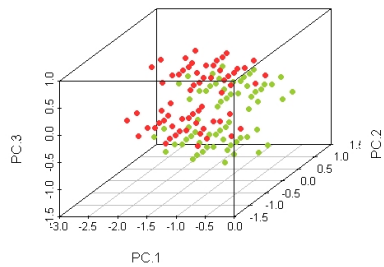
Finalmente, recordando que las variables con mayor influencia en la tercer componente corresponden al de rezago educativo (signo positivo) y alimentación

(signo negativo) indica que la población no pobre tiende a tener rezago educativo pero no la carencia de acceso a la alimentación.

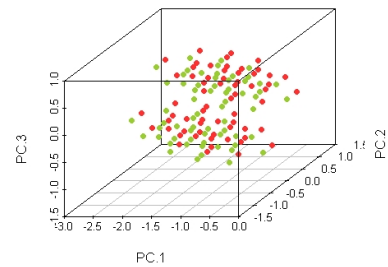
Ahora bien, si en lugar de identificar mediante colores la clasificación de la pobreza se identifica para cada una de las carencias, respetando las primeras tres componentes, observamos lo siguiente:

Figura 3.2.4: Población diferenciada por carencia social con los primeros tres componentes

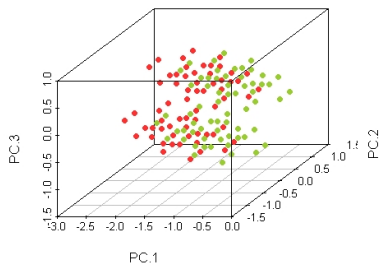
(a) Alimentación



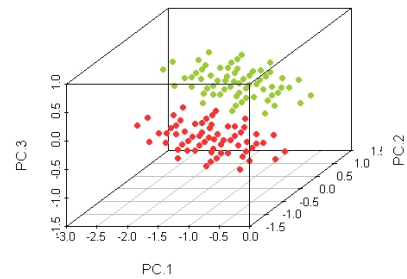
(b) Servicios de salud



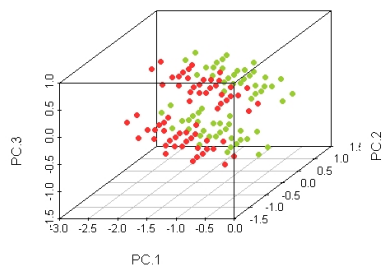
(c) Calidad y espacios de la vivienda



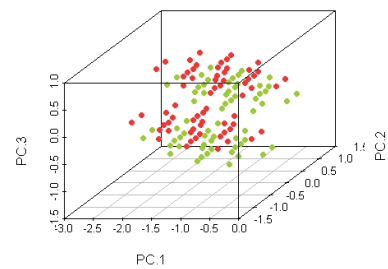
(d) Educación



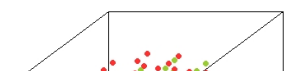
(e) Servicios básicos de la vivienda



(f) Seguridad social



(g) Ingreso



Las gráficas anteriores se componen con las primeras tres componentes, donde la más relevante es la de la carencia por rezago educativo, como se puede observar en la Figura 3.2.4d divide a la población en dos grupos, uno de ellos tiene la carencia mientras que el otro no tiene ésta carencia social, además es la tiene mayor peso en la tercer componente.

En conclusión, el Análisis de Componentes Principales nos muestra que al hacer una combinación lineal entre las variables originales es posible hacer una representación gráfica en dos o tres dimensiones.

Es importante mencionar que el resultado del análisis de componentes principales depende de la escala de las variables originales, es decir, si se llega a cambiar de escala una variable (dejando las demás fijas) las componentes se modificarán, cambiando incluso la posible interpretación, es aquí donde se encuentra la diferencia entre el uso de la matriz de correlación y la matriz de covarianza, ya que con la matriz de covarianzas se trabaja cuando las variables de estudio tienen la misma escala mientras que la matriz de correlación trabaja con las variables estandarizadas, sin embargo, si dichas variables tienen la misma escala y el análisis es utilizado con el fin de identificar la estructura de los datos, entonces ambas matrices se pueden utilizar de manera indiferente.

Capítulo 4

Modelos de regresión

Uno de los aspectos más relevantes de la Estadística es el análisis de la relación o dependencia entre variables. Frecuentemente resulta de interés conocer el efecto que una o varias variables pueden causar sobre otra e incluso el poder predecir los valores de una variable a partir de otra u otras.

Los métodos de regresión lineal estudian la construcción de modelos para explicar o representar la dependencia de una variable denominada de respuesta o dependiente y de las variables explicativas o independientes que pueden ser tanto continuas (covariables) como categóricas (factores). A continuación mostraremos la formulación de estos modelos.

Un modelo de regresión lineal clásico para estimar Y_i en función de X es:

$$E[Y] = \alpha + \beta \mathbf{X} \tag{4.0.1}$$

Donde el parámetro α es la ordenada al origen del modelo (punto de corte con el eje Y) y β un parámetro desconocido a estimar es la pendiente, que puede interpretarse como el incremento de la variable dependiente por cada incremento en la variable independiente.

Sin embargo el modelo que plantea la ec. 4.0.1 supone que las variables explicativas tienen una distribución normal con varianza constante, ante esto se presenta una alternativa cuando los supuestos mencionados anteriormente no se cumplen, este enfoque es conocido como Modelos Lineales Generalizados (MLG), los cuales consisten en modelar a Y_i de la siguiente forma:

$$Y_i = F(\alpha + \beta \mathbf{X}) \tag{4.0.2}$$

con F función monótona creciente, o bien en términos de su función inversa

como:

$$F^{-1}(E[Y_i]) = \alpha + \beta \mathbf{X} \tag{4.0.3}$$

es decir, se busca una función F cuya inversa transforme $E[Y_i]$ y posteriormente, modelar linealmente esta transformación.

Según se elija una determinada F, se tiene distintas formulaciones. A continuación se enlistan las más conocidas:

- **Transformación Logit.**

$$\ln \frac{p(x)}{1-p(x)} = \alpha + \beta x \tag{4.0.4}$$

donde

$$p(x) = F(\alpha + \beta x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \tag{4.0.5}$$

desarrollando, en términos de $F^{-1}(p(x))$, se tendría que:

$$p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

$$p(x) \cdot [1 + \exp(\alpha + \beta x)] = \exp(\alpha + \beta x)$$

$$p(x) + p(x) \cdot [\exp(\alpha + \beta x)] = \exp(\alpha + \beta x)$$

$$p(x) = \exp(\alpha + \beta x) - p(x) \cdot [\exp(\alpha + \beta x)]$$

$$p(x) = \exp(\alpha + \beta x)[1 - p(x)]$$

$$\frac{p(x)}{1-p(x)} = \exp(\alpha + \beta x)$$

$$\ln \frac{p(x)}{1-p(x)} = \alpha + \beta x$$

por lo tanto:

$$F^{-1}\left(p(x)\right) = \ln \frac{p(x)}{1-p(x)}$$

Con lo que se asegura que $p\left(x\right)$ está acotada entre 0 y 1.

- **Transformación probit**

La transformación probit consiste en utilizar como función de transformación la inversa de la función de distribución de una normal estándar $\mathcal{N}(0,1)$.

$$F(x) = \int_{-\infty}^{x} \frac{x}{\sqrt{2\pi}} \cdot e^{\frac{1}{2}t^2} dt \tag{4.0.6}$$

donde:

$$F^{-1}(p\left(x\right)) = \alpha + \beta x \tag{4.0.7}$$

La función *probit* se acerca más rápido a probabilidades de 0 y 1 que la función *logit*.

- **Transformación cloglog**

$$\ln(-\ln(1-p(x))) \tag{4.0.8}$$

donde:

$$p(x) = F(\alpha+\beta x) = 1 - \exp[-\exp(\alpha+\beta x)] \tag{4.0.9}$$

desarrollando en términos de $F^{-1}(p(x))$, se tendría que:

$$p(x) = 1 - \exp[-\exp(\alpha+\beta x)]$$

$$\exp[-\exp(\alpha+\beta x)] = 1 - p(x)$$

$$-\exp(\alpha+\beta x) = \ln(1-p(x))$$

$$\alpha + \beta x = \ln(-\ln(1 - p(x)))$$

por lo tanto:

$$F^{-1}(p(x)) = \ln(-\ln(1 - p(x)))$$

- **Transformación LogLog**

$$\ln(-\ln(p(x))) \tag{4.0.10}$$

$$p(x) = F(\alpha + \beta x) = \exp[-\exp(\alpha + \beta x)] \tag{4.0.11}$$

en términos de $F^{-1}(p(x))$, se tendría que:

$$p(x) = \exp[-\exp(\alpha + \beta x)]$$

$$\ln(p(x)) = -\exp(\alpha + \beta x)$$

$$\exp(\alpha + \beta x) = -\ln(p(x))$$

$$\alpha + \beta x = \ln(-\ln(p(x)))$$

por lo tanto:

$$F^{-1}(p(x)) = \ln(-\ln(p(x)))$$

Hasta ahora hemos considerado que sólo tenemos una variable explicativa, sin embargo en la mayoría de los fenómenos que se estudian mediante los modelos lineales generalizados se tienen k variables donde la esperanza de la variable de respuesta Y es:

$$g[E(y_i)] = g(\mu_i) = \alpha + \beta x$$

El modelo lineal generalizado para la esperanza condicional es de la forma[1]:

$$g[\mu(x)] = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

dónde $\mu(x) = E[Y \mid X_1 = x_1, ..., X_k = x_k]$ y g es una función suave, invertible y con segunda derivada, conocida como función liga.

En el caso de las variables de respuesta binaria, su distribución condicional a la combinación de los valores de las variables explicativas sigue una distribución de probabilidades Bernoulli.

4.1. Modelo de regresión logística simple

Derivado de que el presente trabajo se enfoca en la condición de pobreza la cual es una variable dicotómica, se utilizará un modelo de regresión logística, ya que es el recurso más eficiente para representar la probabilidad de ocurrencia de un evento binario (éxito/fracaso) en función de una serie de variables predictoras e independientes y además es una técnica analítica que forma parte de los llamados modelos lineales generalizados[2].

Entonces, si se tiene una variable dependiente Y, que toma valores $Y = 1$ generalmente para presencia de una característica u otra categoría de la variable y $Y = 0$ para la ausencia de la característica o la otra categoría de la variable, y considerando una variable explicativa, entonces se hace referencia a un modelo de regresión logística simple cuya formulación se plantea de la siguiente manera: en función de la transformación logit,

$$\ln\left[\frac{E\left[Y=1\right]}{1-E\left[Y=1\right]}\right] = \alpha + \beta x$$

En términos de la esperanza de Y

$$E\left[Y=1\right] = \frac{exp(\alpha+\beta x)}{1+exp(\alpha+\beta x)} = \frac{1}{1+exp[-(\alpha+\beta x)]} \tag{4.1.1}$$

4.1.1. Interpretación de los parámetros

1. El signo de β indica el sentido del cambio en la probabilidad a los cambios en X.

[1] H.M and DOUGLAS C. MONTGOMERY, (2012) Generalized Linear Models pp. 451
[2] H.M and DOUGLAS C. MONTGOMERY, (2012) Generalized Linear Models pp. 422

2. Si $\beta = 0$ entonces $E[Y]$ no depende de X y por lo tanto es correcto afirmar que la variable Y es independiente de X.

3. α es el valor del logaritmo de la ventaja de respuesta $Y = 1$ frente a $Y = 0$.

4. La razón entre la probabilidad de éxito y la probabilidad de fracaso se trata de un número que expresa cuánto más probable es obtener la ocurrencia de un evento frente a que no se produzca el evento de estudio, lo anterior se define como el momio asociado a un evento, y se expresa de la siguiente manera:
$$\frac{E[Y=1]}{1-E[Y=1]} = exp[\alpha + \beta x] = e^{\alpha} \cdot e^{\beta x} \tag{4.1.2}$$

5. El cociente de momios para dos valores distintos de X es:
$$\theta(x_1, x_2) = \frac{exp(\alpha + \beta x_1)}{exp(\alpha + \beta x_2)} = e^{\beta(x_1 - x_2)} \tag{4.1.3}$$

donde la exponencial representa el incremento en momios, el cual se interpreta como el incremento de la probabilidad de éxito asociada a una unidad de cambio en la variable explicativa.

4.2. Modelo de regresión logística múltiple

Ahora, si se tiene una variable dependiente Y, que toma valores $Y = 1$ generalmente para presencia de una característica u otra categoría de la variable y $Y = 0$ para la ausencia de la característica o la otra categoría de la variable, considerando n variables explicativas y bajo el supuesto de que las respuestas son independientes, se tiene que para cada combinación de dichas variables, sigue una distribución Binomial, es decir:

$$Y = \mathbf{X}_i'\beta \sim B(1,\, p(x_1, x_2, ..., x_n)) \tag{4.2.1}$$

donde $\mathbf{X} = (X_1, X_2, ..., X_n)$ es el vector de variables explicativas, $\beta' = (\beta_0, ..., \beta_n)$ es el vector de parámetros desconocidos a estimar.

De esta manera, la esperanza condicional se expresaría de la manera siguiente:

$$E[Y \mid X_1 = x_1, ..., X_n = x_n]. \tag{4.2.2}$$

El modelo de regresión logística para Y en términos de los valores de las n variables $\mathbf{X} = (X_1, X_2, ..., X_n)$, se puede modelar como:

$$E\left[Y \mid \mathbf{X}\right] = \frac{exp(\alpha + \sum_{i=1}^{n} \beta_i X_i)}{1 + exp(\alpha + \sum_{i=1}^{n} \beta_i X_i)}$$

Si $\alpha = \beta_0$ y $X_0 = 1$, entonces:

$$E\left[Y \mid \mathbf{X}\right] = \frac{exp(\sum_{i=0}^{n} \beta_i \mathbf{X}_i)}{1 + exp(\alpha + \beta \mathbf{X})}$$

En términos matriciales:

$$E\left[Y \mid \mathbf{X}\right] = \frac{exp(\boldsymbol{\beta}^t \mathbf{X})}{1 + exp(\boldsymbol{\beta}^t \mathbf{X})} \tag{4.2.3}$$

donde $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_n)'$.

En función de la transformación logit:

$$logit\left[E\left[Y = 1\right]\right] = \ln\left[\frac{E(\mathbf{Y} = 1)}{1 - E(\mathbf{Y} = 1)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n = \sum_{i=0}^{n} \beta_i x_i$$

$$E(Y = 1) = \frac{exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)}{1 + exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n)}$$

4.2.1. Interpretación de los coeficientes

1. Si se tiene que $\beta_i = 0$ para $i = 1, 2, ...n$, se concluye que la variable de estudio Y es independiente de las variables explicativas ya que:

$$P(Y = 1) = \frac{exp(\beta_0)}{1 + exp(\beta_0)}. \tag{4.2.4}$$

2. En término de momios:

$$O(E) = \frac{P(Y = 1)}{1 - P(Y = 1)} \tag{4.2.5}$$

3. Si todos los coeficientes de las carencias permanecen igual para dos individuos con excepción de una, el coeficiente representará el cambio en el cociente de momios es decir, será un incremento en el riesgo de pasar de no tener la carencia a tenerla manteniendo todas las demás constantes.

4. Ahora bien, la razón entre dos sujetos, los cuales tienen configuraciones de variables explicativas como $x_1 = (1, x_{11}, ..., x_{1n})$ y $x_2 = (1, x_{21}, ..., x_{2n})$, expresa el riesgo relativo de que ocurra el evento de estudio cuando se tienen las condiciones del sujeto x_1 respecto de cuando se está en las condiciones del sujeto x_2, de manera analítica se tendría que:

$$\theta(x_1,x_2) = \frac{exp(\sum_{i=0}^{n} \beta_i x_{1i})}{exp(\sum_{i=0}^{n} \beta_i x_{2i})} = exp(\sum_{i=0}^{n} \beta_i (x_{1i} - x_{2i})) \tag{4.2.6}$$

Si el valor de $\theta(x_1, x_2) = 1$, entonces es posible concluir que no guardan relación, por otro lado si $\theta(x_1, x_2) > 1$, indica que el sujeto x_1 tiene mayor posibilidad de éxito que x_2.

4.3. Bondad de ajuste

En regresión logística existen varias medidas de bondad de ajuste para comparar la diferencia entre valores pronosticados y valores observados, sin embargo la medida más utilizada se base en el cálculo de la devianza.

Previamente, se considera a la función de verosimilitud de una regresión logística dada por:

$$L(\beta_0, \beta_1) = \prod_{i=1}^{k} \left(\begin{array}{c} n_i \\ y_i \end{array} \right) \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \tag{4.3.1}$$

y su función de log-verosimilitud

$$\log L(\beta_0, \beta_1) = \sum_{i=1}^{k} \left\{ \log \left(\begin{array}{c} n_i \\ y_i \end{array} \right) + y_i \log(\pi_i) + (n_i - y_i) \log(1 - \pi_i) \right\} \tag{4.3.2}$$

Sea $L_c = L\left(\hat{\beta}_0, \hat{\beta}_1\right)$, los estimadores máxima verosimilitud de los parámetros, es decir el modelo completo entonces, bajo el modelo ajustado la función de verosimilitud resulta:

$$L_c = \sum_{i=1}^{k} \left\{ \log \left(\begin{array}{c} n_i \\ y_i \end{array} \right) + y_i \log(\hat{\pi}_i) + (n_i - y_i) \log(1 - \hat{\pi}_i) \right\} \tag{4.3.3}$$

donde $\hat{\pi}_i = (\hat{y}_i/n_i)$, la probabilidad estimada para $Y_i = 1$.

Por otro lado, se define al Modelo Saturado L_f, como aquel modelo que se ajusta perfectamente a los datos, es decir, las frecuencias de respuesta $Y = 1$ estimadas por el modelo coinciden con las observadas, y tiene tantos parámetros desconocidos como variables explicativas, donde su función de verosimilitud es:

$$L_f = \sum_{i=1}^{k} \left\{ \log \binom{n_i}{y_i} + y_i \log(\tilde{\pi}_i) + (n_i - y_i) \log(1 - \tilde{\pi}_i) \right\} \tag{4.3.4}$$

donde $\tilde{\pi}_i = (y_i/n_i)$, la proporción observada para $Y_i = 1$.

Con lo anterior, la devianza se define como 2 veces la razón de las log-verosimilitudes, es decir:

$$D = 2 \log \frac{L_f}{L_c} = 2 \left(\log(L_f) - \log(L_c) \right) \tag{4.3.5}$$

$$\iff \begin{array}{l} 2 \left[\sum_{i=1}^{k} \left\{ \log \binom{n_i}{y_i} + y_i \log(\tilde{\pi}_i) + (n_i - y_i) \log(1 - \tilde{\pi}_i) \right\} \right. \\ \left. - \sum_{i=1}^{k} \left\{ \log \binom{n_i}{y_i} + y_i \log(\hat{\pi}_i) + (n_i - y_i) \log(1 - \hat{\pi}_i) \right\} \right] \end{array}$$

$$\iff 2 \sum_{i=1}^{k} \{ y_i \log(\tilde{\pi}_i) - y_i \log(\hat{\pi}_i) + (n_i - y_i) \log(1 - \tilde{\pi}_i) - (n_i - y_i) \log(1 - \hat{\pi}_i) \}$$

$$\iff 2 \sum_{i=1}^{k} \left\{ y_i \log \left(\frac{\tilde{\pi}_i}{\hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{(1 - \tilde{\pi}_i)}{(1 - \hat{\pi}_i)} \right) \right\} \tag{4.3.6}$$

Cuando los datos se ajustan adecuadamente a un modelo de regresión logística, la devianza se distribuye aproximadamente a una χ^2 con $n-p$ grados de libertad, donde p es el número de parámetros en el modelo, se sabe que a pequeños valores de la devianza implica que los datos se ajustan bien al modelo, en caso contrario, muestra que el modelo no es adecuado. No obstante, en el caso de datos desagregados, es decir en el que contamos con la respuesta individual esta distribución asintótica no se cumple.

Ahora bien, como una primera idea de bondad de ajuste se puede utilizar el hecho de que los grados de libertad corresponden a la esperanza de la estadística

de bondad de ajuste y que dos veces los grados de libertad corresponden a la varianza. Existen otras manera como utilizar la estadística de Hosmer, sin embargo, la diferencia de ellas no necesariamente sería interpretable.

Una referencia para confirmar si los datos están bien ajustados es dividir la devianza entre los grados de libertad

$$\frac{Deviance}{grados\ de\ libertad} \approx 1$$

o bien, si se obtiene la diferencia de devianzas de los modelos, es decir:

$$Dev_1 - Dev_2 \approx gl_1 - gl_2 \tag{4.3.7}$$

Donde Dev_1 es la devianza del modelo sin variables independientes, solo con la constante (*null deviance*), y Dev_2 es la devianza del modelo con variable predictora (*residual deviance*). Ahora bien, si la diferencia de devianzas es pequeña, indica que el modelo sin variables independientes es tan bueno como el modelo con variable predictora, por lo tanto el parámetro del modelo con variable predictora es igual a cero, es decir, la variable predictora no es importante para la construcción del modelo.

4.4. Prueba de hipótesis para coeficientes de variables individuales

En esta sección se expondrá la teoría de inferencia en base a la prueba de hipótesis sobre los coeficientes del modelo individual tal como:

$$\begin{aligned} H_0 &: \beta_j = 0 \\ H_1 &: \beta_j \neq 0 \end{aligned} \tag{4.4.1}$$

Para muestras grandes, la distribución de un estimador de máximo verosimilitud es aproximadamente una normal con poco o ningún sesgo, por lo tanto las varianzas y covarianzas de un conjunto de estimadores máximo verosimilitud pueden encontrarse con la segunda derivada parcial de la función de la log verosimilitud con respecto a los parámetros, lo anterior, se conoce como estadístico de Wald.

Entonces, sea $\mathbf{G}$ la matriz de tamaño $p \times p$ la segunda derivada parcial de la función de log-verosimilitud, esto es:

$$G_{ij} = \frac{\partial \mathcal{L}(\beta)}{\partial \mathcal{L}(\beta_{\mathrm{i}})\, \partial \mathcal{L}(\beta_{\mathrm{j}})}\ donde\ i, j = 0, 1, ..., k \tag{4.4.2}$$

donde **G** es llamada la matriz Hessiana. Si los elementos de esta son evaluados en el estimador máximo verosímil, $\beta = \hat{\beta}$, donde la aproximación de la matriz covarianzas de los coeficientes de la regresión logística es:

$$Var\left(\hat{\beta}\right) = -\mathbf{G}\left(\hat{\beta}\right)^{-1} = \left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)^{-1}$$

donde **V** es una matriz diagonal de tamaño $n \times n$ la estimación de la varianza de cada observación, es decir, el $i - ésimo$ elemento de la diagonal de **V** es:

$$V_{ii} = n_i\hat{\pi}_i\left(1 - \hat{\pi}_i\right)$$

Esto es la matriz de covarianza de $\hat{\beta}$,donde la raíz cuadrada de los elementos de la diagonal son los errores estándar de los coeficientes, entonces, el estadístico para la prueba de hipótesis es

$$z_0 = \frac{\hat{\beta}_j}{SE\hat{\beta}_j} \sim N\left(0, 1\right) \tag{4.4.3}$$

Capítulo 5

Regresión Logística con los datos sobre Pobreza

El ejercicio se realizó con la variable denominada «pobreza» como la variable dependiente la cual toma el valor de 1 si la persona se encuentra en condición de pobreza y 0 si no se encuentra en esta condición.

De manera preliminar, se presenta el ajuste de modelos de regresión logística simple utilizando cada una de las carencias como predictor, más adelante se presenta el ajuste del modelo de regresión logística múltiple considerando todas las carencias como variables predictivas.

- Resultado del ajuste de un modelo de regresión logística con la pobreza como variable respuesta y la carencia de rezago educativo como variable predictor:

Cuadro 5.1: Resultados de una regresión logística simple con el rezago educativo como predictor

Salida

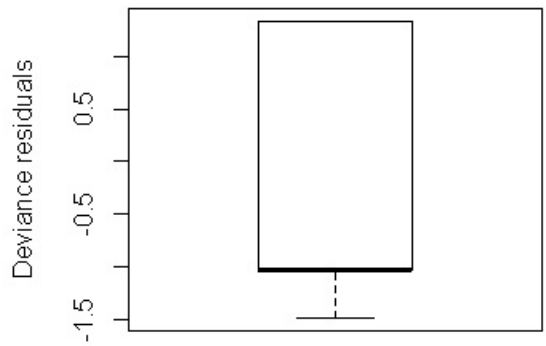
```
Call:
glm(formula = pobreza ~ ic_rezedu, family = binomial, data = pobreza)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.492  -1.026  -1.026   1.337   1.337

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.368347   0.004851  -75.93   <2e-16 ***
ic_rezedu    1.084105   0.011650   93.06   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 289003  on 216207  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 289007

Number of Fisher Scoring iterations: 4
```

El modelo se representa de la siguiente forma:

$$\ln\left[\frac{E\left[Y\right]}{1-E\left[Y\right]}\right] = -0.368 + 1.084\, ic\_rezedu$$

En la línea donde nos indican los *Deviance Residuals*: se obtienen cinco estadísticos sobre la distribución de los residuos del modelo: el valor mínimo, 1er, 2do y 3er cuartil así como el valor máximo. Ahora bien, como la mediana es -1.026, significa que los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura 5.0.1.

Figura 5.0.1: Diagrama de caja rezago educativo



Por lo tanto como la mediana no es cercana a cero, el modelo no es un buen ajuste para los datos ya que las probabilidades no son simétricas alrededor del 0.5, esto indica que quizá otra función liga sería más adecuada para un mejor ajuste.

El resultado de *Coefficients*: muestra los coeficientes estimados por el modelo, es decir, los parámetros $\beta_0 \, y \, \beta_1$, donde β_0 es la ordenada al origen (*Intercept*) en la escala de la transformación *logit*, en un modelo lineal la ordenada al origen señala el *logit* del valor de la variable independiente denominada Y cuando las variables independientes son igual 0, es decir cuando una persona no cuenta con ninguna carencia y en el caso de β_1, se interpreta como el cambio en el logit de la variable resultado, es decir, el cambio de que la persona este en situación de pobreza asociado al cambio de categoría en la variable predictora, como consecuencia la probabilidad de que una persona este en situación de pobreza dado que tiene rezago educativo es $e^{(1.084)} = 2.95$ veces superior en comparación a una persona que no tiene rezago educativo.

El *error estándar* de cada coeficiente es la raíz cuadrada de la varianza estimada del estimador entre el número de observaciones, el valor de z corresponde al valor del coeficiente entre su error estándar asumiendo que su comportamiento es conforme a una distribución normal con media cero y varianza 1 es decir, es el estadístico de Wald, en este caso como la probabilidad de este valor es muy pequeña y el valor absoluto del estadístico es mayor que el punto crítico $z_{\alpha/2} = 1.96$, se rechaza la hipótesis nula, lo que significa que β_2 es significativamente diferente de cero, es decir, la variable explicativa si hace una contribución significativa a la predicción del resultado, con esto finalmente se concluye que ésta carencia si está relacionada con la situación de pobreza.

En el estudio observacional, es común que el *p-value* sea mayor al comúnmente usado (0.05), por ello de manera alternativa se utiliza la diferencia entre las devianzas, en este caso será la diferencia entre el valor de la devianza para el modelo que solo incluye la constante *(null deviance)* y la devianza para el modelo donde solo incluye la constante *(residual deviance*), se tiene que:

$$DN - DR = 298,206 - 289,003 = 9,203$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Finalmente, como el valor de la diferencia de las devianzas es grande, se puede concluir que la variable de rezago educativo mejora la predicción de la condición de pobreza.

A continuación se presentan los resultados de las carencias restantes:

- Resultado del ajuste de un modelo de regresión logística con la pobreza como variable respuesta y la carencia de salud como variable predictor:

Cuadro 5.2: Resultados de una regresión logística simple con servicios de salud como predictor

Salida

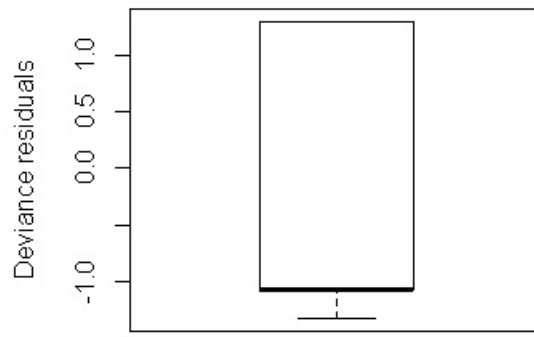
```
Call:
glm(formula = pobreza ~ ic_asalud, family = binomial, data = pobreza)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.327  -1.064  -1.064   1.296   1.296

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.273859   0.004765  -57.47   <2e-16 ***
ic_asalud    0.619616   0.011614   53.35   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 295314  on 216207  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 295318

Number of Fisher Scoring iterations: 4
```

La representación del modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E\left[Y\right]}{1-E\left[Y\right]}\right] = -0.273 + 0.619\, ic\_asalud$$

Los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura 5.0.2.

Figura 5.0.2: Diagrama de caja - servicio de salud



El parámetro asociados a la carencia de salud es significativamente distinto de 0, ya que el valor del estadístico de Wald es de 53.35 el cual es mayor al punto crítico $z_{\alpha/2} = 1.96$, entonces se rechaza la prueba de hipótesis nula donde $\beta_1 = 0$.

En este caso podemos decir que la probabilidad de que una persona este en situación de pobreza dado que es carente de servicios de salud es $e^{(0.619)} = 1.85$ veces superior en comparación a una persona que no es carente de servicios de salud.

Calculando la diferencia de devianzas *(residual deviance y null deviance)*, se tiene que:

$$DN - DR = 298,206 - 295,314 = 2,892$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Como resultado se tiene que la carencia por acceso a los servicios de salud, es significativa para la construcción del modelo ya que el valor de la diferencia de las devianzas es grande.

- Resultado del ajuste de un modelo de regresión logística con la pobreza

como variable respuesta y la carencia de seguridad social como variable predictor:

Figura 5.0.3: Resultados de una regresión logística simple con la seguridad social como predictor

Salida

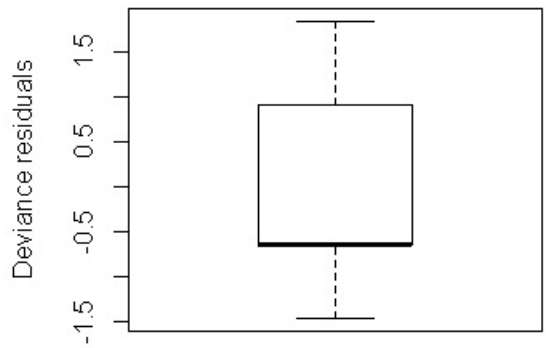
```
Call:
glm(formula = pobreza ~ ic_segsoc, family = binomial, data = pobreza)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4591  -0.6338  -0.6338   0.9197   1.8460

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.503046   0.008637  -174.0   <2e-16 ***
ic_segsoc    2.144659   0.010473   204.8   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 247984  on 216207  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 247988

Number of Fisher Scoring iterations: 4
```

El modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E[Y]}{1-E[Y]}\right] = -1.503 + 2.144\, ic\_segsoc$$

Los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura 5.0.4.

Figura 5.0.4: Diagrama de caja - seguridad social



Calculando la diferencia de devianzas *(residual deviance y null deviance)*, se tiene que:

$$DN - DR = 298,206 - 247,984 = 50,222$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Como resultado se tiene que la carencia por seguridad social, es significativa para la construcción del modelo.

- Resultado del ajuste de un modelo de regresión logística con la pobreza como variable respuesta y la carencia de calidad y espacios de la vivienda como variable predictor:

Figura 5.0.5: Resultados de una regresión logística simple con calidad y espacios de la vivienda como predictor

Salida

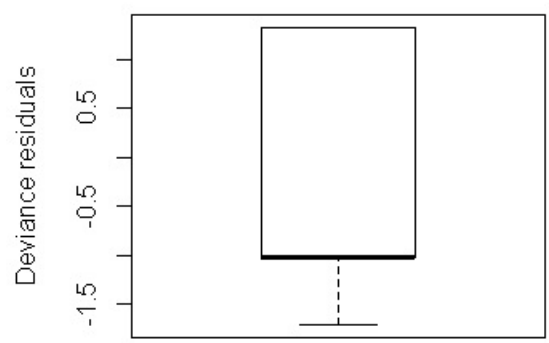
```
Call:
glm(formula = pobreza ~ ic_cv, family = binomial, data = pobreza)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.716  -1.032  -1.032   1.330   1.330

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.352560   0.004674  -75.43   <2e-16 ***
ic_cv        1.563477   0.015112  103.46   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 285495  on 216207  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 285499

Number of Fisher Scoring iterations: 4
```

El modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E\left[Y\right]}{1-E\left[Y\right]}\right] = -0.352 + 1.563\, ic\_cv$$

Los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura 5.0.6.

Figura 5.0.6: Diagrama de caja - calidad y espacios de la vivienda



Calculando la diferencia de devianzas *(residual deviance y null deviance)*, se tiene que:

$$DN - DR = 298,206 - 285,495 = 12,711$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Como resultado se tiene que la carencia por calidad y espacios de la vivienda, es significativa para la construcción del modelo.

- Resultado del ajuste de un modelo de regresión logística con la pobreza como variable respuesta y la carencia de servicios básicos de la vivienda como variable predictor:

Cuadro 5.3: Resultados de una regresión logística simple con los servicios básicos de la vivienda como predictor

Salida

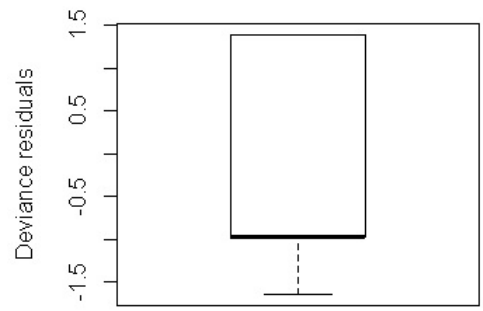
```
Call:
glm(formula = pobreza ~ ic_sbv, family = binomial, data = pobreza)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6480  -0.9716  -0.9716   1.3983   1.3983

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.505624   0.005037  -100.4   <2e-16 ***
ic_sbv       1.566202   0.011568   135.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 277451  on 216207  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 277455

Number of Fisher Scoring iterations: 4
```

El modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E\left[Y\right]}{1-E\left[Y\right]}\right] = -0.505 + 1.566\, ic\_sbv$$

Los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura 5.0.7.

Figura 5.0.7: Diagrama de caja - servicios básicos de la vivienda



Calculando la diferencia de devianzas *(residual deviance y null deviance)*, se tiene que:

$$DN - DR = 298,206 - 277,451 = 20,755$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Como resultado se tiene que la carencia por servicios básicos de la vivienda, es significativa para la construcción del modelo.

- Resultado del ajuste de un modelo de regresión logística con la pobreza como variable respuesta y la carencia de alimentación como variable predictor:

Cuadro 5.4: Resultados de una regresión logística simple con alimentación como predictor

Salida

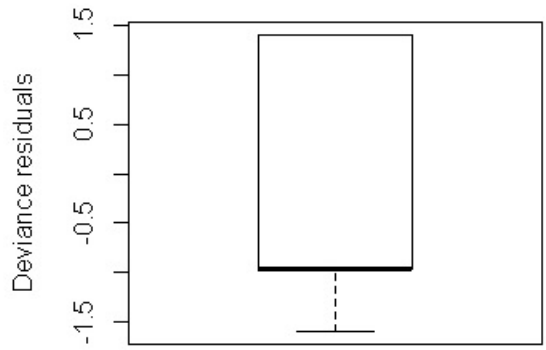
```
Call:
glm(formula = pobreza ~ ic_ali, family = binomial, data = pobreza)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.594  -0.967  -0.967   1.404   1.404

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.517376   0.005116  -101.1   <2e-16 ***
ic_ali       1.457414   0.010942   133.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 278642  on 216207  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 278646

Number of Fisher Scoring iterations: 4
```

El modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E\left[Y\right]}{1-E\left[Y\right]}\right] = -0.517 + 1.457\, ic\_ali$$

Los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura 5.0.8.

Figura 5.0.8: Diagrama de caja - alimentación



Calculando la diferencia de devianzas *(residual deviance y null deviance)*, se tiene que:

$$DN - DR = 298,206 - 278,642 = 19,564$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Como resultado se tiene que la carencia por acceso a la alimentación es significativa para la construcción del modelo.

- Resultado del ajuste de un modelo de regresión logística con la pobreza como variable respuesta y la carencia de ingresos como variable predictor:

Figura 5.0.9: Resultados de una regresión logística simple con el ingreso como predictor

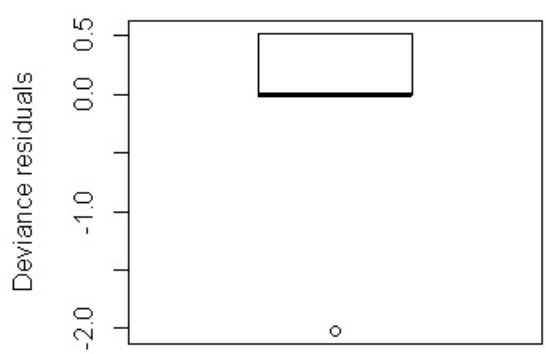
```
Call:
glm(formula = pobreza ~ plb, family = binomial, data = pobreza)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.02462  -0.00005  -0.00005   0.52512   0.52512

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -20.57      55.37  -0.371    0.710
plb            22.48      55.37   0.406    0.685

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance:  87324  on 216207  degrees of freedom
AIC: 87328
```

El modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E\left[Y\right]}{1-E\left[Y\right]}\right] = -20.57 + 22.48\, plb$$

Los residuos no se distribuyen de manera simétrica alrededor del cero como se muestra en la Figura. 5.0.10

Figura 5.0.10: Diagrama de caja - ingresos



Calculando la diferencia de devianzas *(residual deviance y null deviance)*, se tiene que:

$$DN - DR = 298,206 - 87,324 = 210,882$$

con grados de libertad igual a:

$$DFN - DFR = 216,208 - 216,207 = 1$$

Como resultado se tiene que la variable del ingreso por debajo de la línea de bienestar es significativa para la construcción del modelo, sin embargo como el estadístico z es menor al punto crítico y además su p-value es de 0.68, se concluye que éste coeficiente es igual a cero, por lo tanto el resultado de la regresión logística con la variable de ingreso como variable predictora no es un buen ajuste para los datos.

Luego en el Cuadro 5.5, se encuentra el resumen con las devianzas obtenidas en cada uno de los modelos de regresión logística simple.

Cuadro 5.5: Resumen de devianzas

Variable predictora	**Null deviance**	**Residual deviance**	**Diferencia**
Alimentación	298,206	278,642	19,564
Servicios de salud	298,206	295,314	2,892
Calidad y espacios	298,206	285,495	12,711
Rezago educativo	298,206	289,003	9,203
Servicios básicos	298,206	277,451	20,755
Seguridad Social	298,206	247,984	50,222
Ingreso	298,206	87,324	210,882

Debido a que la diferencia de las devianzas son muy grandes, esto implicó que de manera individual todas las variables fueran importantes para la construcción de cada modelo.

Ahora bien, se presenta el resultado de la aplicación de la regresión logística con las carencias (rezago educativo, servicios de salud, seguridad social, calidad y espacios de la vivienda, servicios básicos en la vivienda, y alimentación) como variables independientes, no se omite mencionar que al incluir la linea de bienestar no se lograba convergencia por tal motivo no se incluyó en este ejercicio.

Especificando que al igual que la variable pobreza, las variables de las carencias toman el valor de 1 para aquellos individuos que la presenten y 0 para los que no presenten la carencia:

Cuadro 5.6: Resultados de regresión logística múltiple

Salida

```
Call:
glm(formula = pobreza ~ ic_rezedu + ic_asalud + ic_segsoc + ic_cv +
    +ic_sbv + ic_ali, family = binomial, data = pobreza)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7323  -0.7478  -0.4717   0.8695   2.2130

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.13991    0.01059 -202.09   <2e-16 ***
ic_rezedu    1.00850    0.01427   70.67   <2e-16 ***
ic_asalud   -0.21850    0.01376  -15.88   <2e-16 ***
ic_segsoc    2.07306    0.01212  171.00   <2e-16 ***
ic_cv        0.84466    0.01796   47.03   <2e-16 ***
ic_sbv       0.69253    0.01368   50.62   <2e-16 ***
ic_ali       1.22965    0.01286   95.59   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 298206  on 216208  degrees of freedom
Residual deviance: 219132  on 216202  degrees of freedom
  (41 observations deleted due to missingness)
AIC: 219146

Number of Fisher Scoring iterations: 4
```

Bajo el resultado obtenido, el modelo quedaría de la siguiente forma:

$$\ln\left[\frac{E[Y])}{1-E[Y]}\right] = -2.13 + 1.008\,\textit{rezago educativo} - 0.218\,\textit{salud} \\ +2.073\,\textit{seguridad social} + 0.844\,\textit{calidad y espacios} \\ +0.692\,\textit{servicios básicos} + 1.229\,\textit{alimentación}$$

Dado que el *p-value* de la prueba de hipótesis del estadístico z también indica el valor de la estadística de prueba, entonces se rechaza la hipótesis nula de que el valor de los coeficientes de regresión son cero, como consecuencia se tiene que todas las variables de las carencias sociales utilizadas son significativas para la construcción del modelo.

Ahora bien, las estimaciones de los componentes se resume en el Cuadro 5.7:

Cuadro 5.7: Comparación de los exponenciales de los coeficientes

	exponencial de los coeficientes obtenidos en el modelo conjunto	exponencial de los coeficientes obtenidos en el modelo por separado
Rezago educativo	2.7414	2.9567
Servicios de salud	0.8037	1.8581
Seguridad Social	7.9486	8.5386
Calidad y espacios	2.3270	4.7750
Servicios básicos	1.9987	4.7884
Alimentación	3.4198	4.2947

En el caso de la aplicación del modelo individual con cada una de las carencias, los coeficientes representan un incremento en el coeficiente de momios de ser pobres al cambiar de no tener la carencia a tenerla. Con lo anterior, y con el fin de ejemplificar el resultado, notamos que un individuo que tiene rezago educativo tiene un incremento en el coeficiente de momios de 195 % de ser pobre.

En el modelo múltiple la interpretación es similar, dado que para dos individuos con las mismas carencias, cada coeficiente representa el cambio por carencia, lo que se puede ver es que en este modelo todas las variables, a excepción de la de servicios de salud bajan su importancia, esto es lógico por los efectos marginales, sin embargo, las dos carencias que se mantienen mas fuertes son las de seguridad social y rezago educativo. Para la variable de servicios de salud lo que se puede concluir es que pasa de ser un factor de riego a un factor protector, debido a que su coeficiente incrementó en el modelo múltiple, sin embargo esto puede ser debido a la correlación que tenga con las otras variables.

Posteriormente, con la finalidad de poder comparar el resultado del modelo contra la clasificación que se tiene para cada individuo, se introdujo una variable dicotómica la cual toma el valor de 1 si $p(x) \geq 0.5$ y toma el valor de 0 cuando $p(x) < 0.5$ donde el 1 resulta ser una persona que probablemente tenga la condición de pobreza y 0 en el caso contrario.

Una vez especificando lo anterior, se obtiene el siguiente cuadro:

Cuadro 5.8: Porcentajes de especificidad y sensibilidad del modelo.

		Pobreza CONEVAL	
		no pobre	pobre
Pobreza modelo	no pobre	97,433	58,293
		(83 %)	(37 %)
	pobre	19,739	99,037
		(17 %)	(63 %)
	Total	117,172	157,330
		(100 %)	(100 %)

Calculando los porcentajes por columna, se tiene que para la coincidencia de no pobres es de un 83 %, mientras que la coincidencia para los pobres es de un 63 %.

Conclusiones

Como resultado de la construcción del Análisis de Componentes Principales a partir de la matriz de covarianzas, se encontró que en cuanto a la proporción de varianza obtenida entre la primera (32.2 %) y la segunda componente (16.5 %) se tiene el mayor descenso de varianza con un 15 %, mientras que las demás van disminuyendo de manera casi constante sin embargo, es importante que a pesar del descenso mencionado ninguna de las varianzas de las componentes supera el valor a 1.

Posteriormente se generó una representación gráfica para la interpretación de los resultados obtenidos con las primeras tres componentes las cuales en conjunto explican un 62.3 % del total de la información y cuyas variables más influyentes por componente se enuncian a continuación:

- 1° Componente: acceso a la seguridad social e ingresos por debajo de la línea de bienestar.
- 2° Componente: acceso a la seguridad social y acceso a los servicios de salud.
- 3° Componente: ingresos por debajo de la línea de bienestar y acceso a servicios básicos de la vivienda.

Por otro lado, y con el fin de hacer el ejercicio estandarizando los datos se hizo el análisis de componentes principales con la matriz de correlación policórica, donde el descenso importante en la proporción de varianza también se da entre la primera componente (41.1 %) y la segunda componente (19.8) con un 21.3 % de diferencia, además en este caso las primeras dos componentes tienen una varianza mayor a uno.

De la misma manera que el ejercicio con la matriz de covarianzas, se generó una gráfica de los resultados obtenidos donde las variables que más influyen en la construcción de los primeros tres componentes son:

- 1° Componente: seguridad social, acceso a servicios básicos de la vivienda e ingresos por debajo de la línea de bienestar.

- 2° Componente: acceso a servicios de la salud y la seguridad social.
- 3° Componente: rezago educativo y acceso a la alimentación.

Estas tres componentes explican un 73.1 % del total de la información.

En general, el análisis de componentes principales permitió visualizar el comportamiento de la población respecto a las carencias sociales y situación de pobreza en dos y tres dimensiones, siendo el ejercicio con la matriz de covarianzas la que con tres dimensiones obtuvo más pérdida de información ya que explica el 62.3 % respecto a un 73.1 % que proporcionó el ejercicio con la matriz de correlación, la diferencia no es significativa dado que las variables con las que se realizó el ejercicio tienen las mismas unidades, (variables dicotómicas con valores de 0 y 1) además de que el ejercicio se realizó solo con el fin de identificar la estructura de los datos.

El resultado en ambos ejercicios con tres componentes como ejes, se identificó que la población se agrupa en dos grupos donde las personas pobres y pobres extremas tienden a tener más las carencias sociales que los no pobres y vulnerables por carencias, lo cual resulta congruente respecto a la formulación de clasificación de la pobreza conforme a la metodología del CONEVAL.

Por otro lado para el análisis de regresión logística se elaboraron dos ejercicios donde la variable a predecir consiste en conocer si el individuo esta en condiciones de pobreza o no; es importante mencionar que debido a que el tamaño de la muestra que se utilizó en el presente trabajo es muy grande, tuvo efecto en la significancia de las pruebas haciéndolas mas significativas, sin embargo el valor puntal no se vio afectado.

El primero de ellos consistió en hacer modelos individuales con cada una de las variables de las carencias sociales como variables independientes, donde se determinó mediante el uso de la devianza que estas variables son significativas para la construcción de cada modelo individual, además se analizaron los coeficientes β_1, los cuales se interpretan como el incremento en el riesgo de ser pobres al tener la carencia a diferencia de no tenerla, mostrando los resultados en el Cuadro 5.9.

Cuadro 5.9: β_1'modelos individuales

Carencias	β_1'**modelos individuales**	**cociente de momios**
rezago educativo	2.9567	195 %
servicios de salud	1.8581	185 %
seguridad social	8.5386	153 %
calidad y espacios de la vivienda	4.7750	177 %
servicios básicos de la vivienda	4.7884	178 %
alimentación	4.2947	129 %

Sin embargo, en todos los casos los residuos no se distribuyen de manera simétrica alrededor del cero por lo que el ajuste por regresión logística simple no es el adecuado, en estos casos una regresión Loglog, parecería ser más adecuada ya que este tipo de modelos se suele utilizar cuando la relación no es lineal.

El segundo ejercicio consistió en elaborar la regresión logística múltiple con todas las variables de las carencias sociales como variables independientes, donde al igual que el modelo individual, el resultado de los coeficientes representa el cambio por carencia. En este ejercicio el valor de sus coeficientes disminuye, es decir baja su importancia debido a las interacciones que se realizan, sin embargo todas se mantienen como variables relevantes para la predicción de pobreza de un individuo.

Además, se realizó una comparación entre el resultado del modelo versus los resultados de la clasificación que realiza el CONEVAL introduciendo una variable dicotómica la cual toma el valor de 1 si $p(x) \geq 0.5$ y 0 en otro caso, donde 1 representa al individuo que probablemente se encuentre en condición de pobreza y 0 en caso contrario, bajo esta premisa de detectó que existe un 83 % de coincidencia para la población no pobre y un 63 % de coincidencia para la población en condiciones de pobreza.

Finalmente, es importante mencionar que una posible mejora para conocer la estructura de la condición de pobreza en la población y por lo tanto poder construir una mejor estimación, podría ser la posibilidad de inferir en la intensidad de las carencias, así como la construcción de factores que prevean la variación de los precios u otros factores que afecten el cálculo de las líneas de bienestar.

Bibliografía

[1] ANDERSON, T.W. «An introduction to multivariate statistical analysis» in Principal Components. (New York: John Wiley & Sons, 2003)

[2] F. DÍAZ, F. CORTÉS, A. ESCOBAR, M. FREYERMUTH, J. RIVERA y G. TERUEL. «Metodología para la medición multidimensional de la pobreza en México». (Ciudad de México: CONEVAL, 2018)

[3] COX, D., and E. SNELL. «The Analysis of Binary Data» in Special Logistic Analyses. (New York: Chapman and Hall/CRC, 1989) .

[4] D. COLLETT. «Modelling Binary Data» in Models for Binary and Binomial Data. (London: Chapman and Hall, 1991)

[5] FRIEDBERG, S.H. «Linear Algebra» in Inner Product Spaces. (New Jersey: Prentice Hall, 2003)

[6] JOAKIM EKSTRÖM. A Generalized Definition of the Polychoric Correlation Coefficient.

[7] RAYMOND H.M and DOUGLAS C. MONTGOMERY: Generalized Linear Models (New Jersey: John Wiley & Sons, 2012).

Apéndices

Apéndice A

Descripción de base de datos

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres
1	sexo	carácter	{1,2}	
2	anac_e	numérico	{1917,2014}	Año de nacimiento
3	inas_esc	numérico	{0,1}	Inasistencia a la escuela
				0. Si asiste
				1. No asiste
4	niv_ed	numérico	{0,2}	Nivel educativo
5	ic_rezedu	numérico	{0,1}	Indicador de rezago educativo
				0. No presenta la carencia
				1. Presenta carencia
6	serv_sal	numérico	{0,7}	Servicios de salud
				0. No cuenta con servicios médicos
				1. Seguro Popular
				2. IMSS
				3. ISSSTE o ISSSTE estatal
				4. Pemex, Defensa o Marina
				5. Otros servicios médicos por seguridad social
				6. Seguro privado de gastos médicos
				7. Otros
7	ic_asalud	numérico	{0,1}	Indicador de carencia por acceso a los servicios de la salud
				0. No presenta la carencia

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres
				1. Presenta carencia
8	tipo_trab1	numérico	{1,2,3}	Tipo de trabajo principal
				1. Subordinado
				2. Independiente que recibe un pago
				3. Independiente que no recibe un pago
9	inclab1	numérico	{0,1}	Incapacidad en caso de enfermedad o maternida
				con goce de sueldo en el trabajo principal
				0. No cuenta con prestación
				1. Sí cuenta con prestación
10	aforlab1	numérico	{0,1}	Prestación de SAR o Afore en el trabajo
				principal
				0. No cuenta con prestación
				1. Sí cuenta con prestación
11	smlab1	numérico	{0,1}	Servicios médicos por prestación laboral en
				ocupación principal
				0. Sin servicios médicos
				1. Con servicios médicos
12	smcv	numérico	{0,1}	Servicios médicos por contratación voluntaria
				0. Sin servicios médicos
				1. Con servicios médicos
13	aforecv	numérico	{0,1}	Afore por contratación voluntaria
				0. Sin afore
				1. Con afore
14	pea	numérico	{0,2}	Población económicamente activa
				0. PNEA
				1. PEA ocupada
				2. PEA desocupada
15	jub	numérico	{0,1}	Población pensionada o jubilada
				0. No pensionado o jubilado
				1. Pensionado o jubilado
16	ss_dir	numérico	{0,1}	Acceso directo a la seguridad social
				0. Sin acceso
				1. Con acceso
				Integrantes que tienen acceso por otros miembro
17	par	numérico	{1,6}	

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres
				1. Jefe/a del hogar
				2. Cónyuge del jefe/a
				3. Hijo del jefe/a
				4. Padre o madre del jefe/a
				5. Suegro del jefe/a
				6. Sin parentesco directo
18	jef_ss	numérico	{0,1}	Acceso directo a la seguridad social de la jefatur
				del hogar
				0. No cuenta
				1. Si cuenta
19	cony_ss	numérco	{0,1}	Acceso directo a la seguridad social de cónyuge
				de la jefatura del hogar
				0. No cuenta
				1. Si cuenta
20	hijo_ss	numérico	{0,1}	Acceso directo a la seguridad social de hijos(as)
				de la jefatura del hogar
				0. No cuenta
				1. Si cuenta
21	s_salud	numérico	{0,1}	Servicios médicos por otros núcleos familiares o
				por contratación propia
				0. No cuenta
				1. Si cuenta
22	pam	numérico	{0,1}	Programa de adultos mayores
23	ic_segsoc	numérico	{0,1}	Indicador de carencia por acceso a la seguridad
				social
				0. No presenta la carencia
				1. Presenta carencia
24	tam_loc	numérico	{1,4}	Tamaño de la localidad
				1. Localidades con 100 000 y más habitantes
				2. Localidades con 15 000 a 99 999 habitantes
				3. Localidades con 2 500 a 14 999 habitantes
				4. Localidades con menos de 2500 habitantes
25	rururb	numérico	{0,1}	Identificador de localidades rurales
				0. Localidades urbanas

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres 1. Localidades rurales
26	Ict	numérico	{0,3697166}	Ingreso corriente total del hogar
27	ictpc	numérico	{0,1858807}	Ingreso corriente total per cápita
28	plb_m	numérico	{0,1}	Población con un ingreso menor a la línea de bienestar mínimo 0. Ingreso superior a la línea de bienestar mínim 1. Ingreso inferior a la línea de bienestar mínimo
29	plb	numérico	{0,1}	Población con un ingreso menor a la línea de bienestar 0. Ingreso superior a la línea de bienestar 1. Ingreso inferior a la línea de bienestar
30	tot_resid	numérico	{1, 21}	Número de residentes de la vivienda
31	num_cuarto	numérico	{1,20}	Número de cuartos en la vivienda
32	icv_mat_pisos	numérico	{0,1}	Índice de carencia del material de piso de la vivienda 0. No presenta carencia 1. Presenta carencia
33	icv_mat_techos	numérico	{0,1}	Índice de carencia del material de techos de la vivienda 0. No presenta carencia 1. Presenta carencia
34	icv_mat_muros	numérico	{0,1}	Índice de carencia del material de muros de la vivienda 0. No presenta carencia 1. Presenta carencia
35	icv_hac	numérico	{0,1}	Índice de carencia por índice de hacinamiento d la vivienda 0. No presenta carencia 1. Presenta carencia
36	ic_cv	numérico	{0,1}	Indicador de carencia por calidad y espacios de la vivienda 0. No presenta la carencia 1. Presenta carencia
				Disponibilidad de agua
37	sb_agua	numérico	{1,7}	

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres
				1. Agua de un pozo, río, lago, arroyo u otro
				2. Agua de pipa
				3. Agua entubada que acarrean de otra vivienda
				4. Agua entubada de la llave pública o hidrante
				5. Captadores de agua de lluvia
				6. Agua entubada fuera de la vivienda, dentro del terreno
				7. Agua entubada dentro de la vivienda
38	sb_dren	numérico	{1,5}	Destino del drenaje
				1. No tiene drenaje
				2. Una tubería que va a dar a un río, lago o mar
				3. Una tubería que va a dar a una barranca o grieta
				4. Una fosa séptica
				5. La red pública
39	sb_luz	numérico	{1,5}	Disponibilidad eléctrica
				1. No tiene luz eléctrica
				2. De panel solar u otra fuente
				3. De una planta particular
				4. Del servicio público
40	isb_combus	numérico	{0,1}	Indicador de carencia por acceso a los servicios de combustible
				0. No presenta carencia
				1. Presenta carencia
41	ic_sbv	numérico	{0,1}	Indicador de carencia de acceso a servicios básicos de la vivienda
				0. No presenta la carencia
				1. Presenta carencia
42	id_men	numérico	{0,1}	Hogares con población de 0 a 17 años
				0. Sin población de 0 a 17 años
				1. Con población de 0 a 17 años
43	ia_1ad	numérico	{0,1}	Algún adulto tuvo una alimentación basada en poca variedad
				0. No

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres 1. Sí
44	ia_2ad	numérico	{0,1}	Algún adulto dejó de desayunar, comer o cenar 0. No 1. Sí
45	ia_3ad	numérico	{0,1}	Algún adulto comió menos de lo que debía come 0. No 1. Sí
46	ia_4ad	numérico	{0,1}	El hogar se quedó sin comida 0. No 1. Sí
47	ia_5ad	numérico	{0,1}	Algún adulto sintió hambre pero no comió 0. No 1. Sí
48	ia_6ad	numérico	{0,1}	Algún adulto solo comió una vez al día o dejó d comer todo un día 0. No 1. Sí
49	ia_7ad	numérico	{0,1}	Alguien de 0 a 17 años tuvo una alimentación basada en muy poca variedad de alimentos 0. No 1. Sí
50	ia_8ad	numérico	{0,1}	Alguien de 0 a 17 años comió menos de lo que debía 0. No 1. Sí
51	ia_9ad	numérico	{0,1}	Se tuvo que disminuir la cantidad servida en las comidas a alguien de 0 a 17 años 0. No 1. Sí
52	ia_10ad	numérico	{0,1}	Alguien de 0 a 17 años sintió hambre pero no comió 0. No 1. Sí
53	ia_11ad	numérico	{0,1}	Alguien de 0 a 17 años se acostó con hambre

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres
				0. No
				1. Sí
54	ia_12ad	numérico	{0,1}	Alguien de 0 a 17 años comió una vez al día o dejó de comer todo un día
				0. No
				1. Sí
55	tot_iaad	numérico	{0,6}	Escala de Inseguridad Alimentaria para hogares sin menores de 18 años
56	tot_iamen	numérico	{0,12}	Escala de Inseguridad Alimentaria para hogares con menores de 18 años
57	ins_ali	numérico	{0,3}	Grado de inseguridad alimentaria 0. Seguridad alimentaria
				1. Inseguridad alimentaria leve
				2. Inseguridad alimentaria moderada
				3. Inseguridad alimentaria severa
58	ic_ali	numérico	{0,1}	Indicador de carencia de acceso a la alimentació
				0. No presenta la carencia
				1. Presenta carencia
59	ent	numérico	{1,32}	Identificador de la entidad federativa
60	i_privacion	numérico	{0,6}	Índice de Privación Social
61	pobreza	numérico	{0,1}	Pobreza
				0. No pobre
				1. Pobre
62	pobreza_e	numérico	{0,1}	Pobreza extrema
				0. No pobre extremo
				1. Pobre extremo
63	pobreza_m	numérico	{0,1}	Pobreza moderada
				0. No pobre moderado
				1. Pobre moderado
64	vul_car	numérico	{0,1}	Población vulnerable por carencias
				0. No vulnerable
				1. Vulnerable
65	vul_ing	numérico	{0,1}	Población vulnerable por ingresos
				0. No vulnerable

	Nombre	Tipo	Longitud	Descripción
73	int_pob	numérico	{0,1}	Intensidad de la privación social: pobres
				1. Vulnerable
66	no_pobv	numérico	{0,1}	Población no pobre y no vulnerable
				0. Pobre o vulnerable
				1. No pobre y no vulnerable
67	carencias	numérico	{0,1}	Población con al menos una carencia
				0. Población sin carencias
				1. Población con carencias
68	carencias3	numérico	{0,1}	Población con tres o más carencias
				0. Población con menos de tres carencias
				1. Población con tres o más carencias
69	cuadrantes	numérico	{1,4}	Cuadrantes de Bienestar y Derechos Sociales
				1. Pobres
				2. Vulnerables por carencias
				3. Vulnerables por ingresos
				4. No pobres y no vulnerables
70	prof_b1	numérico	{0,1}	Índice FGT con alfa igual a 1 (línea de bienestar
71	prof_bm1	numérico	{0,1}	Índice FGT con alfa igual a 1 (línea de bienesta
				mínimo)
72	profun	numérico	{0,1}	Profundidad de la privación social
74	int_pobe	numérico	{0,1}	Intensidad de la privación social: pobres
				extremos
75	int_vulcar	numérico	{0,1}	Intensidad de la privación social: población
				vulnerable por carencias
76	int_caren	numérico	{0,1}	Intensidad de la privación social: población
				carenciada

Apéndice B

Construcción de los indicadores

Indicador de Rezago Educativo

El valor 1 identifica a la población en situación de rezago educativo, mientras que 0 a la población carente de este indicador.

$$\text{ic\_rezedu}_i = \begin{cases} 1 & si\, anac\_e_i \geq 1998\, y\, edad_i \geq 3\, y\, edad_i \leq 21\, y\, inas\_esc_i = 1\, y\, niv\_ed_i < \\ 1 & si\, anac\_e_i \geq 1982\, y\, anac\_e_i \leq 1997\, y\, edad_i \geq 16\, niv\_ed_i < 2 \\ 1 & si\, anac\_e_i \leq 1981\, y\, edad_i \geq 16\, y\, niv\_ed_i = 0 \\ 1 & si\, anac\_e_i \geq 1998\, y\, edad_i \geq 22\, y\, niv\_ed_i < 3 \\ 0 & si\, edad_i \leq 2 \\ 0 & si\, anac\_e_i \geq 1998\, y\, edad_i \geq 3\, y\, edad_i \leq 21\, y\, inas\_esc_i = 0 \\ 0 & si\, edad_i \geq 3\, y\, edad_i \leq 21\, y\, niv\_ed_i = 3 \\ 0 & si\, edad_i \geq 16\, y\, anac\_e_i \geq 1982\, y\, anac\_e_i \leq 1997\, y\, niv\_ed_i \geq 2 \\ 0 & si\, edad_i \geq 16\, y\, anac\_e_i \leq 1981\, y\, niv\_ed_i \geq 1 \end{cases}$$

Donde:

$edad_i = edad\ que\ reporta\ la\ persona\ al\ momento\ de\ la\ entrevista$

$anac\_e_i = año\ de\ medición - edad_i$

$$inas\_esc_i = \begin{cases} 0 & si\ la\ persona\ asiste\ a\ alguna\ institución\ del\ Sistema\ Educativo\ Nacion \\ 1 & si\ la\ persona\ no\ asiste\ a\ alguna\ institución\ del\ Sistema\ Educativo\ Naci \end{cases}$$

$$niv\_ed_i = \begin{cases} 0 & si\ la\ persona\ i\ cuenta\ con\ primaria\ incompleta\ o\ menos \\ 1 & si\ la\ persona\ i\ cuenta\ con\ primaria\ completa\ o\ secundaria\ incompleta \\ 2 & si\ la\ persona\ i\ cuenta\ con\ secundaria\ completa\ o\ medias\ uperior\ incomple \\ 3 & si\ la\ persona\ i\ cuenta\ con\ media\ superior\ completa\ o\ mayor\ nivel\ educati \end{cases}$$

Indicador de carencia por acceso a los servicios de salud

$$\text{ic\_asalud}_i = \begin{cases} 1 & si\ serv\_sal_i = 0 \\ 0 & si\ serv\_sal_i \geq 1 \end{cases}$$

Donde:

$$\text{serv\_sal}_i = \begin{cases} 1 & si\ la\ persona\ i\ cuenta\ con\ Seguro\ Popular \\ 2 & si\ la\ persona\ i\ cuenta\ con\ servicios\ médicos\ del\ IMSS \\ 3 & si\ la\ persona\ i\ cuenta\ con\ servicios\ médicos\ del\ ISSSTE\ o\ ISSSTE\ estata \\ 4 & si\ la\ persona\ i\ cuenta\ con\ servicios\ médicos\ de\ Pemex,\ Defensa\ o\ la\ Mari \\ 5 & si\ la\ persona\ i\ cuenta\ con\ otros\ servicios\ médicos \\ 0 & si\ la\ persona\ i\ no\ cuenta\ con\ ninguno\ de\ los\ servicios\ médicos\ refereidos \end{cases}$$

Indicador de carencia por acceso a la Seguridad Social

$$ic\_ss_i = \begin{cases} 0 & si\ ss\_dir_i = 1 \\ 0 & si\ par_i = 1\ y\ cony\_ss_n = 1 \\ 0 & si\ par_i = 1\ y\ pea_i = 0\ e\ hijo\_ss_{n=1} \\ 0 & si\ par_i = 2\ y\ jefe\_ss_n = 1 \\ 0 & si\ par_i = 2\ y\ pea_i = 0\ e\ hijo\_ss_n = 1 \\ 0 & si\ par_i = 3\ y\ edad_i < 16\ y\ jefe\_ss_n = 1 \\ 0 & si\ par_i = 3\ y\ edad_i < 16\ y\ cony\_ss_n = 1 \\ 0 & si\ par_i = 3\ y\ edad_i \in [16,25]\ y\ inas\_esc_i = 0\ y\ jefe\_ss_n = 1 \\ 0 & si\ par_i = 3\ y\ edad_i \in [16,25]\ y\ inas\_esc_i = 0\ y\ cony\_ss_n = 1 \\ 0 & si\ par_i = 4\ y\ pea_i = 0\ y\ jefe\_ss_{n=1} \\ 0 & si\ par_i = 5\ y\ pea_i = 0\ y\ cony\_ss_{n=1} \\ 0 & si\ s\_salud_i = 1 \\ 0 & si\ pam_i = 1 \\ 1 & en\ otro\ caso \end{cases}$$

Donde

$$ss\_dir_i = \begin{cases} 1 & si\ tipotrab_i = 1\ e\ smlab_i = 1 \\ 1 & si\ tipotrab_i = 2\ y\ (aforlab_i = 1\ o\ aforecv_i = 1)\ y\ (smlab_i = 1\ o\ smcv_i = 1) \\ 1 & si\ tipotrab_i = 3\ y\ aforecv_i = 1\ y\ (smlab_i = 1\ o\ smcv_i = 1) \\ 1 & si\ jub_i = 1 \\ 0 & en\ otro\ caso \end{cases}$$

$$tipo\_trab_i = \begin{cases} 1 & si\ pea_i = 1\ y\ la\ persona\ i\ trabaja\ (con\ o\ sin\ pago)\ para\ una\ unidad\ económ \\ & en\ la\ que\ depende\ de\ un\ patrón\ jefe\ o\ superior \\ 2 & si\ pea_i = 1\ y\ la\ persona\ i\ trabaja\ en\ un\ negocio\ propio\ en\ el\ que\ no\ depende \\ & jefe\ o\ superior\ y\ recibe\ o\ tiene\ asignado\ un\ sueldo \\ 3 & si\ pea_i = 1\ y\ la\ persona\ i\ trabaja\ en\ un\ negocio\ propio\ en\ el\ que\ no\ depende \\ & jefe\ o\ superior\ y\ no\ recibe\ o\ tiene\ asignado\ un\ sueldo \end{cases}$$

$$pea_i = \begin{cases} 1 & si\ la\ persona\ i\ es\ ocupada\ y\ edad_i \geq 16 \\ 2 & si\ la\ persona\ i\ es\ desocupada\ y\ edad_i \geq 16 \\ 0 & si\ la\ persona\ i\ pertenece\ a\ la\ pnea\ y\ edad \geq 16 \end{cases}$$

$$smlab_i = \begin{cases} 1 & si\ la\ persona\ i\ cuenta\ con\ acceso\ a\ servicios\ médicos\ como\ prestación\ laboral \\ 0 & si\ la\ persona\ i\ no\ cuenta\ con\ acceso\ a\ servicios\ médicos\ como\ prestación\ labor \end{cases}$$

$$aforlab_i = \begin{cases} 1 & si\ la\ persona\ i\ cuenta\ con\ acceso\ a\ un\ sistema\ de\ jubilación\ o\ pensión\ para\ e \\ 0 & si\ la\ persona\ i\ no\ cuenta\ con\ acceso\ a\ un\ sistema\ de\ jubilación\ o\ pensión\ par \end{cases}$$

$$aforecv_i = \begin{cases} 1 & si\ la\ persona\ i\ cuenta\ con\ acceso\ a\ un\ sistema\ de\ jubilación\ o\ pensión\ para \\ & el\ retiro\ por\ contratación\ voluntaria \\ 0 & si\ la\ persona\ i\ no\ cuenta\ con\ acceso\ a\ un\ sistema\ de\ jubilación\ o\ pensión \\ & para\ el\ retiro\ por\ contratación\ voluntaria \end{cases}$$

$$smcv_i = \begin{cases} 1 & si\ la\ persona\ i\ cuenta\ con\ acceso\ a\ servicios\ médicos\ por\ contratación\ voluntar \\ 0 & si\ la\ persona\ i\ no\ cuenta\ con\ acceso\ a\ servicios\ médicos\ por\ contratación\ volun \end{cases}$$

$$jub_i = \begin{cases} 1 & si\ la\ persona\ i\ es\ jubilada\ o\ pensionada \\ 1 & si\ la\ persona\ i\ declaró\ recibir\ jubilaciones\ o\ pensiones\ originadas\ dentro \\ & del\ país\ o\ provenientes\ del\ extranjero \\ 1 & si\ la\ persona\ i\ declaró\ contar\ con\ servicios\ médicos\ por\ jubilación \\ 0 & en\ otro\ caso \end{cases}$$

$$par_i = \begin{cases} 1 & si\ la\ persona\ i\ es\ jefe(a)\ del\ hogar \\ 2 & si\ la\ persona\ i\ es\ cónyuge\ del\ jefe(a) \\ 3 & si\ la\ persona\ i\ es\ hijo(a)\ del\ jefe(a) \\ 4 & si\ la\ persona\ i\ es\ padre/madre\ del\ jefe(a) \\ 5 & si\ la\ persona\ i\ es\ suegro(a)\ del\ jefe(a) \\ 6 & en\ otro\ caso \end{cases}$$

$$cony\_ss_n = \begin{cases} 1 & si\ la\ o\ el\ cónyuge\ del\ jefe(a)\ del\ hogar\ tiene\ acceso\ directo\ a\ la\ seguridad\ s \\ 0 & en\ otro\ caso \end{cases}$$

$$hijo\_ss_n = \begin{cases} 1 & si\ algún\ hijo(a)\ del\ jefe(a)\ del\ hogar\ tiene\ acceso\ directo\ a\ la\ seguridad\ so \\ & y\ no\ es\ jubilado\ ni\ pensionado \\ 1 & si\ algún\ hijo(a)\ del\ jefe(a)\ del\ hogar\ tiene\ acceso\ directo\ a\ la\ seguridad\ so \\ & es\ jubilado\ o\ pensionado\ y\ edad_i \geq 25 \\ 0 & en\ otro\ caso \end{cases}$$

$$jefe\_ss_n = \begin{cases} 1 & si\ el\ jefe(a)\ del\ hogar\ tiene\ acceso\ directo\ a\ la\ seguridad\ social \\ 0 & en\ otro\ caso \end{cases}$$

$$s\_salud_i = \begin{cases} 1 & si\ la\ persona\ i\ declara\ contar\ con\ servicios\ médicos\ de\ alguna\ institución\ d \\ & seguridad\ social\ pora\ algún\ familiar\ del\ hogar\ o\ de\ otro\ hogar,\ por \\ & muerte\ del\ asegurado\ o\ por\ contratación\ propia \\ 0 & en\ otro\ caso \end{cases}$$

$$pam_i = \begin{cases} 1 & si\ la\ persona\ i\ tiene\ 65\ años\ o\ más\ y\ recibe\ algún\ programa\ de\ adultos\ mayores \\ & ingreso\ mensual\ que\ recibe\ por\ este\ apoyo\ es\ igual\ o\ mayor\ al\ valor\ promedio \\ & de\ la\ canasta\ alimentaria \\ 0 & si\ la\ persona\ i\ tiene\ 65\ años\ o\ más\ y\ recibe\ algún\ programa\ de\ adultos\ mayores \\ & el\ ingreso\ mensual\ que\ recibe\ por\ este\ apoyo\ es\ menor\ al\ valor\ promedio\ de\ la \\ & canasta\ alimentaria \end{cases}$$

Indicador de carencia por Calidad y Espacios de la Vivienda

$$ic\_cv_{ihv} = \begin{cases} 1 & si\, icv\_pisos_{ihv} = 1\, o\, icv\_techos_{ihv} = 1\, o\, icv\_muros_{ihv} = 1\, o\, icv\_hac_{ihv} = \\ 0 & si\, icv\_pisos_{ihv} = 0\, e\, icv\_techos_{ihv} = 0\, e\, icv\_muros_{ihv} = 0\, e\, icv\_hac_{ihv} = \end{cases}$$

Donde

$$cv\_pisos_{ihv} = \begin{cases} 1 & si\, la\, vivienda\, v\, tiene\, piso\, de\, tierra \\ 2 & si\, la\, vivienda\, v\, tiene\, piso\, de\, cemento\, o\, firme \\ 3 & si\, la\, vivienda\, v\, tiene\, piso\, de\, linóleum\, congóleum\, o\, vinil \\ 4 & si\, la\, vivienda\, v\, tiene\, piso\, laminado \\ 5 & si\, la\, vivienda\, v\, tiene\, piso\, de\, mosaico,\, mármol\, o\, vitropiso \\ 6 & si\, la\, vivienda\, v\, tiene\, piso\, de\, madera,\, duela\, o\, parquet \end{cases}$$

$$icv\_pisos_{ihv} = \begin{cases} 1 & si\, cv\_pisos_{ihv} = 1 \\ 0 & si\, icv\_pisos_{ihv} > 1 \end{cases}$$

$$cv\_techos_{ihv} = \begin{cases} 1 & si\, la\, vivienda\, v\, tiene\, techos\, de\, desechos \\ 2 & si\, la\, vivienda\, v\, tiene\, techos\, de\, lámina\, de\, cartón \\ 3 & si\, la\, vivienda\, v\, tiene\, techos\, de\, lámina\, metálica \\ 4 & si\, la\, vivienda\, v\, tiene\, techos\, de\, lámina\, de\, asbesto \\ 5 & si\, la\, vivienda\, v\, tiene\, techos\, de\, palma\, o\, paja \\ 6 & si\, la\, vivienda\, v\, tiene\, techos\, de\, madera,\, o\, tejamanil \\ 7 & si\, la\, vivienda\, v\, tiene\, techos\, de\, teja \\ 8 & si\, la\, vivienda\, v\, tiene\, techos\, de\, terrado\, con\, viguería \\ 9 & si\, la\, vivienda\, v\, tiene\, techos\, de\, losa\, de\, concreto\, viguetas\, con\, bovedilla \end{cases}$$

$$icv\_techos_{ihv} = \begin{cases} 1 & si\, cv\_techos_{ihv} \leq 2 \\ 0 & si\, icv\_techos_{ihv} > 2 \end{cases}$$

$$cv\_muros_{ihv} = \begin{cases} 1 & si\ la\ vivienda\ v\ tiene\ muros\ de\ desechos \\ 2 & si\ la\ vivienda\ v\ tiene\ muros\ de\ lámina\ de\ cartón \\ 3 & si\ la\ vivienda\ v\ tiene\ muros\ de\ lámina\ metálica\ o\ de\ asbesto \\ 4 & si\ la\ vivienda\ v\ tiene\ muros\ de\ carrizo,\ bambú\ o\ palma \\ 5 & si\ la\ vivienda\ v\ tiene\ muros\ de\ embarro\ bajareque \\ 6 & si\ la\ vivienda\ v\ tiene\ muros\ de\ madera \\ 7 & si\ la\ vivienda\ v\ tiene\ muros\ de\ adobe \\ 8 & si\ la\ vivienda\ v\ tiene\ muros\ de\ tabique,\ ladrillo,\ block,\ piedra\ o\ concre \end{cases}$$

$$icv\_muros_{ihv} = \begin{cases} 1 & si\ cv\_muros_{ihv} \leq 5 \\ 0 & si\ icv\_muros_{ihv} > 5 \end{cases}$$

$$cv\_hac_{ihv} = \frac{num\_ind_{ihv}}{num\_cua_{ihv}}$$

Donde:

num_ind_{ihv}es el número de residentes en la vivienda

num_cua_{ihv}es el número de cuartos en la vivienda

$$icv\_hac_{ihv} = \begin{cases} 1 & si\ cv\_hac_{ihv} > 2.5 \\ 0 & si\ icv\_hac_{ihv} \leq 2.5 \end{cases}$$

Indicador de carencia por acceso a los Servicios Básicos en la Vivienda

$$ic\_sbv_{ihv} = \begin{cases} 1 & si\ isb\_agua_{ihv} = 1\ o\ isb\_dren_{ihv} = 1\ o\ isb\_luz_{ihv} = 1\ o\ isb\_combus_{ihv} = \\ 0 & si\ isb\_agua_{ihv} = 0\ e\ isb\_dren_{ihv} = 0\ e\ isb\_luz_{ihv} = 0\ e\ isb\_combus_{ihv} = \end{cases}$$

$$sb\_agua_{ihv} = \begin{cases} 1 & si\ la\ vivienda\ v\ obtiene\ agua\ de\ un\ pozo,\ río,\ lago,\ arroyo\ u\ otro \\ 2 & si\ la\ vivienda\ v\ obtiene\ agua\ de\ una\ pipa \\ 3 & si\ la\ vivienda\ v\ obtiene\ agua\ entubada\ que\ acarrea\ de\ otra\ vivienda \\ 4 & si\ la\ vivienda\ v\ obtiene\ agua\ entubada\ de\ la\ llave\ pública\ o\ hidrante \\ 5 & si\ la\ vivienda\ v\ obtiene\ agua\ a\ través\ de\ captores\ de\ agua\ de\ lluvia \\ 6 & si\ la\ vivienda\ v\ tiene\ agua\ entubada\ fuera\ de\ la\ vivienda\ pero\ dentro\ del \\ 7 & si\ la\ vivienda\ v\ tiene\ agua\ entubada\ dentro\ de\ la\ vivienda \end{cases}$$

$$sb\_agua\_captad_{ihv} = \begin{cases} 1 & si\ la\ vivienda\ v\ cuenta\ con\ captador\ de\ agua\ de\ lluvia\ que\ cumpl \\ & con\ la\ normatividad\ establecida\ por\ CONAGUA \\ 0 & en\ otro\ caso \end{cases}$$

$$isb\_agua_{ihv} = \begin{cases} 1 & si\ sb\_agua_{ihv} < 5 \\ 0 & si\ sb\_agua_{ihv} = 5\ y\ sb\_agua\_captad_{ihv} = 1 \\ 0 & si\ sb\_agua_{ihv} > 5 \end{cases}$$

$$sb\_dren_{ihv} = \begin{cases} 1 & si\ la\ vivienda\ v\ no\ tiene\ drenaje \\ 2 & si\ la\ vivienda\ v\ tiene\ drenaje\ conectado\ a\ una\ tubería\ que\ va\ a \\ & dar\ a\ un\ río,\ lago\ o\ mar \\ 3 & si\ la\ vivienda\ v\ tiene\ drenaje\ conectado\ a\ una\ tubería\ que\ va\ a \\ & dar\ a\ una\ barranca\ o\ grieta \\ 4 & si\ la\ vivienda\ v\ tiene\ drenaje\ conectado\ a\ una\ fosa \\ & séptica\ o\ tanque\ séptico \\ 5 & si\ la\ vivienda\ v\ tiene\ drenaje\ conectado\ a\ la\ red\ pública \end{cases}$$

$$isb\_dren_{ihv} = \begin{cases} 1 & si\ sb\_dren_{ihv} \leq 3 \\ 0 & si\ sb\_dren_{ihv} > 3 \end{cases}$$

$$sb\_luz_{ihv} = \begin{cases} 1 & si\ la\ vivienda\ v\ no\ tiene\ luz\ eléctrica \\ 2 & si\ la\ vivienda\ v\ obtiene\ luz\ eléctrica\ del\ panel\ solar\ o\ de\ otra\ fuente \\ 3 & si\ la\ vivienda\ v\ obtiene\ luz\ eléctrica\ de\ una\ planta\ particular \\ 4 & si\ la\ vivienda\ v\ obtiene\ luz\ eléctrica\ del\ servicio\ público \end{cases}$$

$$isb\_luz_{ihv} = \begin{cases} 1 & si\ sb\_luz_{ihv} = 1 \\ 0 & si\ sb\_luz_{ihv} > 1 \end{cases}$$

$$sb\_combus_{ihv} = \begin{cases} 1 & si\ en\ la\ vivienda\ v\ utilizan\ leña\ o\ carbón\ sin\ chimenea\ para\ cocinar \\ 2 & si\ en\ la\ vivienda\ v\ utilizan\ leña\ o\ carbón\ con\ chimenea\ para\ cocinar \\ 3 & si\ en\ la\ vivienda\ v\ utilizan\ gas\ de\ tanque\ para\ cocinar \\ 4 & si\ en\ la\ vivienda\ v\ utilizan\ gas\ natural\ o\ de\ tubería\ para\ cocinar \\ 5 & si\ en\ la\ vivienda\ v\ utilizan\ electricidad\ para\ cocinar \end{cases}$$

$$isb\_combus_{ihv} = \begin{cases} 1 & si\ sb\_combus_{ihv} \leq 1 \\ 0 & si\ sb\_combus_{ihv} > 1 \end{cases}$$

Indicador de carencia por acceso a la Alimentación Nutritiva y de Calidad

$$ic\_ali\_nc = \begin{cases} 1 & si\ lca = 1\ o\ ic\_ali = 1 \\ 0 & si\ lca = 0\ y\ ic\_ali = 0 \end{cases}$$

Donde:

$$lca = \begin{cases} 1 & limitado,\ si\ dch = 0\ o\ dch = 2 \\ 0 & no\ limitado\ si\ dch = 3 \end{cases}$$

$$dch = \begin{cases} 1 & si\ tot\_cpond \leq 28,\ hogar\ con\ dieta\ pobre \\ 2 & si\ 28 < tot\_cpond \leq 42,\ hogar\ con\ dieta\ limítrofe \\ 3 & si\ tot\_cpond > 42,\ hogar\ con\ dieta\ aceptable \end{cases}$$

Apéndice C

Código en R del Análisis de Componentes Principales

```
pobreza <- read.table("C:/Users/POBREZA2014/Base final/pobreza_VF.csv
header=TRUE, sep=",",na.strings="NA", dec=".", strip.white=TRUE)

##########CÁLCULO DE ACP COVARIANZAS##########

ACP <- princomp(~ic_ali+ic_asalud+ic_cv+ic_rezedu+ic_sbv+ic_segsoc+pl
       cor=FALSE,data=pobreza)
cat("\nComponent loadings:\n")
VecP<- print(unclass(loadings(ACP)))
cat("\nComponent variances:\n")
ValP<- print(ACP$sd^2)
  cat("\n")
Summary<- print(summary(ACP))
screeplot(ACP)
pobreza <<- within(pobreza, {
    PC4 <- ACP$scores[,4]
    PC3 <- ACP$scores[,3]
    PC2 <- ACP$scores[,2]
    PC1 <- ACP$scores[,1]
  })

##########GRÁFICA DE SEDIMENTACIÓN##########

screeplot(ACP, main="Componentes Principales", xlab='Componentes',
          axes=F, col="red")
plot(ACP, type="lines", main="Sedimentación")
```

```
##########GRÁFICA 2 DIMENSIONES##########

plot(ACP$scores[, 1], ACP$scores[, 2], col = ifelse(pobreza$cat_2 ==
"yellowgreen", ifelse(pobreza$cat_2==2,"blue", ifelse(pobreza$cat_2 =
"darkgoldenrod1", "firebrick1"))), pch = 16, xlab = "PC.1", ylab = "P

legend("bottomleft", legend=c("Pobre extremo", "Pobre",
"Vulnerable", "No pobre"),col=c("firebrick1", "darkgoldenrod1",
"blue", "yellowgreen"), pch=19)

##########GRÁFICA 3 DIMENSIONES##########

library(scatterplot3d)

scatterplot3d(ACP$scores[, 1], ACP$scores[, 2], ACP$scores[, 3],
color = ifelse(pobreza$cat_2 == 3, "yellowgreen",
ifelse(pobreza$cat_2 == 2, "blue",ifelse(pobreza$cat_2 == 1,
"darkgoldenrod1", "firebrick1"))),
pch=16, xlab = "PC.1", ylab = "PC.2", zlab= "PC.3")

legend("bottomleft", legend=c("Pobre extremo", "Pobre",
"Vulnerable", "No pobre"), col=c("firebrick1", "darkgoldenrod1",
"blue", "yellowgreen"), pch=19)

##########CÁLCULO DE ACP CORRELACIÓN##########

pobreza <- read.table("C:/Users/POBREZA2014/Base final/pobreza_VF.csv
header=TRUE, sep=",",na.strings="NA", dec=".", strip.white=TRUE)

carencias <- pobreza[,c(6:12)]

########## MATRIZ DE CORRELACIÓN POLICÓRICA########

library(polycor)

mpolicorica <- function(carencias) {
  dd <- numeric()
  length(dd) <- length(carencias) * length(carencias)
  attr(dd, "dim") <- c(length(carencias), length(carencias))

for(x in c(1:(length(carencias)-1)))
 for(y in c((x+1):length(carencias)))
 dd[x,y]<-(polychor(carencias[,x],carencias[,y]))
   for (x in c(1:length(carencias)))
```

```
   dd[x,x]<- 1    for(x in c(1:(length(carencias)-1)))
     for(y in c((x+1):length(carencias)))
     dd[y,x]<-dd[x,y]       print(dd)}
(*Aportaciones del Software libre R, Rafael Jódar Anchía)

policorica <- mpolicorica(carencias)

######Cálculo de vectores y valores propios#####

eigen_p <- eigen(policorica) help("eigen")

### Componentes Principales###

vector_p <- eigen_p$vectors vector_p

###Varianza por Componente principal###

valor_p <- eigen_p$values valor_p

###Gráfica de barras varianza###

barplot(valor_p, col = "red", ylab = "variances",
names=c("Comp.1","Comp.2","Comp.3","Comp.4","Comp.5","Comp.6","Comp.7

##Cálculo de scores##

tvector_p <- t(vector_p)
t_carencias <- t(carencias)
pc_scores <- tvector_p%*%t_carencias
scores <- t(pc_scores) scores
```

Apéndice D

Código en R del Análisis de Regresión Logística

```
pobreza <- read.table("C:/Users/POBREZA2014/Base final/pobreza_VF.2.c
header=TRUE, sep=",",na.strings="NA", dec=".", strip.white=TRUE)

##########APLICACIÓN DE REGRESIÓN LOGÍSTICA##########

modelo.1 <- glm(pobreza ~ ic_rezedu + ic_asalud
            + ic_segsoc + ic_cv + + ic_sbv + ic_ali,
data = pobreza, family = binomial) summary(modelo.1)

modelo.2 <- glm(pobreza ~ ic_rezedu, data = pobreza, family = binomia
summary(modelo.2)

modelo.3 <- glm(pobreza ~ ic_asalud, data = pobreza, family = binomia
summary(modelo.3)

modelo.4 <- glm(pobreza ~ ic_segsoc, data = pobreza, family = binomia
summary(modelo.4)

modelo.5 <- glm(pobreza ~ ic_cv, data = pobreza, family = binomial)
summary(modelo.5)

modelo.6 <- glm(pobreza ~ ic_sbv, data = pobreza, family = binomial)
summary(modelo.6)

modelo.7 <- glm(pobreza ~ ic_ali , data = pobreza, family = binomial)
summary(modelo.7)

##########TABLA DE FRECUENCIAS##########
```

```
fv.modelo.1 <- ifelse(modelo.1$fitted.values >= 0.5, 1,0)

tabla <-data.frame(pobreza$pobreza, fv.modelo.1)
prop.table(table(fv.modelo.1,pobreza$pobreza), margin=1)
length(fv.modelo.1)
length(pobreza$pobreza)

tabla <- within(tabla, {
       fv.modelo.1 <- factor(fv.modelo.1, labels=c('no pobre','pobre')
})

tabla <- within(tabla, {
     pobreza.pobreza <- factor(pobreza.pobreza, labels=c('no pobre','
})

local({
     .Table <- with(tabla, table(fv.modelo.1))
     cat("\ncounts:\n")
     print(.Table)
     cat("\npercentages:\n")
     print(round(100*.Table/sum(.Table), 2))
})

local({
     .Table <- with(tabla, table(pobreza.pobreza))
     cat("\ncounts:\n")
     print(.Table)
     cat("\npercentages:\n")
     print(round(100*.Table/sum(.Table), 2))
})

local({
     .Table <- xtabs(~fv.modelo.1+pobreza.pobreza, data=tabla)
     cat("\nFrequency table:\n")
     print(.Table)
     .Test <- chisq.test(.Table, correct=FALSE)
     print(.Test)
})
```