

**APLICACIÓN DE COMPONENTES PRINCIPALES EN EL
DESARROLLO DE ÍNDICES: SOCIAL, EDUCATIVO Y ECONÓMICO
PARA LA CLASIFICACIÓN DE LOS MUNICIPIOS DE PUERTO RICO**

Por
RICELA FELICIANO SEMIDEI

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de
MAESTRÍA EN CIENCIAS
en
MATEMÁTICAS (ESTADÍSTICA)

UNIVERSIDAD DE PUERTO RICO
RECINTO MAYAGÜEZ
2013

Aprobada por:

Edgardo Lorenzo González, Ph.D.
Presidente, Comité Graduado

Fecha

Dámaris Santana Morant, Ph.D.
Miembro, Comité Graduado

Fecha

Olgamary Rivera Marrero, Ph.D.
Miembro, Comité Graduado

Fecha

Moisés Orengo Avilés, Ph.D.
Represente de Estudios Graduados

Fecha

Omar Colón Reyes, Ph.D.
Director del Departamento

Fecha

Abstract of Thesis Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Master of Science

**APPLICATION OF PRINCIPAL COMPONENT FOR THE
DEVELOPMENT OF SOCIAL, EDUCATIVE AND ECONOMIC
INDEXES FOR THE CLASSIFICATION OF THE MUNICIPALITIES
OF PUERTO RICO**

By: RICELA FELICIANO SEMIDEI

Chair: Edgardo Lorenzo González

Major Department: Department of Mathematical Sciences

Several findings in this research could be relevant to decision making by government agencies of Puerto Rico. This decision making depends on knowledge of the social, educational and economic aspects of the municipalities. To quantify these aspect of the municipalities of Puerto Rico, indexes were created, and were calculated with the data from 2010. For the construction of these indexes, the multivariate technique of principal components was applied, then municipalities were grouped using partitioning around medoids (PAM) classification technique in each of the aspects. In addition, a quality of life index by municipalities was calculated by averaging the social, educational and economic indexes.

In previous studies, Pedroso [26] calculated the Human Development Index deve-

veloped by the United Nations for Puerto Rico with 2002 data, which measures social, educational and economic aspects in the country. For their part, Correa and Quintana [8], created a socioeconomic profile of each municipality in Puerto Rico. In the present study, indexes that measure the social, educational and economic for each municipality were created. The social aspect was measured using the following variables: Ageing Index, the number of part I crimes, number of murders, quantity of health facilities, percent of population under age 18 who live with a couple in charge, with their grandparents or with a female head of the household. To measure the educational aspect of the municipalities the following variables were used: percent of high school graduates, number of students per school, number of years the school has been in plan of improvement, number of students per teacher, number of schools and results of Puerto Rican Academic Achievement Tests. To measure the economic aspect the following variables were used: per capita income, unemployment rate, percent of population below poverty level, Gini Index, percent of houses occupied by renters, number of establishments dedicated to construction, municipal income, surplus or deficit and public debt of local municipality.

It was found that municipalities with higher social indexes are in the West area, the one with higher economic indexes are in the metropolitan area and the higher educational index in the Northeast area. By calculating the index of quality of life, the municipalities were categorized in high, medium or low. Most municipalities were categorized with a medium index. The municipalities with the highest quality of life index are concentrated in the metropolitan area.

Resumen de Tesis Presentada a Escuela Graduada
de la Universidad de Puerto Rico como Requisito Parcial de los
Requerimientos para el Grado de Maestría en Ciencias

**APLICACIÓN DE COMPONENTES PRINCIPALES EN EL
DESARROLLO DE ÍNDICES: SOCIAL, EDUCATIVO Y ECONÓMICO
PARA LA CLASIFICACIÓN DE LOS MUNICIPIOS DE PUERTO RICO**

Por: RICELA FELICIANO SEMIDEI

Consejero: Edgardo Lorenzo González

Departamento: Departamento de Ciencias Matemáticas

En esta investigación se presentan hallazgos relevantes que podrían ayudar a la toma de decisiones informadas en las agencias gubernamentales de Puerto Rico. Esta toma de decisiones depende del conocimiento de los aspectos sociales, educativos y económicos de los municipios. Para cuantificar cada aspecto se crearon índices para cada municipio de Puerto Rico que se calcularon con datos del año 2010. Para la construcción de estos índices se aplicó la técnica multivariada de componentes principales y se agruparon los municipios utilizando la técnica de clasificación de particiones alrededor de medoides (PAM) en cada uno de los aspectos. Además se calculó un índice de calidad de vida por municipio, promediando los índices social, educativo y económico.

En trabajos previos, Pedroso [26] calculó el Índice de Desarrollo Humano creado por la Organización de las Naciones Unidas para Puerto Rico con datos del 2002, el cual mide los aspectos: social, educativo y económico en el país. Por su parte, Correa

y Quintana [8], crearon un perfil socioeconómico de cada municipio de Puerto Rico. En esta investigación se crearon índices que miden los aspectos social, educativo y económico para cada municipio. Para medir el aspecto social se utilizaron las variables: Índice de Envejecimiento, la cantidad de delitos tipo I, cantidad de asesinatos, de establecimientos de salud, por ciento de la población menor de 18 años que vive con una pareja a cargo, con sus abuelos o con una mujer cabeza del hogar. Para medir el aspecto educativo de los municipios se utilizaron las variables: por ciento de personas graduadas de escuela superior, cantidad de estudiantes por escuela, cantidad de años en plan de mejoramiento de la escuela, cantidad de estudiantes por maestro, cantidad de escuelas y resultados de las Pruebas Puertorriqueñas de Aprovechamiento Académico. Para medir el aspecto económico se utilizaron las variables: ingreso per cápita, tasa de desempleo, por ciento de población bajo nivel de pobreza, Índice de Gini, por ciento de viviendas ocupadas por inquilinos, cantidad de establecimientos dedicados a la construcción, ingreso municipal, superávit o déficit municipal y deuda pública municipal.

Se encontró que los municipios con mayor índice social se encuentran en el área Oeste, los de mayor índice económico en el área metropolitana y los de mayor índice educativo en el área Noreste. Al calcular el índice de calidad de vida, se categorizaron los municipios en alto, medio o bajo. La mayoría de los municipios se categorizó con un índice medio. Los municipios de mayor índice de calidad de vida se concentran en el área metropolitana.

Dedicado a:

Mis padres: Lilliam I. Semidei Delgado y Ramón Feliciano Sepúlveda, por su ejemplo, inspiración, amor incondicional, entrega y apoyo en todas las etapas de mi vida.

AGRADECIMIENTOS

- Al Dr. Edgardo Lorenzo González, mi asesor, por su guía y apoyo.
- A la Dra. Olgamary Rivera Marrero y a la Dra. Dámaris Santana Morant, por su disposición durante esta tesis.
- A mis compañeros estudiantes graduados: Isnardo Arenas, César Barreto, Jesús Cajigas, Karina Gelis, Héctor Moyano y Xavier Terán por su ayuda incondicional durante la redacción y realización de la tesis.
- A todos mis compañeros estudiantes graduados, por su amistad y aliento durante la realización de mi grado de maestría, en especial a Abner Ortiz.
- Al personal administrativo, por su disposición.
- A mis estudiantes, por su comprensión y apoyo.

Índice general

ABSTRACT ENGLISH	II
RESUMEN ESPAÑOL	IV
AGRADECIMIENTOS	VII
LISTA DE CUADROS	XI
LISTA DE FIGURAS	XII
1. Introducción	1
1.1. Justificación	1
1.2. Antecedentes	4
1.3. Objetivos	9
1.3.1. Objetivo General	9
1.3.2. Objetivos Específicos	9
2. Revisión de Literatura	10
2.1. Técnicas Multivariadas	10
2.1.1. Variables múltiples	10

2.1.2.	Análisis de componentes principales	11
2.2.	Clasificación no supervisada	14
2.3.	Índices	20
3.	Metodología	22
3.1.	Variables	22
3.1.1.	Definiciones	23
3.1.2.	Descripción de las variables sociales	24
3.1.3.	Descripción de las variables educativas	25
3.1.4.	Descripción de las variables económicas	26
3.2.	Procedimiento	30
3.2.1.	Observación acerca del procedimiento	31
3.3.	Limitaciones y delimitaciones	31
4.	Resultados	33
4.1.	Análisis descriptivo de las variables	33
4.1.1.	Análisis descriptivo de las variables sociales	34
4.1.2.	Análisis descriptivo de las variables educativas	37
4.1.3.	Análisis descriptivo de las variables económicas	41
4.2.	Análisis de componentes principales	46
4.2.1.	Componentes principales de las variables sociales	46
4.2.2.	Componentes principales de las variables educativas	49
4.2.3.	Componentes principales de las variables económicas	52
4.3.	Índices	55
4.3.1.	Índice social	56

4.3.2.	Índice educativo	59
4.3.3.	Índice económico	61
4.3.4.	Índice de calidad de vida	62
4.4.	Análisis de <i>clusters</i>	65
4.4.1.	<i>Clusters</i> sociales	66
4.4.2.	<i>Clusters</i> económicos	69
4.4.3.	<i>Clusters</i> educativos	72
5.	Conclusión y trabajos futuros	79
5.1.	Conclusiones	79
5.2.	Trabajos futuros	81
	REFERENCIAS	83
	A. Gráficas de Análisis Descriptivo	87
	B. Gráficas relacionadas a índices	94
	C. Gráficas de análisis de <i>clusters</i>	96
	D. Fuentes de las variables utilizadas	100

Índice de cuadros

3.1. Fuentes de variables utilizadas para el año 2010. Ver Apéndice D para explicaciones detalladas.	28
4.1. Medidas de tendencia central y variabilidad del aspecto social I_{Soc} . . .	34
4.2. Medidas de tendencia central y variabilidad del aspecto educativo I_{Ed} .	38
4.3. Medidas de tendencia central y variabilidad del aspecto económico I_{Econ}	41
4.4. Análisis de componentes principales de las variables sociales	47
4.5. Análisis de componentes principales de las Variables Educativas	50
4.6. Análisis de componentes principales de las variables económicas	53
4.7. Medidas de tendencia central y variabilidad de I_{Soc}, I_{Ed}, I_{Ec}	56
4.8. Medidas de tendencia central y variabilidad de I	63
4.9. Tabla de similaridad para los <i>clusters</i> sociales	66
4.10. Tabla de similaridad para los <i>clusters</i> económicos	69
4.11. Tabla de similaridad para los <i>clusters</i> educativos	73
4.12. Índices para los municipios de Puerto Rico (Adjuntas a Florida)	76
4.13. Índice para los municipios de Puerto Rico (Guánica a Orocovis)	77
4.14. Índice para los municipios de Puerto Rico (Patillas a Yauco)	78

Índice de figuras

4.1. Gráfico de puntos de Índice de Envejecimiento y asesinatos por cada 10,000 habitantes. Cada punto representa un municipio.	35
4.2. Diagrama de caja de delitos tipo 1 y asesinatos por cada 10,000 habitantes	35
4.3. Diagrama de caja de las Pruebas Puertorriqueñas de Aprovechamiento Académico	39
4.4. Gráfico de puntos de por ciento de la población mayor de 25 años graduada de escuela superior. Cada punto representa un municipio.	40
4.5. Gráfico de puntos de ingreso municipal y superávit o déficit. Cada punto representa un municipio.	42
4.6. Gráfico de puntos de ingreso per cápita y por ciento bajo nivel de pobreza. Cada punto representa un municipio.	43
4.7. Diagrama de sedimentación de variables sociales para determinar la cantidad de componentes principales.	47
4.8. Análisis de componentes principales de variables sociales, gráfico de cargas.	48
4.9. Diagrama de sedimentación de variables educativas para determinar la cantidad de componentes principales.	50

4.10. Análisis componentes principales de las variables educativas, gráfico de cargas.	51
4.11. Diagrama de sedimentación de variables económicas para determinar la cantidad de componentes principales.	54
4.12. Análisis de componentes principales de variables económicas, gráfico de cargas.	54
4.13. Diagrama de caja de los índices I_{Soc}, I_{Ed}, I_{Ec}	56
4.14. Gráfico de puntos de índices I_{Soc}, I_{Ed}, I_{Ec} . Cada punto representa un municipio.	57
4.15. Gráficas del índice de calidad de vida	63
4.16. Mapa de la categorización de ICV Alto, Mediano o Bajo	65
4.17. Ancho de Silueta en <i>clusters</i> sociales	67
4.18. Diagramas de caja I_{Soc} en <i>clusters</i> sociales	68
4.19. Mapa de <i>clusters</i> sociales. Los números 1, 3, 2 y 4 en la leyenda corresponden al número del <i>cluster</i> al cual pertenece el municipio.	68
4.20. Ancho de Silueta en <i>clusters</i> económicos	70
4.21. Diagramas de caja I_{Ec} en <i>clusters</i> económicos	71
4.22. Mapa de <i>clusters</i> económicos. Los números 4, 3, 2 y 1 en la leyenda corresponden al número del <i>cluster</i> al cual pertenece el municipio.	72
4.23. Ancho de Silueta en <i>clusters</i> educativos	73
4.24. Diagramas de caja I_{Ed} en <i>clusters</i> educativos	74
4.25. Mapa de <i>clusters</i> educativos. Los números 5, 3, 1, 2 y 4 en la leyenda corresponden al número del <i>cluster</i> al cual pertenece el municipio.	74
A.1. Diagramas de caja de variables sociales	88

A.2. Gráficos de puntos de variables sociales. Cada punto representa un municipio.	89
A.3. Diagramas de caja de variables educativas	90
A.4. Gráficos de puntos de variables educativas. Cada punto representa un municipio.	91
A.5. Diagramas de caja de variables económicas.	92
A.6. Gráficos de puntos de variables económicas. Cada punto representa un municipio.	93
B.1. Gráficas de líneas de los índices I_{Soc}, I_{Ed}, I_{Ec} en cada municipio	95
C.1. Intervalos de confianza para la media de los índices de los municipios en sus <i>clusters</i>	96
C.2. Intervalos de confianza para la media de PC sociales	97
C.3. Intervalos de confianza para la media de PC educativos	98
C.4. Intervalos de confianza para la media de PC económicos	99

Capítulo 1

Introducción

1.1. Justificación

Un elemento fundamental para la toma de decisiones de un país es la data que registra los aspectos sociales, educativos y económicos del mismo. Esta data debe ser recopilada y organizada con precisión, para luego ser analizada de manera estadística. De esta forma se garantiza reflejar la calidad y certeza que se necesita para la toma de decisiones adecuadas. Para analizar estos aspectos, se pretende crear en esta investigación índices: social, educativo, económico y de calidad de vida.

Los aspectos sociales, educativos y económicos están claramente interrelacionados o correlacionados y cada uno de ellos incide en la calidad de vida de un pueblo. Un ejemplo de la interrelación es el desempleo, la salud y el grado escolar alcanzado por un individuo. En Puerto Rico, una de las formas en que estas variables están interrelacionadas es, por ejemplo, una persona desempleada no tiene suficiente dinero para

atender todas sus necesidades de salud, a pesar de las ayudas del gobierno. De igual manera, la economía y la educación se interrelacionan pues a una persona con poco dinero se le hace difícil costear servicios educativos complementarios para sus hijos, como por ejemplo, tutorías personalizadas. Otro ejemplo de interrelación entre las variables es que una persona con grado escolar alto no se espera que esté desempleada. Este último ejemplo lo evidencian datos del año 2006 del Departamento del Trabajo y Recursos Humanos, en el cual se indica que el 82.2% de la fuerza laboral de Puerto Rico tiene diploma de escuela superior o más. Muchas otras interrelaciones se ven en los aspectos económicos, sociales y educativos de Puerto Rico, como por ejemplo, en la drogadicción, la criminalidad y la deserción escolar.

A pesar de la existencia de las interrelaciones en los aspectos mencionados, es inadecuada la recopilación de información conjunta por las autoridades de Puerto Rico, pues se recopila y se trabaja la información de manera aislada. Una de las aportaciones de esta tesis será el análisis de estos aspectos en conjunto. Además de pensar en cada criterio por separado, se quiere desarrollar una medida o índice que cuantifique o refleje la calidad de vida de los individuos en los distintos municipios. La calidad de vida según Rubén Ardila [3] se define como: “un estado de satisfacción general, derivado de la realización de las potencialidades de la persona. Posee aspectos subjetivos y objetivos. Es una sensación subjetiva de bienestar físico, psicológico y social. Incluye aspectos como la intimidad, la expresión emocional, la seguridad percibida, la productividad personal y la salud objetiva. Como aspectos objetivos el bienestar material, las relaciones armónicas con el ambiente físico y social y con la comunidad, y la salud objetivamente percibida.” El enfoque de interés en esta investigación será la parte objetiva de la definición de calidad de vida.

El Programa de Naciones Unidas para el Desarrollo (PNUD) ha hecho un esfuerzo por desarrollar índices de calidad de vida a nivel mundial por países. En particular, han desarrollado un índice de calidad de vida que tiene el nombre de Índice de Desarrollo Humano (IDH) y se registra en el Informe sobre Desarrollo Humano, publicado anualmente por la PNUD. El IDH es el resultado de un análisis multivariado que considera variables demográficas, sociales y económicas. La importancia de este índice es el énfasis en el individuo como protagonista del desarrollo y la calidad de vida de un país. El IDH resulta de la combinación de tres aspectos que describen la calidad de la vida humana:

1. Cantidad y calidad de vida que permitan disfrutar de una vida larga y saludable.
2. Disponer de los beneficios que genera la educación.
3. Contar con las condiciones económicas apropiadas para alcanzar un nivel de vida digno.

Notemos que el aspecto 1 está relacionado a los factores sociales y de salud, mientras que el aspecto 2 atiende los factores educativos y el aspecto 3 refleja las condiciones económicas de la sociedad.

A nivel mundial este índice es importante para comparar la situación de calidad de vida entre los países, desarrollar ayudas a niños que lo necesiten para su manutención, y hacer estudios de impacto económico y social según la calidad de vida de los países, entre otros. Este índice, para Puerto Rico, fue calculado por primera vez en el año 2008 por Teresa Pedroso de la Universidad Nacional Autónoma de México, cuyos detalles se presentan en la Sección 1.2.

El propósito en esta tesis es calcular los índices: I_{Soc} , I_{Ed} , I_{Ec} para los aspectos sociales, educativos y económicos de los municipios de Puerto Rico para el año 2010. Luego se creará un índice I de calidad de vida que dependa a su vez de estos tres índices. Estos índices: I_{Soc} , I_{Ed} , I_{Ec} encaminarán una clasificación de los municipios de Puerto Rico, mediante técnicas de agrupación. Estos índices y la clasificación resultante, serán una herramienta adicional disponible relacionada a la calidad de vida de los municipios. Además, sería interesante reflexionar sobre las causas para obtener dichos resultados y sugerir acciones para mejorar la calidad de vida de todos los pueblos. Así mismo se puede determinar si hay uniformidad o diferencia marcada entre la calidad de vida en los distintos municipios. En adición, los resultados obtenidos pueden ser utilizados por agencias gubernamentales para tomar acciones pertinentes de manera informada que promuevan una mejor calidad de vida, según las necesidades de cada municipio.

1.2. Antecedentes

El índice IDH, mencionado anteriormente, desarrollado por Teresa Pedroso en el 2008 para Puerto Rico, corresponde a datos del 2002. Fue calculado como parte de un esfuerzo por tener una comparación de cada uno de los países latinoamericanos [26]. Pedroso obtuvo datos educativos del Censo, la esperanza de vida al nacer del Informe de Estadísticas Vitales del Departamento de Salud y el producto interno bruto de la Junta de Planificación. El IDH tiene tres categorizaciones: Desarrollo Humano Alto (IDH entre 0.800 a 1.000), Desarrollo Humano Medio (IDH entre 0.500 y 0.799) y Desarrollo Humano Bajo (IDH inferior de 0.500). El IDH de Puerto Rico calculado fue de 0.853 para los datos del año 2002, lo que clasifica a Puerto Rico en un país con

Desarrollo Humano Alto, esto quiere decir que Puerto Rico tiene buenas condiciones de vidas relacionadas a la sociedad, su educación y salud. Este índice coloca a Puerto Rico en posición 34, con el mismo índice que Argentina, con respecto a los demás 177 países del mundo para los que se calculó el índice en el 2002. Según el Reporte de Desarrollo Humano publicado en el 2004 [19], el IDH correspondiente al año 2002 de Estados Unidos fue de 0.939, el de República Dominicana de 0.738, el de Cuba 0.809, en promedio el de Latinoamérica fue de 0.777 y el del mundo 0.73. Lo que coloca a Puerto Rico con un índice mayor a los países vecinos y al índice promedio de Latinoamérica. Sin embargo, el índice de Puerto Rico es menor al de Estados Unidos.

En el desarrollo de esta tesis, se adapta un índice similar utilizando variables pertinentes a los aspectos: social, educativo y económico de Puerto Rico. Para ésto, se modificó el IDH al considerar algunas variables nuevas que Pedroso no utilizó en su trabajo y aplicando técnicas como análisis de componentes principales y clasificación no supervisada. Los indicadores recomendados por la PNUD y sus definiciones en el Informe de Desarrollo Humano de 2011 para medir estos tres aspectos son:

IR_1 : Esperanza de vida al nacer: años que vivirá un recién nacido si los patrones de mortalidad por edades imperantes en el momento de su nacimiento siguieran siendo los mismos a lo largo de toda su vida.

IR_2 : Años promedio de escolaridad: años promedio de escolaridad que reciben las personas de 25 años y más, según los niveles de logros educacionales de la población de acuerdo con la duración de cada nivel.

IR_3 : Años esperados de escolarización: años de instrucción que un menor en edad se espera que reciba si los patrones vigentes de las tasas de matrícula específicas

por edad se mantuviesen constantes durante toda su vida.

IR_4 : Ingreso Nacional Bruto (INB) per cápita: el ingreso total de una economía generado por su producción y la propiedad de los factores de producción, menos los ingresos pagados por el uso de los factores de producción de propiedad del resto del mundo, convertido a dólares estadounidenses usando las tasas de la paridad del poder adquisitivo (PPA), dividido por la población a mitad del año.

IR_5 : Clasificación según el IDH menos la clasificación según el INB per cápita: diferencia entre la clasificación del INB y la clasificación del IDH. Una cifra negativa indica que el país tiene una mejor clasificación en el INB que en el IDH.

IR_6 : IDH no referido a ingresos: valor del IDH calculado sólo a partir de los indicadores de esperanza de vida y educación.

Estos indicadores recomendados, junto a otros antecedentes, fueron un punto de partida para el proceso de selección de nuestras variables.

Un esfuerzo particular para medir el aspecto socioeconómico fue realizado por Yovani Correa y el Dr. Julio Quintana [8], de la Universidad de Puerto Rico en Mayagüez. En esta investigación se creó un perfil socioeconómico de los pueblos de Puerto Rico para el año 2000, utilizando 19 variables socioeconómicas y la técnica de reducción de variables de análisis de componentes principales. Las variables más importantes, según su peso en los componentes principales utilizados, resultaron ser las siguientes:

1. Por ciento de población mayor de 65 años
2. Por ciento de población empleada como gerencial o profesional

3. Por ciento de población bajo el nivel de pobreza
4. Tasa bruta de mortalidad por cada mil habitantes
5. Índice de Envejecimiento
6. Por ciento de hogares con mujer cabeza del hogar
7. Índice Vital

En estas variables se incluyen dos índices: Índice de Envejecimiento e Índice Vital. Además de estos índices, Correa utilizó el Índice de Gini, el cual no fue significativo en términos estadísticos. Sin embargo, el Índice de Gini, en particular, es importante para calcular el índice económico de los municipios, pues mide la justa distribución de las riquezas. Es decir, su interpretación contextual en esta investigación es relevante. Por este motivo, incluiremos el Índice de Gini a nuestras variables, junto al Índice de Envejecimiento.

Los índices son importantes porque permiten hacer comparaciones a través del tiempo y en diferentes territorios. Los tres índices que se utilizaron se calculan de la siguiente manera:

- Índice de Envejecimiento: Mide el porcentaje de adultos mayores de 65 años con respecto a los niños menores de 15 años.

$$I_{Envejecimiento} = \frac{población > 65}{población < 15} * 100 \quad (1.1)$$

- Índice Vital: Total de nacimientos en proporción al total de muertes por cada municipio.

$$I_{Vital} = \frac{Total_{nacimientos}}{Total_{defunciones}} \quad (1.2)$$

- Índice de Gini: Mide la distribución de las riquezas en la sociedad y es uno de los indicadores aplicados habitualmente en los estudios de desigualdad realizados por distintos organismos internacionales, como por ejemplo, las Naciones Unidas. Pérez [28] lo define de la siguiente forma: “Para una distribución de N rentas (X) con los valores ordenados en sentido creciente, esto es, $x_1 \leq x_2 \leq \dots \leq x_N$, el Índice de Gini, viene dado por la siguiente expresión:

$$I_{Gini} = \frac{\sum_{i=1}^{N-1} (p_i - q_i)}{\sum_{i=1}^{N-1} p_i} \quad (1.3)$$

donde p_i : es la proporción acumulada de rentistas tal que $p_i = \frac{i}{N}$, ésta es la proporción que suponen respecto al total los i rentistas que perciben rentas más pequeñas, o sea, cuya renta es menor o igual que x_i ; q_i : la proporción acumulada de rentas, tal que $q_i = \frac{A_i}{A_N}$, donde $A_i = \sum_{j=1}^i x_j$ es la renta acumulada por los i primeros individuos; y $A_N = \sum_{j=1}^N x_j$ es la renta total. Este valor q_i nos indica la proporción sobre el valor total de la renta acumulada por los primeros i individuos con rentas inferiores o iguales a x_i (i individuos menos ricos) tiene el valor total de la renta, la parte de renta q_i acumulada por la proporción p_i de individuos menos ricos nunca superará el valor p_i ” [28]. El Índice de Gini es un valor entre 0 y 100, donde 0 significa la absoluta uniformidad de las riquezas y 100 significa la absoluta desigualdad de riquezas. Por ejemplo, en el año 2009, Chile obtuvo un Índice de 22.6, mientras que Brasil obtuvo un Índice de 53.9 [18]. Estos datos señalan que la distribución de riquezas de Chile es mucho más uniforme que la de Brasil.

1.3. Objetivos

1.3.1. Objetivo General

El objetivo general de esta investigación es desarrollar índices sociales, educativos, económicos y de calidad de vida total que nos permitan entender la realidad de los municipios de Puerto Rico para el año 2010.

1.3.2. Objetivos Específicos

Los objetivos específicos para esta investigación son:

- Aplicar la técnica de reducción de dimensionalidad de componentes principales a variables ya identificadas en los aspectos sociales, educativos y económicos.
- Para cada uno de los aspectos sociales, educativos y económicos, utilizar estos componentes principales como variables en la clasificación de los municipios.
- Generar los índices I_{Soc} , I_{Ed} , I_{Ec} para cada aspecto social, educativo y económico utilizando un promedio ponderado de los componentes previamente calculados.
- Desarrollar un índice de calidad de vida total I para los municipios de Puerto Rico que dependa de los índices I_{Soc} , I_{Ed} , I_{Ec} .
- Categorizar los municipios de Puerto Rico a base de su puntuación en el índice de calidad de vida.

Capítulo 2

Revisión de Literatura

2.1. Técnicas Multivariadas

2.1.1. Variables múltiples

Las técnicas multivariadas se enfocan en los siguientes aspectos:

1. Simplificación de la estructura de datos - Intentan transformar las variables de interés para conseguir un conjunto de menor dimensión.
2. Clasificación - Intenta encontrar una ubicación espacial de los individuos y variables.
3. Interdependencia - Su objetivo es conocer la dependencia entre las variables.
4. Dependencia - Su propósito es encontrar la asociación entre dos conjuntos de variables.

5. Formulación y pruebas de hipótesis - Se propone un modelo estadístico, definido por unos parámetros para estimar, y luego se verifica de acuerdo a la información recopilada.

Entre los enfoques más usados están el de interdependencia y dependencia. El enfoque de dependencia consiste en encontrar relaciones entre conjuntos de variables. Uno de ellos se considera dependiente de otro conjunto. Entre los métodos de dependencia están las siguientes técnicas: regresión múltiple, análisis del discriminante, análisis de correlación canónica, análisis de varianza multivariado y análisis de conjunto. Por su parte, los métodos de interdependencia buscan la manera en que se relacionan las variables, pueden estar relacionadas a interdependencia de las variables o a manera de dependencia con un subconjunto de alguna de ellas. Algunos de los métodos de interdependencia son: análisis de componentes principales, análisis de factores comunes, análisis de correspondencias y análisis de conglomerados. Para esta investigación se utilizó el análisis de componentes principales.

2.1.2. Análisis de componentes principales

Descripción:

El método de interdependencia de análisis de componentes principales (ACP) es una técnica multivariada usada para generar un número menor de nuevas variables, que serán una combinación lineal de las variables originales con una pérdida mínima de la información contenida en las variables originales. En otras palabras, se quiere reducir la dimensionalidad de p variables a un subconjunto $r < p$. Dadas n observaciones de

p variables, se quiere representar esta información en combinaciones lineales de las p variables originales, tales que estas combinaciones expliquen la mayor parte de la variabilidad que se tenía inicialmente.

Este nuevo conjunto de variables más pequeño son combinaciones lineales de las variables que se tenían originalmente, y a éstos se les conoce como componentes principales. éstos serán no correlacionados y en el caso de que exista normalidad, también serían unos componentes independientes.

Algunas finalidades del ACP, según Díaz Monroy [12] es que permite:

- Generar nuevas variables que expresen la información contenida en un conjunto de datos.
- Reducir la dimensión del espacio donde están inscritos los datos.
- Eliminar las variables (si es posible) que aportan poco al estudio del problema.
- Facilitar la interpretación de la información contenida en los datos.

Propiedades de los componentes principales

Para este método no es necesario asumir distribución alguna de las variables. Las variables tienen que ser cuantitativas y es más efectivo aplicar el método de ACP a variables altamente correlacionadas. Los componentes principales, es decir, las combinaciones lineales de las variables originales, tienen las siguientes propiedades según Peña [27]:

1. Conservan la variabilidad inicial. La suma de las varianzas de los componentes principales es igual a la suma de las varianzas de las variables originales.

2. La proporción de variabilidad explicada por un componente es el cociente entre su varianza (el valor propio asociado al vector propio que lo define) y la suma de los valores propios de la matriz. Sea j un componente principal, entonces la varianza de este componente es λ_j y la suma de todas las varianzas de las variables originales es $\sum_{i=1}^p \lambda_i$ y la proporción de la variabilidad explicada es $\frac{\lambda_j}{\sum_{i=1}^p \lambda_i}$.

3. La covarianza entre cada componente principal y cada variable X están dadas por el producto de las coordenadas del vector propio que define el componente por el valor propio. Sea a_i el vector de coeficientes del componente z_i , entonces $Cov(z_i; x_1, \dots, x_p) = \lambda_i \mathbf{a}_i = (\lambda_i a_{i1}, \dots, \lambda_i a_{ip})$.

4. La correlación entre un componente principal y una variable X es proporcional al coeficiente de esa variable en la definición del componente, y el coeficiente de proporcionalidad es el cociente entre la desviación típica del componente y la desviación típica de la variable. Es decir: $Corr(z_i; x_j) = \frac{Cov(z_i x_j)}{\sqrt{Var(z_i)Var(x_j)}} = \frac{\lambda_i a_{ij}}{\sqrt{\lambda_i s_j^2}} = a_{ij} \frac{\sqrt{\lambda_i}}{s_j}$.

5. Los r componentes principales proporcionan la predicción lineal óptima con r variables del conjunto de variables X . Esto ocurre porque la mejor predicción lineal con r variables de las variables originales se obtiene utilizando los primeros r componentes principales.

6. Si estandarizamos los componentes principales, dividiendo cada uno por su desviación típica, se obtiene la estandarización multivariante de los datos originales. Haciendo que $E(X) = 0$ y $Var(X) = 1$.

Selección del número de componentes principales

Existe una serie de sugerencias para seleccionar el número de componentes principales adecuado en un problema según Peña [27] .

1. Hacer una gráfica del valor propio del componente versus el número del componente, es decir, λ_i vs. i . Se comienza a seleccionar los componentes hasta que los restantes tengan aproximadamente el mismo valor de λ_i . Se pretende encontrar un punto al partir del cual los valores propios son aproximadamente iguales.
2. Seleccionar componentes hasta cubrir con cierta cantidad de varianza deseada, usualmente el 80 %. Se debe tener precaución al usar esta sugerencia, pues se debe combinar a incluir ciertos componentes que se quieren estudiar que expliquen la forma que tienen las variables de manera más adecuada.
3. Descartar los componentes que tengan valores propios menores a una cota establecida, esta cota puede ser la varianza media $\frac{\sum_{i=1}^p \lambda_i}{p}$. Esta regla también es arbitraria, pues una variable que sea independiente del resto, suele llevarse un componente principal y puede tener un valor propio mayor que la unidad. Sin embargo, cuando las variables están no-correlacionadas con el resto puede ser que no sea relevante para el análisis.

2.2. Clasificación no supervisada

Las técnicas de agrupación se aplican con el propósito de identificar la estructura natural de los elementos de estudio y de encontrar esquemas útiles que expliquen la manera en que se agrupan los objetos. Requiere que la distribución entre los objetos

sea medida en términos de las variables medidas sobre ellos [12]. Luego de calcular las medidas, se agrupan los objetos y se verifica si realmente las estructuras definidas se encuentran en los datos.

La clasificación no supervisada intenta agrupar los datos sin información a priori de la posible clasificación que las variables tenían. Dentro de la clasificación no supervisada utilizaremos una de sus aplicaciones, la que llamaremos *clusters*.

Un *cluster* se define como un grupo de unidades experimentales con cierto grado de similitud entre sus unidades formados por el proceso de clasificación. La técnica de *clustering* consiste en dividir los datos en grupos a los que se les llama *clusters*, tal que los datos dentro de un grupo sean más similares entre sí que entre los datos que pertenezcan a un grupo diferente al suyo. Por ejemplo, al tomar los datos de los niveles de estrés de ciertas personas se pueden formar *clusters* de personas con características en común, probablemente con condiciones similares de trabajo. El objetivo principal del *clustering* es revelar la organización de patrones en grupos sensibles, que nos permiten descubrir similitudes y diferencias, así como derivar conclusiones útiles acerca de ellos [17]. En el proceso de *clustering*, como parte de la clasificación no supervisada no hay clases predefinidas y no hay ejemplos que puedan demostrar que tipo de relación se desea que sea válida acerca de la data, por esta razón es que se percibe como un proceso no supervisado [17]. Es decir, los grupos no están formados previamente con las características que desean, sino que a partir de los datos se va formando según sus semejanzas y diferencias. Estas comparaciones se establecen aplicando métodos tales como el *K-Means* y el Particionamiento alrededor de medoides (PAM). Luego de esta agrupación, se debe hacer una validez interna de la agrupación.

Pasos para trabajar un problema de *clustering*: El proceso para trabajar un problema de agrupación por *clustering* se resume en los siguientes pasos [17]:

- El primer paso es seleccionar apropiadamente las características en las que un *cluster* va a ser seleccionado, tal que obtenga toda la información posible de interés, a esto se le conoce como selección de características o *feature selection*.
- Luego se selecciona un algoritmo que tenga como resultado un buen esquema de agrupación para la data. Una medida de proximidad y un criterio de *clustering* son los que principalmente caracterizan este tipo de algoritmo, así como su eficiencia para definir un esquema de agrupación que se ajuste a la data. A este paso se le conoce como *clustering algorithm* o algoritmo de *clusters*. Algunos de estos algoritmos son: *K-mean* y PAM, los cuales explicaremos más adelante.
 - a) medida de proximidad - es la medida que cuantifica cuan similares son dos datos. En la mayoría de los casos tenemos que asegurarnos que los atributos seleccionados contribuyen igualmente a los cálculos de medidas de proximidad y que no hay atributos que dominan a otros. Se recomienda que se haga un pre-procesamiento y estandarización de los datos.
 - b) criterio de *clustering* - En este paso tenemos que definir el criterio de agrupación, el cual puede ser expresado vía una función de costo que mide la promedio entre datos. Hay que tener en cuenta el tipo de *clusters* que se espera que ocurran en la data.
- Una vez seleccionado el algoritmo, se procede a la validación de los resultados. La exactitud de los resultados del algoritmo es verificada usando criterios y técnicas

apropiadas, como por ejemplo, el Índice de Silueta. Como el algoritmo define *clusters* que no son conocidos con anterioridad, independientemente del método de *clustering*, la partición final de la data requiere algún tipo de evaluación en la mayoría de las aplicaciones. Entonces, definimos como un buen *clustering*, a una partición que se ajuste bien a la data.

- Finalmente, procedemos a la interpretación de los resultados. En muchos casos, los expertos en el área tienen que integrar los resultados del *clustering* con otra evidencia experimental para contextualizar la interpretación de los resultados.

Algoritmo de *clusters*

Los dos métodos más utilizados para definir un *cluster* se pueden clasificar en los siguientes tipos:

1. Métodos de Particionamiento - Dividen la data en conjuntos disjuntos, intentando determinar la cantidad de particiones que optimice una cierta función de criterio. Esta función de criterio enfatiza la estructura global o local de la data y su optimización en un proceso iterativo.
2. Métodos Jerárquicos - Proceden sucesivamente al unir los *clusters* más pequeños o de separar los *clusters* más grandes. El resultado del algoritmo es un dendograma que muestra la manera en que se relacionan los *clusters*. El dendograma se termina en un nivel deseado y se obtiene un *clustering* de la data en grupos disjuntos.

Para cada uno de estos métodos hay diferentes algoritmos. Nosotros vamos a utilizar el método de particionamiento, porque es mejor trabajando con una gran cantidad de datos [31].

Método de particionamiento

El método de particionamiento se aplica mayormente a data de tipo cuantitativo, aunque puede ser utilizado para data de tipo cualitativo, cuando se le hacen algunas modificaciones. Según Halkidi [17], una de las limitaciones del método de particionamiento es que no maneja de manera adecuada la data con ruidos o *outliers* y que no se le facilita descubrir *clusters* con formas no-convexas. Más bien están basados en ciertas suposiciones para particionar la data. Sin embargo, esta técnica suele usarse por términos de tiempo y costo de data en comparación al Método Jerárquico cuando se tienen muchos datos para el *cluster*. Es más rápida y más conveniente para casos como el nuestro en que tenemos una gran cantidad de datos.

Para los métodos de particionamiento tenemos las siguientes técnicas:

1. **K-Mean:** Es la optimización de una función objetivo que describe la siguiente ecuación:

$$E = \sum_{i=1}^c \sum_{x \in C_i} d(x, m_i) \quad (2.1)$$

Donde m_i es el centro del cluster C_i , y $d(x, m_i)$ es la distancia euclidiana entre el punto x y el punto m_i . Entonces el objetivo de optimización sería encontrar la distancia mínima de cada punto con respecto al centro del *cluster* al cual pertenece. El algoritmo comienza con un conjunto de centros de *clusters* c . Luego asigna cada objeto de la data al *cluster* cuyo centro esté más cerca a este y vuelve a computar los centros. El proceso continúa hasta que todos los centros de los *clusters* permanecen iguales, es decir, al re-calcularlos no hay cambios.

2. **Particionamiento alrededor de medoides (PAM).** El objetivo del PAM es determinar un objeto representativo para cada *cluster*, al que se la llama medoide.

El medoide se define como el objeto perteneciente a un *cluster* cuyo promedio de disimilaridad a todos los objetos en el *cluster* es mínima, es decir que se puede considerar como el punto más céntrico de la agrupación considerada, esto es la mediana [2]. El algoritmo comienza seleccionando un objeto como el medoide de cada uno de los c *clusters*. Luego, cada uno de los objetos que no fueron seleccionados, se agrupan con el medoide que sea más similar a él. El algoritmo de PAM intercambia los medoides con otros objetos que no fueron seleccionados hasta que todos los objetos se califican como medoides [17].

Validación de los resultados:

En el análisis de *cluster*, es importante encontrar la partición que mejor se ajuste a la data. Es por eso, que se han desarrollado técnicas para validar los resultados obtenidos en el proceso de *clustering*. Uno de los problemas principales es decidir la cantidad óptima de *clusters* que se ajusta mejor a la data de interés. Existen varias formas o técnicas para decidir ésta cantidad óptima, por ejemplo, la visualización de la división de la data en diferentes *clusters* en una gráfica de dos dimensiones. Según Figueras [14], las validaciones pueden realizarse de manera externa o interna. La validación interna consiste en dividir los datos en dos grupos y aplicar el algoritmo de clasificación en cada grupo y luego comparar los resultados de cada grupo. Mientras que la externa se realiza comparando los resultados obtenidos con un criterio externo o realizando un análisis de *cluster* con una muestra diferente a la realizada [14].

Dos criterios propuestos para la validación de los clusters, según Halkidi [17], son:

1. Compacidad: Consiste en minimizar la varianza dentro de cada *cluster*, con el propósito de que los elementos de cada *cluster* estén más cerca unos de otros.

2. Separación: Los *clusters* deben estar lo suficientemente separados uno de otro.

La distancia entre dos *clusters* diferentes puede ser medida por:

- a) enlace simple o vecino más próximo: Mide la proximidad entre dos grupos calculando la distancia entre sus elementos más semejantes
- b) enlace completo: Mide la proximidad entre dos grupos calculando la distancia entre sus objetos más lejanos.
- c) enlace medio: Mide la proximidad entre dos grupos calculando la distancia entre las medias de cada grupo.

2.3. Índices

Un índice es una expresión cualitativa o cuantitativa utilizada para describir los comportamientos de la realidad a través de variables o establecer relaciones entre variables. A través de los índices, se puede medir estos comportamientos y relaciones a través del tiempo. Los índices deben mantener un balance adecuado para representar el fin y los medios, marcando una relación que refleje lo que se desea calcular en el índice. El proceso de medición permite a las organizaciones planificar con mayor certeza y confiabilidad soluciones a sus problemas. Además de discernir con precisión las oportunidades de mejorar procesos determinados para explicar y analizar los hechos [16].

Para lograr que las mediciones de los índices sean efectivas, la guía de indicadores DANE [16] indica que es necesario cumplir con que sean pertinentes, relevantes, útiles en el tiempo, precisas, oportunas y económicas. Sin embargo, deben encontrarse disponibles en el periodo de tiempo en que la información es importante. Esta condición

nos limita en el uso de variables, a pesar de la importancia de las mismas.

El índice debe tener las siguientes características según la Guía para el Diseño de Indicadores [16] :

1. Simplificación: El indicador puede ser multidimensional (económico, social, cultural).
2. Medición: Compara la actualidad de una dimensión de estudio en el tiempo.
3. Comunicación: Los índices deben hacerse para la toma de decisiones. Mostrar disponibilidad de los índices a las agencias correspondientes.

En la construcción de un índice se debe determinar el rango que se desea obtener y se debe tener en consideración el peso que se le asignará a cada variable de las cuales depende el índice. Este peso puede ser asignado en términos de la variabilidad. Aunque no existe una regla definitiva acerca de la asignación de los pesos de las variables, Babbie recomienda que se asignen pesos iguales, a menos que haya razones para asignar un peso diferente a cada variable [4].

Una de las formas para el re-escalamiento de un índice, explicada por Schuschny y Soto [30], cuyo propósito es llevar los valores de una variable al intervalo $[0, 1]$, consiste en calcular todos los máximos y mínimos que la variable adquiere considerando todos los datos conjuntamente. Obtenidas por la siguiente ecuación:

$$y_t^i = \frac{x_t^i - \min_{\forall p} x_t^i}{\max_{\forall p} x_t^i - \min_{\forall p} x_t^i} \in [0, 1] \quad (2.2)$$

Capítulo 3

Metodología

El objetivo de esta tesis es calcular los índices sociales, económicos, educativos y de calidad de vida de los municipios de Puerto Rico para el año 2010. Para esto se seleccionó variables adecuadas que representen cada aspecto. En adición, se aplicó el método de componentes principales, para luego realizar los *clusters* pertinentes con respecto a los índices de cada municipio.

3.1. Variables

La selección de variables de interés se realizó tomando en cuenta varios criterios. El primer criterio es la pertinencia, pues las variables seleccionadas deben estar relacionadas a uno de los aspectos medidos (social, económico o educativo), que a su vez miden de alguna forma la calidad de vida de cada municipio. El segundo criterio que se utilizó fue la confiabilidad de la fuente. Es indispensable que las fuentes que proveen

los datos sean oficiales y respetadas a nivel nacional. En nuestro caso, los datos que se utilizaron fueron del 2010 de algunas agencias de gobierno, tales como la Oficina del Censo de Estados Unidos, el Departamento de Educación de Puerto Rico, la Oficina del Contralor, el Departamento de la Policía, entre otros. El tercer criterio es la disponibilidad, pues a pesar de que algunas variables cumplen los criterios mencionados, éstas no están disponibles. Un ejemplo de algunas variables educativas que no estuvieron disponibles son la deserción escolar y el ausentismo de maestros.

3.1.1. Definiciones

Para entender las variables utilizadas, debemos definir algunos términos:

- Los "hogares de familias" se componen de un jefe de hogar y una o más personas emparentadas al jefe de hogar por nacimiento, matrimonio o adopción. No incluyen las parejas del mismo sexo aunque el matrimonio se haya realizado en un estado que otorga licencias de matrimonio a parejas del mismo sexo.¹ Los hogares de parejas del mismo sexo se incluyen en la categoría de hogares de familia si hay, por lo menos, una persona adicional emparentada con el jefe de hogar por nacimiento o adopción. Los hogares de parejas del mismo sexo sin parientes del jefe de hogar presente se tabulan como hogares no en familia. Los "hogares no en familia" se componen de personas que viven solas o en hogares donde no hay nadie emparentado con el jefe de hogar[11].
- Un hogar que tiene al menos un miembro del hogar emparentado con el jefe del hogar por nacimiento, matrimonio o adopción es un "Hogar de familia". Al jefe de

¹Aunque en Puerto Rico no se otorgan licencias de matrimonio entre parejas del mismo sexo.

hogar se le llama “jefe de familia”. En este total se incluyen todas las personas que viven en este hogar sin importar su parentesco con el jefe del hogar. Las parejas del mismo sexo se incluyen en la sección de no emparentados [7].

- El “Cónyuge” representa el cónyuge del jefe de hogar. Esto no refleja todos los cónyuges en un hogar. Las respuestas de “cónyuge del mismo sexo” fueron incluidas en “pareja no casada” [11].

3.1.2. Descripción de las variables sociales

Las siguientes son las variables que utilizamos para medir el aspecto social en cada uno de los municipios:

1. **%Pob<18_MujCabHog**: Por ciento de la población menores de 18 años en hogares, quienes viven a cargo de una mujer, sin su esposo presente.
2. **Ind Env**: Esta variable se calculó mediante las variables: “Personas mayores de 65 años” y “Personas menores de 15 años” del Censo 2010. Se calculó una razón de Personas mayores de 65 años dividido por las personas menores de 15 años multiplicado por 100. Su interpretación, por ejemplo, para un valor de 80 sería que hay 80 personas mayores de 65 años por cada 100 niños menores de 15 años.
3. **%Abu_Nietos**: Población en hogares que es menor de 18 años y son nietos o nietas del jefe de familia.
4. **%Pob<18_HogPareja**: Por ciento de la población menor de 18 años que vive con una pareja a cargo.

5. **Est_salud:** Cantidad de establecimientos en cada municipio dedicados a la salud. Incluye hospitales, centros de vacunación y otros. Fue calculada por cada 10,000 habitantes.
6. **Del_Tip_1:** Cantidad de delitos tipo I. En estos se incluye asesinatos, violaciones, robo, agresión grave, escarmiento, apropiación ilegal y hurto de autos. Este número fue calculado por cada 10,000 habitantes.
7. **Asesinatos:** Cantidad de asesinatos ocurridos en cada municipio. Este número fue calculado por cada 10,000 habitantes.

3.1.3. Descripción de las variables educativas

Las variables que utilizamos para medir el aspecto educativo en cada municipio son:

1. **Grad_25:** Por ciento de la población mayor de 25 años con al menos el grado de escuela superior.
2. **Est_x_esc:** Estudiantes promedio por escuelas. Este dato se calculó a partir de una muestra de aproximadamente el 20% de las escuelas de cada municipio, dividiendo la cantidad de estudiantes entre la cantidad de escuelas.
3. **Yrs_PlanMej:** Años promedio en plan de mejoramiento de la muestra del 20% de las escuelas públicas de cada municipio. Esta variable, cualitativa ordinal, se convirtió a los años que representaba cada categoría. Primer Año, Segundo Año se interpretó, según su descripción, como 1 y 2 años respectivamente. Acción Correctiva y Reestructuración Escolar corresponde a 3 y 4 años respectivamente.

Mientras que Reestructuración 2, 3 y 4 corresponden a 5, 6 y 7 años en Plan de mejoramiento. La Asistencia intensiva se tomó como 10 años, aunque representa 8 años o más en Plan de Mejoramiento. Estas definiciones utilizadas se especifican en la Carta Circular Núm. 1-2012-2013 [5] del Departamento de Educación.

4. **PP_Elem**: Por ciento de estudiantes que aprobaron por materia las Pruebas Puertorriqueñas a nivel elemental. Estos datos se obtuvieron del promedio del 20 % de las escuelas públicas secundarias de cada municipio para el año escolar 2010-2011.
5. **PP_Secundaria**: Por ciento de estudiantes que aprobaron por materia las Pruebas Puertorriqueñas de Aprovechamiento Académico a nivel secundario. Estos datos se obtuvieron del promedio del 20 % de las escuelas públicas secundarias de cada municipio para el año escolar 2010-2011.
6. **Est_x_maest**: Número de Estudiantes Promedio por Maestro. Este dato fue tomado de una muestra del 20 % de las escuelas públicas. Se dividió la cantidad de estudiantes entre la cantidad de maestro muestreados por municipio.
7. **Escuelas_x_10M**: Cantidad total de Escuelas Públicas por cada 10,000 habitantes en el municipio.

3.1.4. Descripción de las variables económicas

Las siguientes son las variables que utilizamos para medir el aspecto económico en cada municipio:

1. **IPC** : Ingreso per cápita. Este es el ingreso de dinero promedio por cada persona en el municipio.
2. **%Pobreza** : Por ciento de la población de 18 años o más que se encuentra bajo el nivel de pobreza en cada municipio.
3. **Tasa_Des** : Tasa de desempleo en por ciento.
4. **%Viv_inq** : Por ciento de viviendas ocupadas por inquilinos
5. **Gini_5** : Índice de Gini estimado para 5 años obtenido de la Encuesta sobre la comunidad de Puerto Rico [13]. El censo solo publica los datos para todos los municipios para muestras de 5 años, ya que las muestras que utiliza para el Puerto Rico Community Survey son muy pequeñas. Esta variable, mide la distribución de riquezas en la sociedad, específicamente en cada municipio. Este índice se calcula en base a datos agupados de población, por ejemplo: por ciento de hogares, por ciento de ingreso, porciento acumulado de hogares y por ciento acumulado de ingresos.
6. **Ing_Mun** : Ingreso total, según los presupuestos aprobados en el municipio. Este número se calculó por cada 10,000 habitantes del municipio.
7. **Sup_Deficit** : Superávit o déficit al cierre del año fiscal. Los valores negativos representan un déficit y los positivos un superávit. Este número fue calculado por cada 10,000 habitantes del municipio.
8. **Deuda_Pub** : Deuda pública Municipal, según el Banco Gubernamental de Fomento para Puerto Rico. Calculada por cada 10,000 habitantes.

9. **Est_Const** : Establecimientos dedicados a la construcción por cada 10,000 habitantes.

Las variables que se utilizan en este trabajo son tomadas de varias fuentes, resumidas en el siguiente Cuadro 3.1.

Variable Social		Variable Educativa		Variable Económica	
Variable	Fuente	Variable	Fuente	Variable	Fuente
%Pob<18_MujCabHog	CENSO DP1	Grad_25	Contralor	IPC	Contralor
Ind_Env	CENSO DP1	Est_x_esc	DEPR	%Pobreza	Contralor
%Abu_Nietos	CENSO QT-P12	Yrs_PlanMej	DEPR	Tasa_Des	Contralor
%Pob<18_HogPareja	CENSO QT-P12	PP_Elem	DEPR	%Viv_inq	Censo QT-H3
Est_salud	County Business	PP_Secundaria	DEPR	Gini_5	Encuesta sobre comunidad
Del_Tip_1	Departamento de Policía	Est_x_maest	DEPR	Ing_Mun	Contralor
Asesinatos	Departamento de Policía	Escuelas_x_10M	DEPR	Sup_Deficit	Contralor
				Deuda_Pub	Contralor
				Est_Const	Contralor

Cuadro 3.1: Fuentes de variables utilizadas para el año 2010. Ver Apéndice D para explicaciones detalladas.

En resumen, entre las variables sociales podemos destacar que hay tres variables relacionadas a situación social en los hogares, dos variables relacionadas a la salud y dos variables relacionadas a la seguridad del individuo. Las variables de cantidad de niños menores de 18 años que viven con la mujer como cabeza del hogar, con sus abuelos o con una pareja a cargo están relacionadas a la situación de los hogares y el núcleo familiar. Mientras que las variables relacionadas a la salud son: Índice de Envejecimiento y establecimientos dedicados a la salud. Un número alto en el Índice de Envejecimiento, puede indicar una buena condición de salud en la población de ese

municipio, y un número alto en la cantidad de los establecimientos de salud, puede indicar la facilidad de los habitantes de ese municipio para atender sus condiciones de salud.

Por otro lado, las variables: delitos tipo I y asesinatos están relacionadas a la seguridad de los individuos del municipio. Esta última, también puede estar relacionada a la definición de calidad de vida de Ardila [3] cuando especifica que la calidad de vida puede ser vista de manera subjetiva, y uno de estos aspectos es la percepción de la seguridad de los individuos. El compartir estos datos a través de los medios de comunicación puede afectar la tranquilidad del individuo y sus relaciones armónicas con el ambiente físico y social.

Una de las clasificaciones para las variables educativas es el desempeño académico, a la cual corresponde: años en plan de mejoramiento y Pruebas Puertorriqueñas. Mientras que hay variables que tienen que ver con los recursos y facilidades, como por ejemplo: cantidad de escuelas, de estudiantes por escuela y de estudiantes por maestro. Y por último la variable de por ciento de población mayor de 25 años graduada de escuela superior, la cual mide la educación general básica del individuo.

Las variables económicas recogen información sobre aspectos económicos en el hogar, en empleos y en la administración municipal. Las primeras son: ingreso per cápita, por ciento de pobreza, Índice de Gini y viviendas ocupadas por inquilinos; las relacionadas a empleos son: tasa de desempleo y establecimientos dedicados a la construcción. Las relacionadas a la administración municipal son: ingreso municipal, superávit o déficit y deuda pública.

3.2. Procedimiento

Para la realización de esta tesis, se trabajaron diferentes etapas dentro de la investigación.

En la primera etapa, se seleccionaron las variables que tienen una relación importante con los tres aspectos: sociales, educativos y económicos. Entre los factores que se tomaron en cuenta para la selección de dichas variables están: la importancia de las variables para cada índice, según la literatura, trabajos previos y la disponibilidad de los datos. En esta etapa, se recolectaron los datos de las variables de cada uno de los municipios de Puerto Rico correspondientes al año 2010.

En la segunda etapa, se aplicó el análisis de componentes principales para las variables asociadas con cada aspecto: social, educativo y económico. Una vez aplicada la técnica multivariada, pasamos a desarrollar el índice correspondiente a cada uno de los aspectos, a éstos les nombramos I_{Soc} , I_{Ed} , I_{Ec} para cada municipio. Luego se desarrolló un índice de calidad de vida I , a través del promedio de los índices I_{Soc} , I_{Ed} , I_{Ec} calculados para cada pueblo.

En la tercera etapa, se clasificaron los municipios para cada uno de los índices: I_{Soc} , I_{Ed} , I_{Ec} ; a través de las técnicas de clasificación y se categorizó el índice de calidad de vida I .

En la última se hizo una comparación entre los municipios según los resultados de los índices calculados. Mostrando los principales hallazgos del análisis según los índices y la agrupación resultante.

3.2.1. Observación acerca del procedimiento

Luego de completar la primera etapa se hizo un análisis exploratorio de la data con estadística descriptiva acerca del comportamiento de esas variables para cada uno de los municipios. Se calcularon medidas de variabilidad y tendencia central, y se representan los datos en gráficas. El propósito de realizar este análisis es detectar algunos comportamientos de las variables en los municipios y tener una base como referencia para el análisis que se hará posteriormente. A través de esto, se encontraron municipios con índices *outliers* y se vio un panorama ilustrativo de cada variable para cada municipio.

3.3. Limitaciones y delimitaciones

La falta de uniformidad en la recopilación de datos estadísticos es un problema en Puerto Rico. Existe poca información recopilada de manera centralizada. El Instituto de Estadística de Puerto Rico, la legislatura y el gobierno de Puerto Rico han hecho varios esfuerzos a través de Leyes para que se organice la data con respecto a varias variables en una sola institución. Sin embargo, estos esfuerzos son recientes y aún no se habían cumplido en su totalidad en el año 2010.

Una de las limitaciones para esta tesis fue con las posibles variables de interés: “ausentismo de maestros” y “deserción escolar”. Aunque los maestros utilizan un sistema electrónico para registrar su asistencia a clases, dichos datos no fueron encontrados a pesar de nuestros intentos de obtenerlos. Con respecto a la deserción escolar, el abandono de la educación por parte de la población es una situación preocupante para el

desarrollo pleno de una sociedad y su calidad de vida. Hubiérase sido muy interesante ver si existe diferencia en los por cientos de deserción escolar por municipios. Además de su reflejo en los aspectos educativos y más aún en la calidad de vida de los municipios. En el 2010 la Cámara de Representantes aprobó el P. de la C. 703 para enmendar el Artículo 5 de la Ley 209 del 28 de agosto de 2003 Ley del Instituto de Estadísticas de Puerto Rico, para mantener actualizado datos confiables de la deserción escolar en Puerto Rico trimestralmente a la Asamblea Legislativa [25]. En el año 2011 se enmendó la Ley Orgánica del Departamento de Educación de Puerto Rico del 29 de julio de 2011 para ordenar al Secretario de Educación a proveer información acerca de la deserción escolar al Instituto de Estadísticas de Puerto Rico, al gobernador y a la Asamblea Legislativa [22]. Pese a todos los esfuerzos, la tesis corresponde a datos del año 2010, por lo que no tuvimos la ventaja de obtener esta información pues todavía no estaban en cumplimiento las enmiendas de las leyes mencionadas.

Capítulo 4

Resultados

En este capítulo se presentan los resultados obtenidos durante esta investigación con relación a los municipios de Puerto Rico para datos del 2010. Se hizo un análisis descriptivo de las variables, análisis de componentes principales y análisis de *cluster* para cada aspecto. Además se presentan los índices calculados para cada uno de los aspectos: social, educativo, económico y de calidad de vida por municipio.

4.1. Análisis descriptivo de las variables

Se hizo un análisis descriptivo de las variables utilizadas, a través de medidas de tendencia central y de dispersión, para lo que se calculó la media, desviación estándar, mediana, mínimo, máximo y coeficiente de variación. Además de diagramas de caja (*Box Plot*) y gráfico de puntos (*Dot Plot*) para representar la distribución de cada una de las variables utilizadas; los cuales nos ayudan a la visualización de los datos *outliers*.

4.1.1. Análisis descriptivo de las variables sociales

Variable	Media	DesvStd	CoefVar	Min	Mediana	Max
%Pob<18_MujCabHog	21.731	2.358	10.85	17.223	21.606	30.332
Ind_Env	70.03	14.92	21.31	42.46	67.82	117.40
%Abu_Nietos	11.474	2.415	21.05	6.894	11.485	18.904
%Pob<18_HogPareja	47.847	4.797	10.02	30.966	48.316	56.206
Est_salud	13.768	8.412	61.10	3.968	10.541	41.258
Del_Tip_1	141.27	67.58	47.84	49.09	129.06	369.85
Asesinatos	2.150	1.757	81.72	0.000	1.801	8.317

Cuadro 4.1: Medidas de tendencia central y variabilidad del aspecto social I_{Soc}

Algunas de las estadísticas descriptivas en el aspecto social que debemos resaltar se resumen en el Cuadro 4.1. Uno de los promedios particularmente relevante es la cantidad de delitos tipo I por cada 10,000 habitantes por municipio. Esta variable indica que hay un promedio de 141.27 delitos tipo I por cada 10,000 habitantes. Sin embargo, esta variable presenta una variabilidad significativa, ya que su coeficiente de variación fue de 47.84%, mayor que 30%, el que se considera alta variabilidad. Su diagrama de caja lo podemos ver en la Figura 4.2. El número de asesinatos por cada 10,000 habitantes tiene media de 2.15 y presenta el mayor grado de variabilidad, como vemos en la Figura 4.1. Lo cual indica un grado de heterogeneidad en el número de asesinatos por cada 10,000 habitantes por municipio. Por su parte, el Índice de Envejecimiento es otra variable que podemos recalcar pues indica que, en promedio, hay 70 personas mayores de 65 años por cada 100 jóvenes menores de 15 años por municipio, su gráfico de puntos se presenta en la Figura 4.1.

Cuando los valores de la media y la mediana difieren significativamente uno del otro, esto puede indicar la existencia de *outliers*. La mediana, al ser también una medida

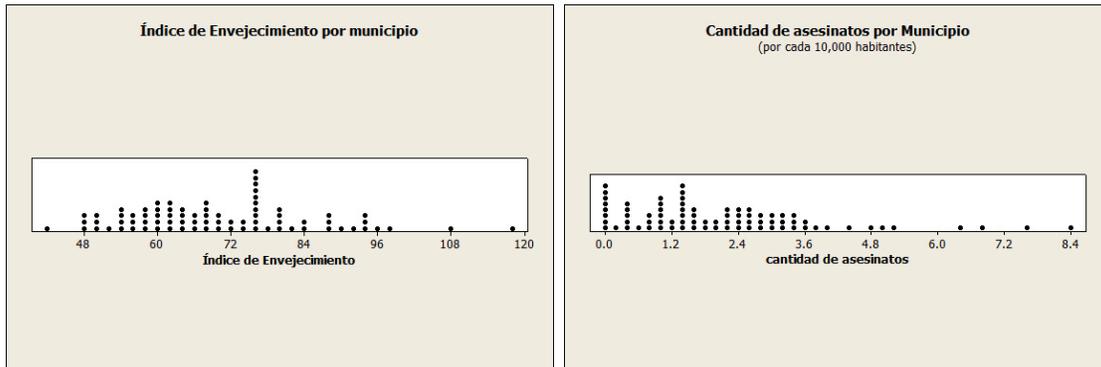


Figura 4.1: Gráfico de puntos de Índice de Envejecimiento y asesinatos por cada 10,000 habitantes. Cada punto representa un municipio.

de posición central no se ve afectada por estos datos inusuales, mientras que la media se afecta con este tipo de datos. En las variables que hay diferencia entre la media y la mediana está: delitos tipo I y asesinatos. En los delitos tipo I los valores son de 141.27 y 129.06 respectivamente. Esto se debe a que en esta variable hay cuatro municipios que son *outliers*, como vemos en la Figura 4.2. Vemos una situación similar en la cantidad de asesinatos por cada 10,000, pues su media y mediana son de 2.15 y 1.80 respectivamente.

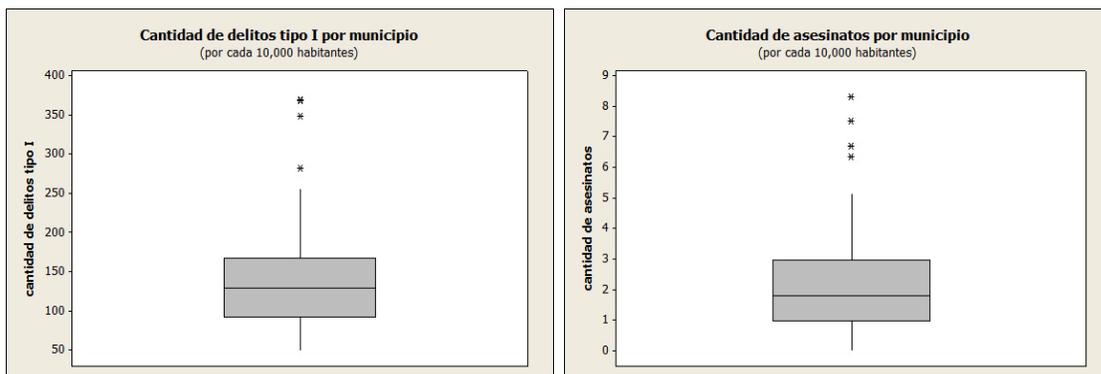


Figura 4.2: Diagrama de caja de delitos tipo 1 y asesinatos por cada 10,000 habitantes

Con respecto a la variación de los datos, podemos notar que las variables: cantidad de establecimientos de salud, delitos tipo I y asesinatos tienen mucha variabilidad. Las variables de: Índice de Envejecimiento, por ciento de niños menores de 18 años que viven con la mujer cabeza del hogar, abuelos o pareja muestran tener un comportamiento homogéneo.

A nivel isla podemos ver que la violencia es bastante heterogénea, al observar la variabilidad en los delitos tipo I y en los asesinatos. Una característica social que a través de la isla mantiene un comportamiento homogéneo es la situación en el hogar de los niños menores de 18 años, en términos del por ciento de estos niños que viven con sus abuelos, con una pareja a cargo o con una mujer cabeza del hogar.

Las variables sociales con mayor correlación son el por ciento de población menor de 18 años que viven con una pareja a cargo del hogar y los delitos tipo I, esta correlación fue de -0.61 . Esto significa, que a mayor cantidad de población menor de 18 años que viven con una pareja a cargo, menor cantidad de delitos tipo I hay en ese municipio. Las otras dos correlaciones mayores lo son la variable de la población menor de 18 años que viven con una pareja comparado contra los que viven con la mujer cabeza del hogar. Esta correlación fue de -0.552 , por lo que a mayor cantidad de niños que no viven con una pareja, mayor cantidad de niños viven con una mujer a cargo del hogar. El por ciento de población menor de 18 años que viven con una mujer cabeza del hogar está correlacionado al por ciento de la población que vive con sus abuelos, su correlación fue de 0.576 .

Hay municipios que se destacan en este análisis descriptivo por sus datos *outliers*, mínimos y máximos. Por ejemplo:

- Loíza es un municipio con varios *outliers*, tuvo un valor considerablemente alto en: el por ciento de niños menores de 18 años que viven con mujeres cabeza del hogar, el por ciento de niños menores de 18 años que viven con sus abuelos y en asesinatos por cada 10,000 habitantes. Siendo el valor máximo en estas últimas dos variables. Esto refleja un grave deterioro en la estructura familiar tradicional¹ y una mayor cantidad de violencia, para este municipio.
- Otro municipio de particular interés es Vieques, el cual refleja aspectos deficientes relacionados a su seguridad. Recordando que la cantidad de asesinatos se calculó en proporción a la población y se puede percibir erróneamente alto por la poca población de Vieques. Sin embargo, Vieques tuvo 7 asesinatos durante el año 2010, al igual que los municipios de Barceloneta, Barranquitas y Cabo Rojo, los cuales en comparación con Vieques, tienen una cantidad de habitantes significativamente mayor. Barceloneta tiene una población de 24,816, Barranquitas de 30,318, Cabo Rojo de 50,917 y Vieques de 9,301.

4.1.2. Análisis descriptivo de las variables educativas

En el Cuadro 4.2, la media del por ciento de personas mayores de 25 años graduada de escuela superior presentó un valor muy parecido al de la mediana. Notemos que la mitad de los municipios tienen menos de un 64 por ciento de personas mayores de 25

¹Se define la familia, en una definición tradicional de Louro Bernal, I (2001): “la célula fundamental de la sociedad, importantísima forma de organización de la vida cotidiana personal, fundada en la unión matrimonial y en los lazos de parentesco, en las relaciones multilaterales entre el esposo y la esposa, los padres y sus hijos, los hermanos y las hermanas, y otros parientes que viven juntos y administran en común la economía doméstica”.

años graduadas de escuela superior. Esta variable, refleja un comportamiento uniforme pues tiene poca variabilidad y no posee *outliers*. El valor mínimo de esta variable es el municipio de Las Marías con un por ciento de 49.3% de graduación de escuela superior para personas mayores de 25 años. El cual es sumamente bajo, le siguen San Sebastián con 52.8 y Maricao con 54.3. Esto puede deberse a que estos pueblos son tradicionalmente agrícolas por lo que es probable que muchos de ellos se quedaran trabajando en el campo en la agricultura y no terminaran su escuela superior. Los valores máximos para esta variable son: Rincón, Carolina, Aguadilla y Guaynabo con 80.7, 77.9, 77.8 y 77.7. Estos valores máximos ocurren en pueblos más industrializados costeros. Esto requieren una mano de obra diestra, que implica que haya una mayor tendencia a terminar al menos el grado de escuela superior.

Variable	Media	DesvStd	CoefVar	Min	Mediana	Max
Grad_25	64.577	6.281	9.73	49.300	63.750	80.700
Est_x_esc	317.5	95.9	30.19	123.0	312.5	650.5
Yrs_PlanMej	5.133	1.962	38.21	1.000	4.875	10.000
PP_Elem	54.65	14.25	26.07	28.00	52.12	91.00
PP_Secundaria	30.28	14.90	49.21	7.25	28.00	77.75
Est_x_maest	14.948	2.197	14.70	10.696	14.711	22.182
Escuelas_x_10M	4.990	1.788	35.83	2.145	4.675	11.001

Cuadro 4.2: Medidas de tendencia central y variabilidad del aspecto educativo I_{Ed}

La variable años en plan de mejoramiento tiene un promedio de 5.1 años en plan de mejoramiento, con una variabilidad alta. Esto es fatal para la educación del país, pues en promedio por municipio las escuelas llevan 5 años en plan de mejoramiento sin alcanzar una mejoría significativa. Cidra fue el municipio con menor cantidad de años, pues en promedio tiene un año en plan de mejoramiento. Mientras que los municipios con mayor cantidad de años en plan de mejoramiento fueron los municipios de Manatí,

Vieques, Ciales, Jayuya y Juana Díaz con 10 años en promedio de años en plan de mejoramiento, lo cual es altamente preocupante.

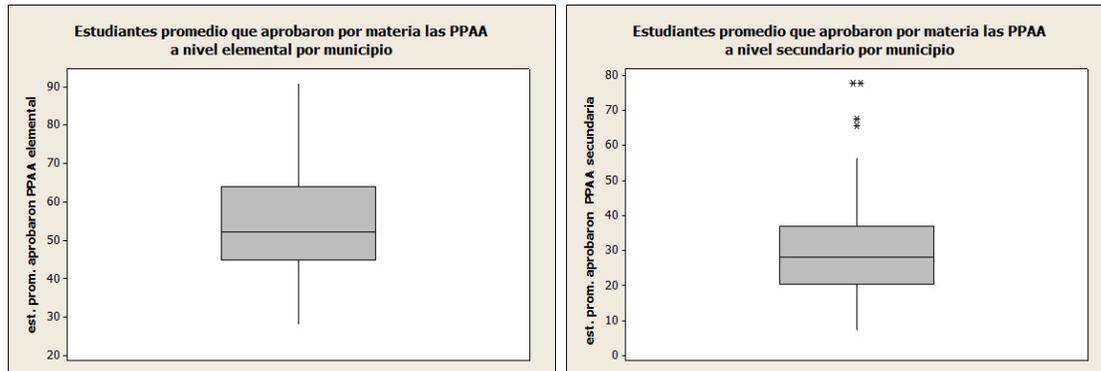


Figura 4.3: Diagrama de caja de las Pruebas Puertorriqueñas de Aprovechamiento Académico

La variable cantidad de estudiantes por escuela tiene una mediana de 312.5 y una media de 317.5, bastante similares. Los municipios con los valores más pequeños son Culebra con 123, Yabucoa con 143 y Arecibo con 150 estudiantes por escuela. Notemos que Culebra es el valor mínimo, pero debemos tener en cuenta que también es el municipio con menor población en Puerto Rico. Los valores máximos son: Fajardo, Ceiba y Loíza. Estos últimos municipios están ubicados en el Este, por lo que vemos una tendencia a una mayor matrícula en esta área. Por otro lado, el número de estudiantes por maestros tuvo un promedio de 14.7 estudiantes. Este número es muy bajo, pero debemos tener en cuenta lo siguiente: esta variable divide la cantidad de estudiantes entre la cantidad de maestros. Esta cantidad de maestros incluye la cantidad de maestros de educación física, artes, música, educación especial y otros. Encontramos también que la cantidad de maestros por escuela incluye el personal de apoyo, tales como: trabajadores sociales, orientadores y bibliotecarios. Esta variable tiene un comportamiento

homogéneo, tiene un máximo de 22.2 en Adjuntas y un mínimo de 10.7 en Culebra. Notemos que Culebra obtuvo el valor mínimo en ambas variables.

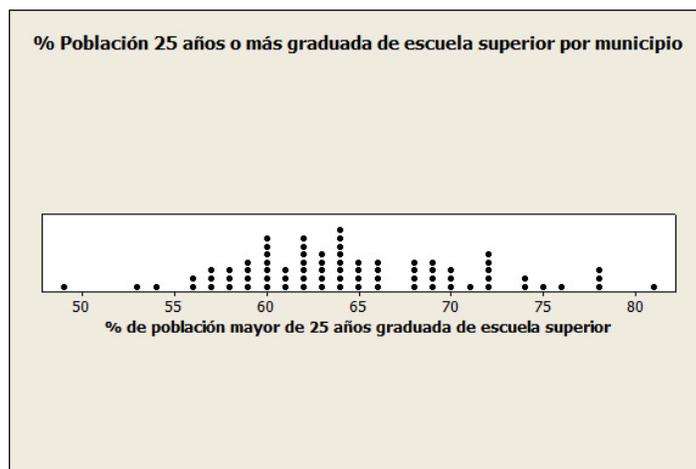


Figura 4.4: Gráfico de puntos de por ciento de la población mayor de 25 años graduada de escuela superior. Cada punto representa un municipio.

Acerca de las Pruebas Puertorriqueñas de Aprovechamiento Académico, en el Cuadro 4.2 notamos una tendencia de una mejor puntuación en las escuelas elementales, aunque en ambos el por ciento de aprobación fue muy bajo. En la escuela elemental el comportamiento de esta variable es más homogéneo que en la secundaria. El valor mínimo en las pruebas de nivel secundario es de 7.25 para el municipio de Loíza, que es inaceptable. En adición a esto, a nivel secundario sólo 8 municipios tienen más de un 50% aprobados por materia, como vemos en el gráfico de puntos en el Apéndice A.4, lo cual es alarmante. Estos municipios son: Añasco, Utuado, Vega Baja, Hatillo, Barceloneta, Salinas, Maricao y Barranquitas. De los cuales tres están entre los primeros diez municipios con menor cantidad de asesinatos.

4.1.3. Análisis descriptivo de las variables económicas

Variable	Media	DesvStd	CoefVar	Min	Mediana	Max
IPC	8,633	2,526	29.26	5,327	7,931	20,409
%Pobreza	46.097	8.658	18.78	25.400	47.500	62.100
Tasa_Des	18.206	3.279	18.01	10.200	18.000	25.300
%Viv_inq	25.676	4.995	19.46	17.000	24.250	45.400
Gini_5	0.50662	0.03309	6.53	0.43070	0.50455	0.59170
Ing_Mun	5,411,012	3,543,256	65.48	2,111,253	4,365,896	23,263,273
Sup_Deficit	-411,178	1,829,563	444.96	-7,449,528	24,149	3,927,733
Deuda_Pub	6,509,773	4,969,485	76.34	615,142	5,017,774	29,520,240
Est_Const ²	5.233	2.564	48.99	0.333	5.111	12.828

Cuadro 4.3: Medidas de tendencia central y variabilidad del aspecto económico I_{Econ}

En el Cuadro 4.3 se presentan algunas de las estadísticas descriptivas de interés sobre el aspecto económico. Entre las medias de particular interés, está la tasa de desempleo a nivel municipal. Este número alcanza el 18 % de personas desempleadas, en promedio, por cada municipio. Otra variable que tiene una media interesante es la de por ciento bajo nivel de pobreza, este valor es de 46.1 %, lo que indica que en promedio, por municipio, casi la mitad de la población está bajo el nivel de pobreza. En adición, ambas características tienen poca variabilidad según el coeficiente de variación. Esto refleja serios problemas económicos de una alta tasa de desempleo y un alto nivel de pobreza en forma homogénea a lo largo y ancho de Puerto Rico.

Entre las variables destacadas en la estadística descriptiva está: superávit o déficit. En esta variable encontramos una media negativa, mientras que su mediana es de \$24,149. Esto nos indica que la media, está siendo afectada por valores atípicos muy bajos, por ejemplo, Maricao, Culebra y Vieques. La mediana, por su parte nos indica

²Las Marías fue estimado en los datos encontrados, entre 0 y 19 establecimientos, utilizamos 10.

que la mitad de los datos está sobre 24, 249. En adición a los resultados del Cuadro 4.3, el 54 % de los municipios está en superávit y el 46 % está en déficit. El gráfico de puntos relacionado a esta variable, se presenta en la Figura 4.5

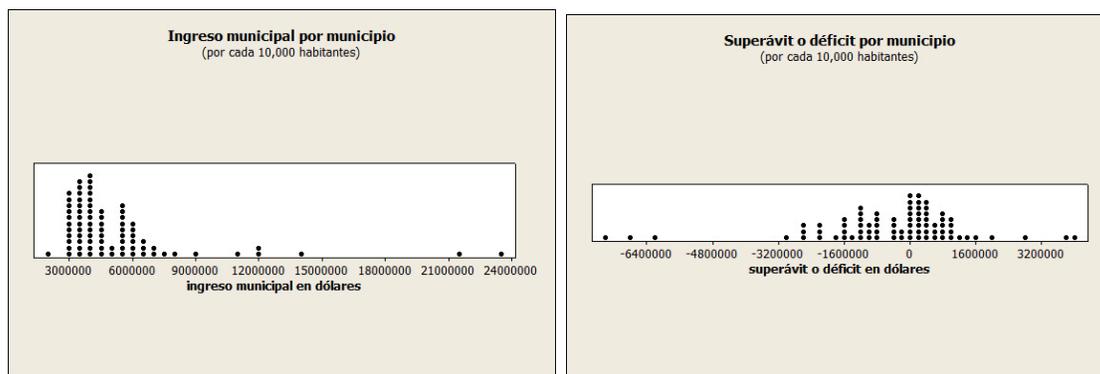


Figura 4.5: Gráfico de puntos de ingreso municipal y superávit o déficit. Cada punto representa un municipio.

El promedio de déficit por municipio es aproximadamente el 13 % del promedio de ingresos municipales. Mientras que el promedio de la deuda pública supera el promedio de los ingresos municipales en un 16 %. Sin embargo, hay una variabilidad muy alta en estas características y la variabilidad en el déficit, 445 %, es alarmante.

El Índice de Gini, por su parte, tiene un comportamiento homogéneo a través de los municipios de Puerto Rico y su promedio es de 0.51. El municipio con un mayor Índice de Gini fue Mayagüez con un valor de 0.5917, mientras que el valor menor de 0.4307 corresponde a Hormigueros, municipio que colinda con Mayagüez. Esto indica que el pueblo con mayor injusticia en la distribución de riquezas es Mayagüez, lo que podrá sugerir nuevos estudios a fines.

Una de las variables con mayor cantidad de *outliers* es ingreso municipal, como se

puede ver en el diagrama de caja del Apéndice A.5, la cual tiene 6 municipios con datos muy altos. Estos son: Culebra, Cataño, Guaynabo, Barceloneta, San Juan y Vieques. Siendo Culebra el municipio con mayor ingreso municipal por cada 10,000 habitantes, meramente reflejo de sus pocos habitantes. Al igual que el municipio de Vieques, es un municipio con muy pocos habitantes. Otra variable con muchos *outliers* es la variable de superávit o déficit, pues tiene dos municipios con valores muy altos en superávit: Fajardo e Isabela; y tres municipios con valores muy altos en déficit (valores negativos de la variable): Maricao, Culebra, Vieques. Para estos últimos tres *outliers* debemos tener en cuenta que estos municipios tienen muy pocos habitantes, y que también fueron los municipios con mayor ingresos municipales por habitante.

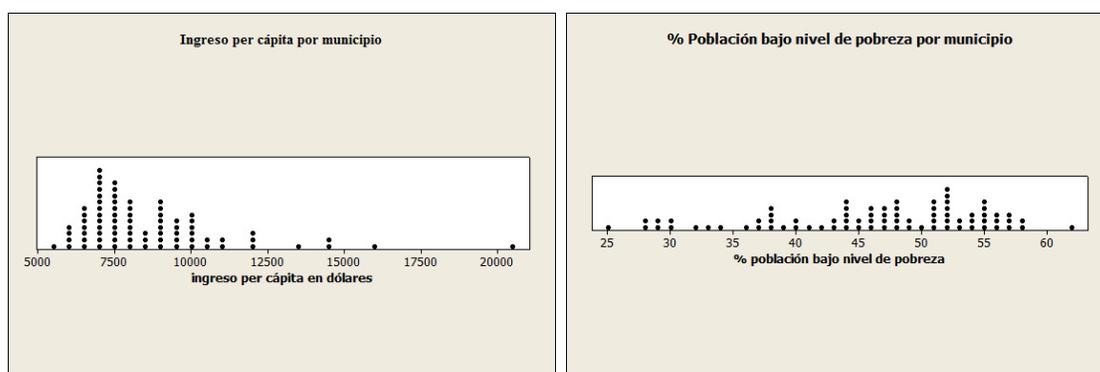


Figura 4.6: Gráfico de puntos de ingreso per cápita y por ciento bajo nivel de pobreza. Cada punto representa un municipio.

La variable ingreso per cápita tiene coeficientes de correlación de Pearson altos con la tasa de desempleo -0.857 , el por ciento de población bajo el nivel de pobreza -0.689 y la deuda pública 0.642 . El ingreso per cápita aumenta cuando la tasa de desempleo y el nivel de pobreza disminuyen. Además, cuando el ingreso per cápita aumenta, la deuda pública también aumenta. Otra correlación alta es entre los ingresos municipales

y la deuda pública. En este caso, el coeficiente de correlación de Pearson fue de 0.738. Esta es una correlación interesante, pues a mayor cantidad de ingresos, el municipio tiene mayor deuda pública.

En el Apéndice A.5 también se puede presenciar la cantidad de *outliers* de cada variable. La variable de Ingreso per Cápita tiene *outliers* en los municipios de Guaynabo, San Juan, Dorado y Trujillo Alto. La tasa de desempleo tiene solamente un valor atípico, Guaynabo, que es un *outlier* inferior. En el por ciento de viviendas ocupadas por inquilinos hay tres *outliers* superiores, estos son: San Juan, Mayagüez y Cataño. Mientras que los de la variable de la Deuda Pública Municipal son: Guaynabo, Barceloneta y Culebra. La variable de establecimientos dedicados a la construcción por cada 10,000 habitantes tiene tres *outliers* superiores y son Naranjito, San Juan y Guaynabo. Los mayores valores del Índice de Gini son Mayagüez, San Juan, Vega Baja, Aguadilla y Cataño. Mientras que los cinco valores menores corresponden a Fajardo, Ceiba, Loíza, Florida y Hormigueros; siendo este último el dato con menor Índice de Gini.

Con respecto a la variación, las variables de ingreso municipal, superávit o déficit, deuda pública y establecimientos dedicados a la construcción por cada 10,000 habitantes fueron las que representaron un comportamiento inconsistente, mientras que las variables: ingreso per cápita, por ciento bajo el nivel de pobreza, tasa de desempleo, viviendas ocupadas por inquilinos e Índice de Gini tienen un comportamiento homogéneo. Superávit o déficit fue la variable con mayor variabilidad, según su coeficiente de variación y la variable Índice de Gini, fue la variable con menor coeficiente de variación.

En resumen, podemos resaltar los siguiente municipios:

- Guaynabo, con el ingreso per cápita más alto y la deuda pública más alta, además de ser el municipio con la menor tasa de desempleo. También, Guaynabo es un dato *outlier* superior en la variable de ingresos municipales y en la variable de cantidad de establecimientos dedicados a la construcción. Esto indica que Guaynabo es un municipio con una actividad económica sólida. Aunque es preocupante la posible injusta distribución de riquezas reflejada en su Índice de Gini de 0.5476, el 11^{mo} municipio con mayor Índice de Gini en Puerto Rico para el 2010.
- San Juan, por su parte es un *outlier* superior en las variables: ingreso per cápita, Índice de Gini, ingreso municipal, establecimientos dedicados a la construcción y el valor máximo en por ciento de viviendas ocupadas por inquilinos. Es decir, San Juan tiene un alto ingreso per cápita, pero a la vez, tiene una tendencia hacia una distribución desigual de las riquezas, pues es el segundo municipio con mayor Índice de Gini. Es también el valor máximo en viviendas ocupadas por inquilinos, esto probablemente esté influenciado por la gran cantidad de universidades en el área de San Juan, incluyendo algunas de las que tienen mayor estudiantado de Puerto Rico, y la gran cantidad de residenciales públicos.
- Mayagüez fue un *outlier* superior en los datos de vivienda ocupadas por inquilinos y en el Índice de Gini. En Mayagüez está una de las universidades más concurridas en Puerto Rico y hay muchos estudiantes que se hospedan en esta área durante el transcurso de sus estudios universitarios. Además, hay un número significativo de residenciales públicos en este municipio.
- Vieques y Culebra son islas municipios que tienen muy poca población. A la hora de realizar un análisis, se debe tomar en cuenta que las variables utilizadas se

calcularon en base a la población, lo que hace que los valores de estos municipios sean muy altos al dividirlo entre un número pequeño correspondiente a su población. Sin embargo, puede ser que realmente estas variables reflejen un comportamiento atípico y se debe estudiar con especial atención ciertos detalles como el costo de transporte de personas y de artículos entre la isla grande y las islas municipio, la facilidad y variedad de empleo y recreación. En la “isla grande”, si un municipio no cuenta con varias facilidades, los habitantes pueden buscar satisfacer sus necesidades sociales, educativas y económicas en otros municipios cercanos, esto se dificulta en las islas municipio.

4.2. Análisis de componentes principales

El análisis de componentes principales se hizo para reducir la cantidad de variables en cada uno de los aspectos. Luego de hacer el análisis y seleccionar el número de componentes principales deseado, se rotularon los componentes según las variables que representaban.

4.2.1. Componentes principales de las variables sociales

El siguiente Cuadro 4.4 muestra el análisis de componentes principales obtenido para el aspecto social.

El valor propio para los primeros cuatro componentes principales en el aspecto social es de 2.62, 2.22, 0.73 y 0.57. Este último valor propio es significativamente menor que uno. Notemos que en el tercer componente ocurre un punto de inflexión notable en el

Valor Propio	2.6154	2.2180	0.7271	0.5710	0.4312	0.3214	0.1160
Proporción	0.374	0.317	0.104	0.082	0.062	0.046	0.017
Acumulado	0.374	0.690	0.794	0.876	0.938	0.983	1.000
Variable	PC1	PC2	PC3				
%Pob<18_MujCabHog	0.437	0.333	0.436				
Ind_Env	0.224	-0.533	0.166				
%Abu_Nietos	0.175	0.541	0.241				
%Pob<18_HogPareja	-0.557	0.127	-0.020				
Est_salud	0.264	-0.416	0.458				
Del_Tip_1	0.437	-0.150	-0.656				
Asesinatos	0.397	0.317	-0.290				

Cuadro 4.4: Análisis de componentes principales de las variables sociales

Diagrama de Sedimentación de las Variables Sociales en la Figura 4.7.

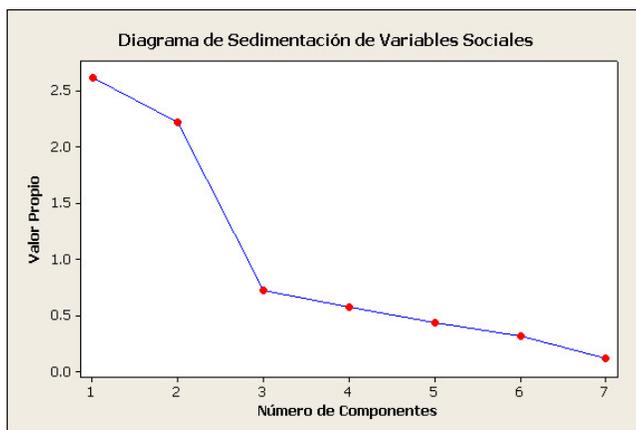


Figura 4.7: Diagrama de sedimentación de variables sociales para determinar la cantidad de componentes principales.

En cuanto a la variabilidad, con dos componentes principales, explicamos el 69% de variabilidad y se busca que sea mayor de 70%. Con tres componentes principales explicamos un 79.4% de variabilidad y con cuatro componentes principales explicamos

un 87.6 %. Además, en la Figura 4.7, notamos un punto de inflexión en tres componentes principales. En nuestro caso, tomamos la decisión de seleccionar tres componentes principales.

El gráfico de cargas de la Figura 4.8 muestra la relación de las variables con los primeros dos componentes.

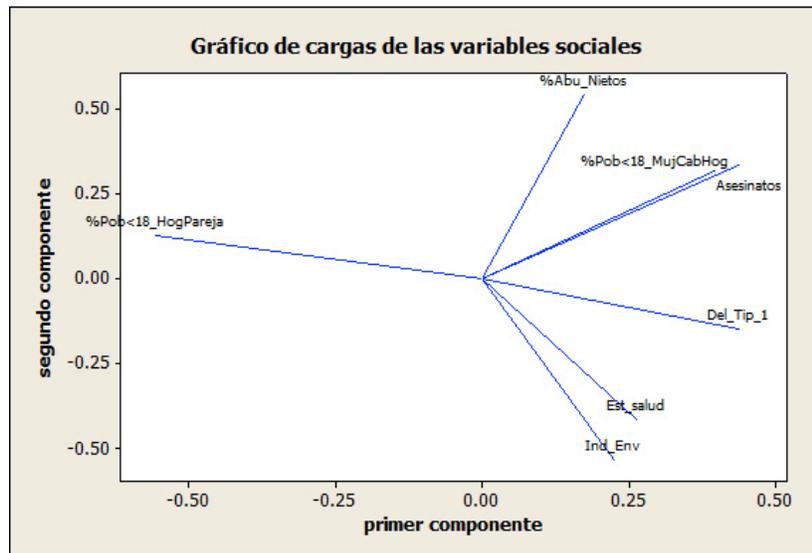


Figura 4.8: Análisis de componentes principales de variables sociales, gráfico de cargas.

Para rotular los componentes se analizó el tipo de variables que se relacionaban mejor con ellos, utilizando los resultados de la Cuadro 4.4. Los tres componentes se rotularon de la siguiente manera:

- Sistema de crianza sustituto: el primer componente principal tiene un peso negativo para la variable por ciento de la población menor de 18 años que vive con una pareja y valores positivos para todas las demás variables sociales. Notamos que entre los valores más altos positivos está el por ciento de la población me-

nor de 18 años que vive con una mujer cabeza del hogar, los delitos tipo I y los asesinatos.

- Apatía hacia la tercera edad: el segundo componente principal tiene pesos negativos para las variables: Índice de Envejecimiento, cantidad de establecimientos dedicados a la salud y cantidad de delitos tipo I, siendo las primeras dos con mayor magnitud que la última. Es decir, los valores negativos corresponden a las variables que miden la salud de los individuos de los municipios. Las variables positivas con mayor magnitud es el por ciento de población menor de 18 años que vive con sus abuelos y con una mujer cabeza del hogar, así como la cantidad de asesinatos.
- Percepción de bienestar social: el tercer componente principal tiene pesos negativos para las variables por ciento de la población menor de 18 años que vive con una pareja, cantidad de delitos tipo I y cantidad de asesinatos, siendo la primera de una magnitud significativa. Entre las mayores magnitudes para las cuales los valores son positivos, tenemos el por ciento de la población menor de 18 años que vive con una mujer cabeza del hogar y establecimientos de salud.

4.2.2. Componentes principales de las variables educativas

El siguiente Cuadro 4.5 muestra el análisis de componentes principales obtenido para el aspecto educativo.

El valor propio para los primeros cuatro componentes principales son: 2.4317, 1.1672, 0.9801 y 0.9261. Estos últimos valores no son mayores de uno, sin embargo, el

Valor Propio	2.4317	1.1672	0.9804	0.9261	0.6319	0.4776	0.3852
Proporción	0.347	0.167	0.140	0.132	0.090	0.068	0.055
Acumulado	0.347	0.514	0.654	0.786	0.877	0.945	1.000
Variable	PC1	PC2	PC3				
Grad_25	0.431	-0.247	-0.313				
Est_x_esc	0.465	-0.079	0.485				
Yrs_PlanMej	0.161	0.587	0.452				
PP_Elem	-0.287	-0.408	0.479				
PP_Secundaria	-0.349	-0.409	0.364				
Est_x_maest	0.456	-0.022	0.302				
Escuelas_x_10M	-0.399	0.504	0.093				

Cuadro 4.5: Análisis de componentes principales de las Variables Educativas

0.98 está muy cerca de uno y tiene una variabilidad explicada de 65%. Mientras que con el valor propio de 0.9261 explicamos el 79% de la variabilidad. En la Figura 4.9 vemos el diagrama de sedimentación que con 2 componentes tiene un punto de inflexión. Sin embargo, es muy poca la variabilidad explicada de 5.4%.

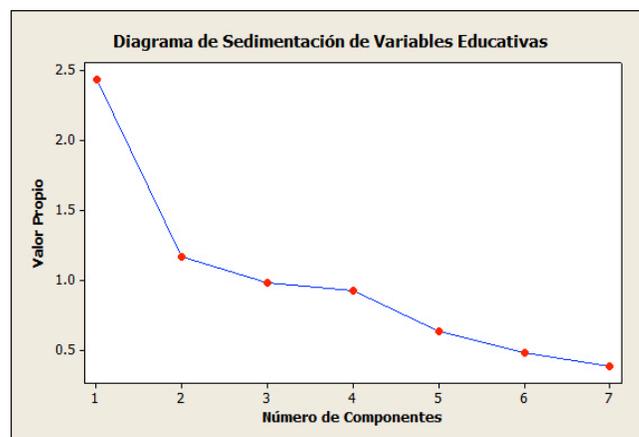


Figura 4.9: Diagrama de sedimentación de variables educativas para determinar la cantidad de componentes principales.

En nuestro caso, decidimos seleccionar tres componentes principales, considerando el 65.4 % como una variabilidad explicada suficiente, basándonos además en el principio de parsimonia de escoger un modelo adecuado que tenga la menor cantidad de variables posibles.

En la Figura 4.10, mostramos el gráfico de cargas de las variables educativas en relación a los primeros dos componentes.

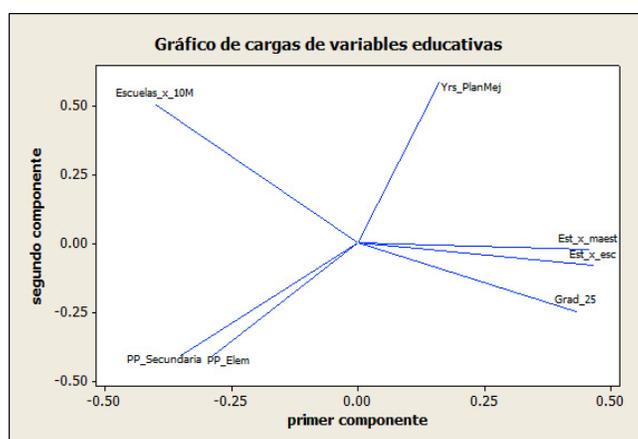


Figura 4.10: Análisis componentes principales de las variables educativas, gráfico de cargas.

Para rotular los componentes se analizó el tipo de variables que se relacionaban mejor con ellos, utilizando los resultados. Los tres componentes se rotularon de la siguiente manera:

- Acceso a la educación masiva: el primer componente principal para las variables educativas tiene valores positivos significativos para las variables del por ciento de personas mayores de 25 años con al menos el grado de escuela superior, la cantidad de estudiantes por escuela y la cantidad de estudiantes por maestro.

Tiene un valor negativo significativo para las escuelas por cada 10 mil habitantes. En esta variable percibimos un deseo por ser educado, pues hay un gran cantidad de estudiantes en las escuelas y muchas personas que se gradúan, a pesar de las pocas escuelas que hay.

- Pobre desempeño académico: el segundo componente principal tiene valores negativos para las variables relacionadas a las Pruebas Puertorriqueñas y un valor positivo en los años en plan de mejoramiento. Además, tiene un valor positivo en la cantidad de escuelas por cada diez mil habitantes.
- Presión del entorno: el tercer componente principal tiene un valor negativo para la variable del por ciento de personas mayores de 25 años graduadas de escuela superior y un valor positivo para las demás variables. Siendo las variables positivas significativas la cantidad de estudiantes por escuela, la cantidad de años en plan de mejoramiento y las Pruebas Puertorriqueñas de Aprovechamiento Académico a nivel elemental. En este componente se refleja un abandono de la escuela, pues en las Pruebas, el nivel alcanzado en la escuela elemental tiene un mayor peso que en las Pruebas a nivel secundario. Además de que hay un efecto negativo de la variable de por ciento de graduación de escuela superior. Esto puede deberse a la deserción escolar por presiones de grupo, pues también se muestra que la variable de cantidad de estudiantes por escuela tiene un efecto positivo.

4.2.3. Componentes principales de las variables económicas

El siguiente Cuadro 4.6 muestra el análisis de componentes principales obtenido para el aspecto económico.

Valor Propio	3.4279	1.8302	1.3874	0.7592	0.5627	0.4978	0.3135	0.1404
Proporción	0.381	0.203	0.154	0.084	0.063	0.055	0.035	0.016
Acumulado	0.381	0.584	0.738	0.823	0.885	0.941	0.975	0.991
Variable	PC1	PC2	PC3					
IPC	0.470	-0.269	-0.005					
%Pobreza	-0.384	0.429	0.151					
Tasa_Des	-0.404	0.224	-0.130					
%Viv_inq	0.194	0.488	0.247					
Gini_5	0.108	0.368	0.528					
Ing_Mun	0.354	0.390	-0.354					
Sup_Deficit	0.010	-0.353	0.641					
Deuda_Pub	0.457	0.167	-0.121					
Est_Const	0.292	0.127	0.263					

Cuadro 4.6: Análisis de componentes principales de las variables económicas

El valor propio para los primeros cuatro componentes principales es de 3.43, 1.83, 1.39 y 0.76. Este último valor propio no es mayor de uno. Notemos en el Cuadro 4.6 que el tercer componente principal tiene un valor propio de 1.39, con una variabilidad explicada de 74 %. Mientras que con el valor propio de 0.76 explicamos el 88 % de la variabilidad. En cuanto a la variabilidad explicada, con dos componentes principales, explicamos el 69 % de variabilidad, con tres componentes principales explicamos 79.4 %. En este caso, decidimos seleccionar tres componentes principales, considerando el 79.4 % como una variabilidad explicada suficiente en nuestro caso. En la Figura 4.11 se presenta el diagrama de sedimentación, que sugiere entre dos y cuatro componentes principales.

En la Figura 4.12, mostramos el gráfico de cargas de las variables económicas en relación a los primeros dos componentes.

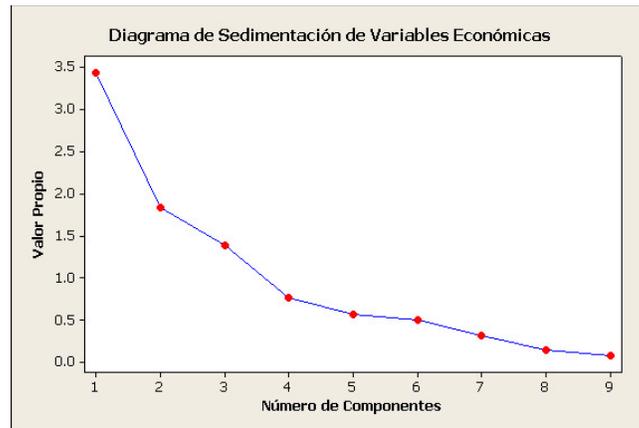


Figura 4.11: Diagrama de sedimentación de variables económicas para determinar la cantidad de componentes principales.

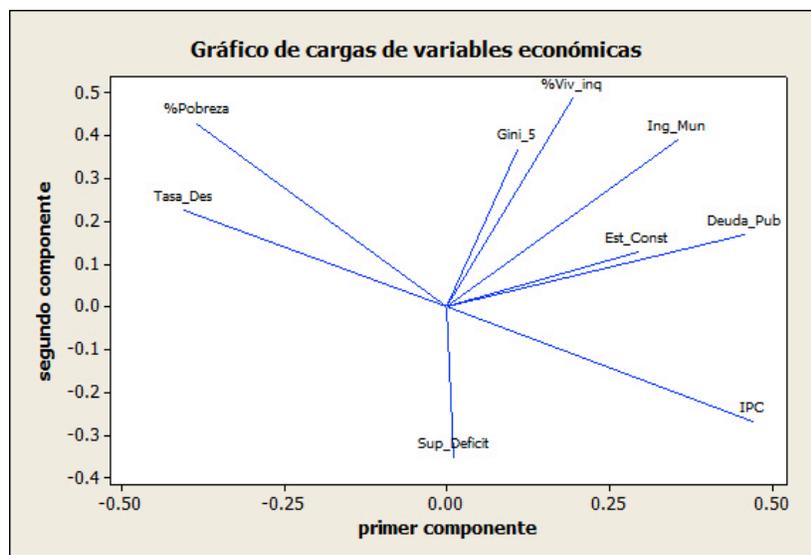


Figura 4.12: Análisis de componentes principales de variables económicas, gráfico de cargas.

Para rotular los componentes se analizó el tipo de variables y la relación de las variables con los componentes. Se rotularon de la siguiente manera:

- Capacidad de potencial económico: el primer componente principal para las variables económicas tiene valores negativos para las variables por ciento de población bajo el nivel de pobreza y tasa de desempleo. Los mayores valores positivos de este componente corresponden a las variables ingreso per cápita y deuda pública. Notemos que las variables con valor negativo corresponden a la capacidad productiva del municipio pues tiene que ver con la pobreza y la tasa de desempleo.
- Marginación socioeconómica: El segundo componente principal tiene valores negativos para las variables ingreso per cápita y superávit o déficit. Los valores de mayor magnitud son por ciento bajo el nivel de pobreza, por ciento de viviendas ocupadas por inquilinos e ingreso municipal.
- Distribución adecuada de los recursos: el tercer componente principal tiene valores negativos para la tasa de desempleo, ingreso municipal y deuda pública. Tiene un valor de casi cero en el ingreso per cápita y valores positivos significativos para superávit o déficit e Índice de Gini.

4.3. Índices

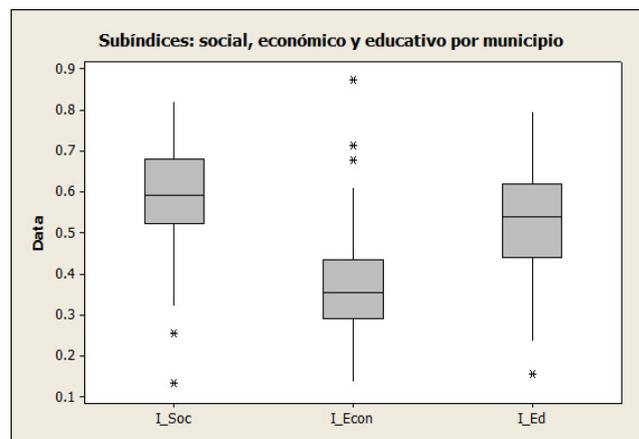
Las estadísticas descriptivas como la media, mediana, mínimo, máximo, desviación estándar y coeficiente de variación de cada índice I_{Soc} , I_{Ed} , I_{Ec} se presentan en el Cuadro 4.7

En la Figura 4.13 mostramos el diagama de caja de los índices I_{Soc} , I_{Ed} , I_{Ec} . Hay dos municipios *outliers* inferiores en el índice social: Vieques y Loíza. Mientras que Guaynabo, San Juan y Carolina son *outliers* superiores del índice económico y Culebra

Variable	Media	DesvStd	CoefVar	Min	Mediana	Max
I_{Soc}	0.5940	0.1237	20.82	0.1339	0.5915	0.8201
I_{Econ}	0.3737	0.1247	33.38	0.1389	0.3563	0.8727
I_{Ed}	0.5341	0.1374	25.73	0.1569	0.5386	0.7968

Cuadro 4.7: Medidas de tendencia central y variabilidad de I_{Soc} , I_{Ed} , I_{Ec}

es un *outlier* inferior del índice educativo.

Figura 4.13: Diagrama de caja de los índices I_{Soc} , I_{Ed} , I_{Ec}

En la Figura 4.14 se presenta el gráfico de puntos para cada uno de los índices I_{Soc} , I_{Ed} , I_{Ec} . En este notamos que el índice económico tienen aproximadamente ocho municipios mayores de 0.50 y estos están dispersos. Mientras que en los índices sociales y educativos hay menos dispersión de los datos.

4.3.1. Índice social

Calculamos el índice social de la siguiente manera: comenzamos con calcular un valor preliminar A_{Soc} del aspecto social en el cual se incluyen los valores de $PC1_{Soc}$,

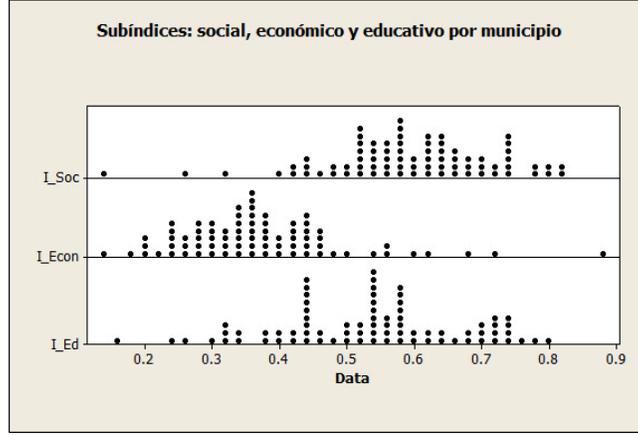


Figura 4.14: Gráfico de puntos de índices I_{Soc} , I_{Ed} , I_{Ec} . Cada punto representa un municipio.

$PC2_{Soc}$ y $PC3_{Soc}$ con sus respectivos pesos. Tomando en cuenta la variabilidad [4], los pesos de estos componentes fueron la proporción de la variabilidad explicada por el componente entre la variabilidad total explicada por los tres componentes en conjunto. Para el caso de los componentes sociales, los pesos fueron los siguientes: $w_{1Soc} = 0.47$, $w_{2Soc} = 0.40$ y $w_{3Soc} = 0.13$.

$$A_{Soc} = -w_{1Soc}PC1_{Soc} - w_{2Soc}PC2_{Soc} + w_{3Soc}PC3_{Soc} \quad (4.1)$$

El signo que lleva cada término, lo decidimos según cada componente afecta la calidad social del municipio, de modo que un valor alto de A_{Soc} corresponda a una mayor calidad social. El sistema de crianza sustituto representado por la variable $PC1$, fue colocado con un valor negativo por la alta cantidad de Delitos y Asesinatos que representa. Apatía hacia la tercera edad representado por $PC2_{Soc}$ fue colocado con un signo negativo, por la baja cantidad de establecimientos dedicados a la salud y un bajo Índice de Envejecimiento. Mientras que percepción de bienestar social se presenta con

un signo positivo para el índice social, debido a que hay poca cantidad de delitos en el municipio para este componente y la población tiene una percepción de bienestar mayor. El cálculo de A_{Soc} es indispensable para el cálculo de nuestro índice social. El cual es un valor entre 0 y 1 y se calculó en base a un valor ínfimo, que lo definimos como la peor situación que puede tener un municipio en el valor de cada uno de los componentes principales sociales multiplicado por el peso que le corresponde al componente. Por ejemplo, la peor situación en el primer componente principal es que el municipio obtenga el valor máximo de esta multiplicación, que es $w_{1Soc}PC1_{Soc} = 2.2725$. En el segundo componente tendría el peor escenario también con el valor máximo obtenido $w_{2Soc}PC2_{Soc} = 1.9667$, mientras que en el tercer componente ocurrió el peor escenario con el valor del componente mínimo $w_{3Soc}PC3_{Soc} = -0.4320$. Esto resulta en un valor ínfimo de:

$$Inf_{Soc} = -(2.2725) - (1.9667) + (-0.4320) = -4.6712 \quad (4.2)$$

Mientras que el valor del supremo es la situación óptima de cada uno de los componentes, por lo que se tomaron los valores mínimos en los primeros dos componentes y el valor máximo en el tercer componente. En esta situación obtuvimos los siguientes valores: -1.4010, -1.5305 y 0.2611. Entonces, el valor del supremo es:

$$Sup_{Soc} = -(-1.4010) - (-1.5305) + (0.2611) = 3.1926 \quad (4.3)$$

Luego calculamos el índice utilizando la siguiente Ecuación [2.2]:

$$I_{Soc} = \frac{A_{Soc} - Inf_{Soc}}{Sup_{Soc} - Inf_{Soc}} \quad (4.4)$$

El índice social tiene un comportamiento homogéneo y tiene una media de 0.59, la cual es la más alta entre los índices I_{Soc} , I_{Ed} , I_{Ec} . El valor mínimo corresponde al municipio

de Loíza con un valor *outlier* inferior, y le sigue el municipio de Vieques. Estos valores son *outliers* inferiores de este índice como se presenta en el diagrama de caja de la Figura 4.13. Los valores mayores que se ven en el gráfico de puntos en la Figura 4.14, corresponden a los municipios de Moca, Las Marías, Aguada y San Sebastián.

Debemos señalar que este índice se ve afectado positivamente por el por ciento de población menor de 18 años que vive con una pareja a cargo, el Índice de Envejecimiento y la cantidad de establecimientos de salud. Mientras que el índice se ve afectado de manera negativa por variables como el por ciento de población menor de 18 años que vive con una mujer a cargo y la cantidad de delitos tipo I. Lo que indica que un municipio con valores altos en este índice social, tiene una tendencia a tener mayor cantidad de familias tradicionales, poca criminalidad, establecimientos de salud accesibles, entre otros. Mientras que hay que atender las necesidades de los municipios con un bajo índice social, entre estos, Loíza, Vieques, Cataño y Santa Isabel.

4.3.2. Índice educativo

El índice educativo, similar al social, fue calculado a partir del valor preliminar A_{Ed} , que incluye los valores de $PC1_{Ed}$, $PC2_{Ed}$ y $PC3_{Ed}$ con sus respectivos pesos. Los pesos fueron asignados tomando en cuenta la proporción de variabilidad explicada por cada componente [4], estos son: $w_{1Ed} = 0.53$, $w_{2Ed} = 0.26$ y $w_{3Ed} = 0.21$.

$$A_{Ed} = w_{1Ed}PC1_{Ed} - w_{2Ed}PC2_{Ed} - w_{3Ed}PC3_{Ed} \quad (4.5)$$

El componente acceso a educación masiva representado por $PC1_{Ed}$, tiene un peso positivo, pues esto afecta la educación de Puerto Rico positivamente al tener una alta participación de estudiantes en las escuelas y un alto por ciento de personas graduadas

de escuela superior mayores de 25 años. Los componentes $PC2_{Ed}$: pobre desempeño académico y $PC3_{Ed}$: presión del entorno tienen un efecto negativo en el índice educativo de un municipio. El cálculo de A_{Ed} para el caso específico de las variables educativas tiene los valores del ínfimo para cada componente del mínimo de $w_{1Ed}PC1_{Ed}$ y los valores máximos de $w_{2Ed}PC2_{Ed}$ y $w_{3Ed}PC3_{Ed}$. Y lo contrario para el cálculo del supremo.

$$Inf_{Ed} = (-2.1528) - (0.8135) - (0.5215) = -3.4878 \quad (4.6)$$

$$Sup_{Ed} = (1.9635) - (-0.6014) - (-0.4780) = 3.0429 \quad (4.7)$$

Luego calculamos el índice utilizando la Ecuación [2.2]:

$$I_{Ed} = \frac{A_{Ed} - Inf_{Ed}}{Sup_{Ed} - Inf_{Ed}} \quad (4.8)$$

El índice educativo tiene un comportamiento homogéneo a través de los municipios de Puerto Rico. En el Cuadro 4.7, este índice tiene una media de 0.53 y un coeficiente de variación de 25.73%. El valor máximo de esta variable corresponde a Fajardo y el mínimo, quien a su vez corresponde a un valor *outlier*, es el municipio de Culebra. Este índice tiene un mayor peso en el primer componente: acceso a la educación masiva, por lo que el índice está midiendo un componente mayormente enfocado al acceso de la educación; la calidad educativa como tal tiene un peso menor en el índice. Volvemos a señalar la importancia de haber considerado variables como el nivel de deserción escolar y la ausencia de los maestros, las cuales no estuvieron disponibles.

Este índice tiene un *outlier* inferior para el municipio de Culebras, presentado en la Figura 4.13. Los tres municipios con valores más altos para este índice son: Fajardo,

Loíza y Carolina, los cuales se presentan en la Figura 4.14. Mientras que los valores más bajos son: Culebra, Las Marías, Maricao y Jayuya.

4.3.3. Índice económico

El índice económico, similar al índice social y el educativo, fue calculado a partir del valor preliminar A_{Econ} , que incluye los valores de $PC1_{Econ}$, $PC2_{Econ}$ y $PC3_{Econ}$ con sus respectivos pesos. Los pesos de estos componentes, fueron obtenidos según la variabilidad de los componentes [4]. Se calcularon según la proporción de la variabilidad representada entre la variabilidad total explicada, en los tres componentes son: $w_{1Econ} = 0.52, w_{2Econ} = 0.28$ y $w_{3Econ} = 0.21$.

$$A_{Econ} = w_{1Econ}PC1_{Econ} - w_{2Econ}PC2_{Econ} + w_{3Econ}PC3_{Econ} \quad (4.9)$$

“Capacidad de potencial económico” representado por la variable $PC1_{Econ}$, fue colocado con un valor positivo para el índice económico. El segundo componente $PC2_{Econ}$ representa “Marginación socioeconómica” y tiene un efecto negativo en el índice económico de un municipio. Mientras que una “Distribución adecuada de los recursos” tiene un signo positivo, pues este componente tiene valores negativos para la tasa de desempleo y deuda pública. El cálculo de A_{Econ} depende de un valor ínfimo y uno supremo, que los definimos como la peor y mejor situación, respectivamente, que puede tener un municipio en el valor de cada uno de los componentes principales económicos. El peor escenario para las variables económicas, toma los valores mínimos de $w_{1Econ}PC1_{Econ}$ y $w_{3Econ}PC3_{Econ}$, mientras que se utiliza el valor máximo de $w_{2Econ}PC2_{Econ}$. En este caso, el ínfimo está dado:

$$Inf_{Econ} = (-1.2250) - (1.0634) + (-0.8711) = -3.1595 \quad (4.10)$$

Mientras que el valor del supremo es la situación óptima de cada uno de los componentes, por lo que se tomaron los valores máximos en el primer y tercer componente y el valor mínimo en el segundo componente. En esta situación obtuvimos el valor del supremo:

$$Sup_{Econ} = (4.001) - (-0.7336) + (0.5611) = 5.2957 \quad (4.11)$$

Luego calculamos el índice utilizando la Ecuación [2.2]:

$$I_{Econ} = \frac{A_{Econ} - Inf_{Econ}}{Sup_{Econ} - Inf_{Econ}} \quad (4.12)$$

En el índice económico hay mucha variabilidad y su media es de 0.37, relativamente inferior comparado con el valor del índice social y el educativo. Lo que puede ser un reflejo de la situación económica de Puerto Rico. El valor mínimo para este índice es Maricao, le siguen Salinas, Guánica, Ciales y Maunabo. Mientras que los tres valores más altos: Guaynabo, San Juan y Carolina, a su vez representan *outliers*. Siendo Guaynabo un municipio con un valor extremadamente alto en el índice económico.

Notemos que en el índice económico hay mayor peso en el primer componente que tiene tendencias a mayor cantidad de ingresos per cápita, poco desempleo y bajo nivel de pobreza. Éstas son características de municipios que mayormente se encuentran en el área metro. Mientras que valores bajos de esta variable deben ser atendidos por las agencias pertinentes para mejorar las características de desempleo, bajo nivel de pobreza e ingreso per cápita, entre otros.

4.3.4. Índice de calidad de vida

El índice de calidad de vida (ICV) fue calculado en base a los índices: social, educativo y económico. En este caso, ponderamos las variables de manera equiproporcional,

al igual que lo hace el Índice de Desarrollo Humano. Esto funciona en nuestro caso, siguiendo la recomendación de Babbie [4] de tomar un peso igual de los aspectos y tomando en consideración que nuestras variables son igualmente prioritarias.

Se calculó de la siguiente forma:

$$I = \frac{I_{Soc} + I_{Econ} + I_{Ed}}{3} \quad (4.13)$$

La estadística descriptiva resultante de este índice se presentan en el Cuadro 4.8. El índice de calidad de vida tiene un promedio de 0.50058 con un comportamiento homogéneo.

Variable	Media	DesvStd	CoefVar	Min	Mediana	Max
ICV	0.50058	0.07890	15.76	0.29196	0.50203	0.73673

Cuadro 4.8: Medidas de tendencia central y variabilidad de I

En la Figura 4.15 se presentan los gráficos correspondientes al índice de calidad de vida de cada municipio.

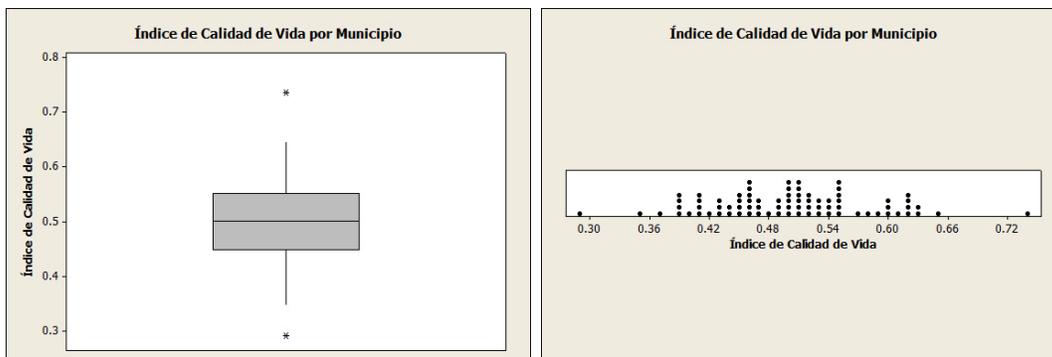


Figura 4.15: Gráficas del índice de calidad de vida

El índice de calidad de vida tiene un *outlier* superior que es el municipio de Guaynabo, y un *outlier* inferior que es el municipio de Vieques como vemos en la Figura 4.15.

Los Municipios con índices menores son: Vieques, Salinas, Maricao y Ciales. Mientras que los valores más altos del índice de calidad de vida corresponden a los municipios de Guaynabo, Carolina, Gurabo, Trujillo Alto, Rincón y Caguas según se presentan en la Figura 4.15.

Para la categorización del índice de calidad de vida, tomamos la media de este índice, 0.50058, y la desviación estándar, 0.07890. Las tres categorías usadas, al igual que en el Índice de Desarrollo Humano, son: alto, medio y bajo. Los índices a una desviación estándar de la media, se categorizaron como medios, es decir entre 0.42168 y 0.57948. Los índices con valores menores a 0.42168 se categorizaron como índices bajos y los valores mayores a 0.57948 se categorizaron como altos.

En la Figura 4.16, los municipios en color azul, representan los municipios con un índice de calidad de vida alto, estos municipios generalmente están en el área metro y algunos en el área oeste como Cabo Rojo, Hormigueros, Rincón y Aguadilla. Los municipios en color amarillo representan un valor bajo en el índice de calidad de vida, entre estos están algunos municipios del área central, además de las islas municipios y otros pueblos costeros como Guánica, Loíza, Salinas, Maunabo y Yabucoa. Mientras que la mayoría de los municipios, en color verde, tienen un índice de calidad de vida medio.

En la Figura 4.16 se muestran la categorización de los municipios según el índice de calidad de vida obtenido.

Al final de este capítulo se presentan los Cuadros: 4.12, 4.13 y 4.14 están los índices de todos los municipios en orden alfabético. En el Cuadro 4.12 se muestran los índices de calidad de vida, social, educativo y económico desde el municipio de Adjuntas hasta

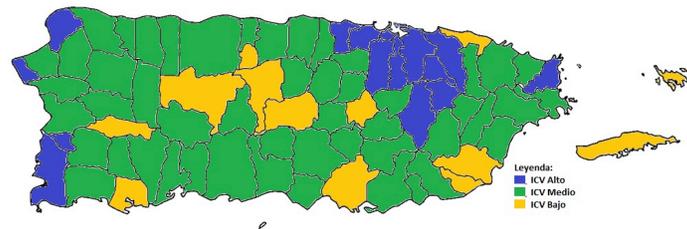


Figura 4.16: Mapa de la categorización de ICV Alto, Mediano o Bajo

el municipio de Florida. El Cuadro 4.13 muestra los índices desde el municipio de Guánica hasta el municipio de Orocovis. Y el Cuadro 4.14 se muestran los índices de calidad de vida, social, educativo y económico desde el municipio de Patillas hasta el municipio de Yauco.

4.4. Análisis de clusters

Para determinar el número adecuado de *clusters* para cada una de las clasificaciones, se utilizó el dendograma junto al ancho de silueta. El dendograma tiene el objetivo de ver características en los datos, como ayuda para elegir el número de *clusters*, pero no para tomar la clasificación definitiva. Por su parte, el ancho de silueta ayuda a seleccionar el número de *clusters* utilizando la clasificación que tenga mayor ancho de silueta. Utilizaremos el método de partición, aplicando el PAM. Los *clusters* de cada uno de los aspectos se analizaron utilizando los componentes principales de cada uno de los aspectos.

4.4.1. Clusters sociales

Para determinar la cantidad de *clusters* sociales analizamos las variables de los componentes principales sociales en un dendograma y vimos el Ancho de Silueta. En el Cuadro 4.9, se presenta un cambio mayor en el nivel de similaridad entre el quinto y cuarto *cluster*, pues sus valores cambian de 77.96 a 73.69. Lo que sugiere que se debe utilizar entre 4 o 5 *clusters* para el aspecto social.

Paso	Núm. <i>clusters</i>	Nivel de Similaridad	Distancia
...
68	10	86.8617	1.16138
69	9	86.6704	1.17828
70	8	84.5650	1.36439
71	7	79.3478	1.82557
72	6	78.0927	1.93652
73	5	77.9696	1.94740
74	4	73.6942	2.32533
75	3	68.6905	2.76764
76	2	68.5127	2.78335
77	1	63.9013	3.19098

Cuadro 4.9: Tabla de similaridad para los *clusters* sociales

El ancho de silueta para ambos, se presenta en la Figura 4.17. El promedio del ancho de silueta es bastante similar para ambas cantidades de *clusters*, en el ancho de silueta para 4 *clusters*, obtuvimos un promedio de 0.23 y para el de 5 *clusters* es de 0.24. Notemos también que el ancho de silueta para C_5 , el quinto *clusters*, es de -0.07 , valor muy bajo en comparación con los demás anchos de silueta. Por lo que decidimos tomar un número de 4 *clusters* para realizar nuestras agrupaciones.

En el Apéndice C.2 se presentan los intervalos de confianza para los promedios de

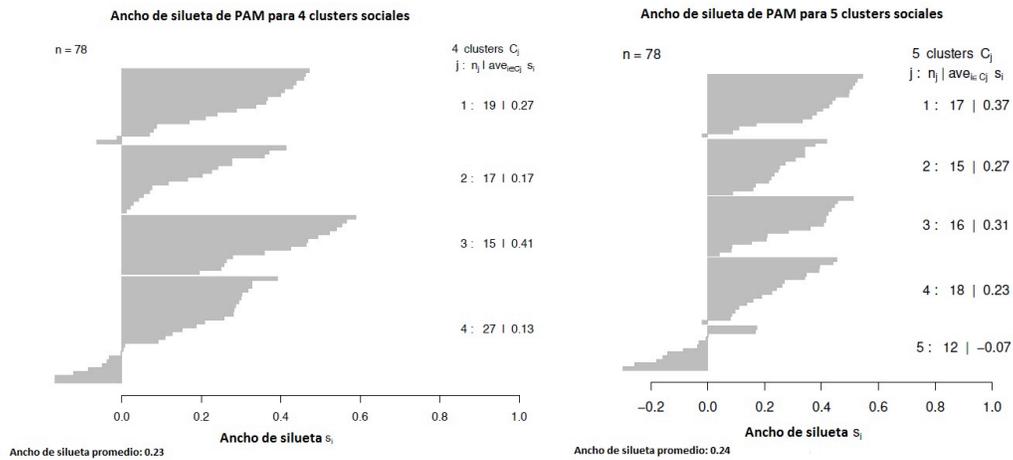


Figura 4.17: Ancho de Silueta en *clusters* sociales

los componentes principales en cada uno de los *clusters* sociales. Con respecto al primer componente: sistema de crianza sustituto, el intervalo de confianza para la media tiene valores altos en el segundo *cluster*, mientras que tiene valores bajos para el primer y el tercer *cluster*. Esto refleja una tendencia en el segundo *cluster* de municipios con mayor cantidad de sistemas de crianza sustitutos. Con respecto al segundo componente: apatía hacia la tercera edad, los municipios con valores altos tienen una tendencia a estar en los últimos dos *clusters*. Mientras que el tercer componente: percepción de bienestar social, tiene valores altos en el primer *cluster* y bajos en el cuarto *cluster*.

En términos de los *clusters* sociales, el primer *cluster* tiene un intervalo de confianza mayor al de los demás *clusters* para la media del componente 3, percepción del bienestar social. Y valores bajos para apatía hacia la tercera edad y sistema de crianza sustituto. Lo que presenta que en el primer *cluster* hay una mejor tendencia a la percepción del bienestar social.

La Figura 4.18 nos muestra que el primer *cluster* tiene valores altos del índice

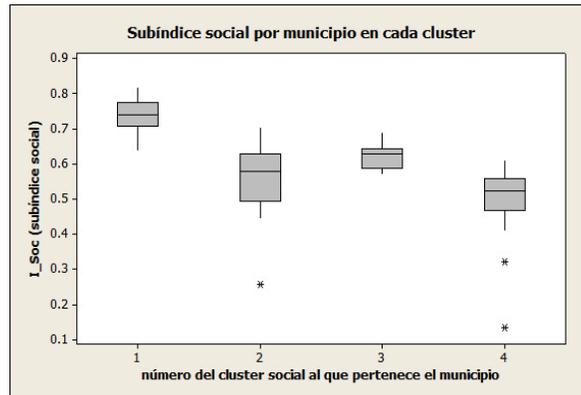


Figura 4.18: Diagramas de caja I_{Soc} en *clusters* sociales

social. En orden descendente del índice social en cada *cluster* tenemos: primero, tercero, segundo y cuarto *cluster*. De esta manera pintamos el mapa presentado en la Figura 4.19.

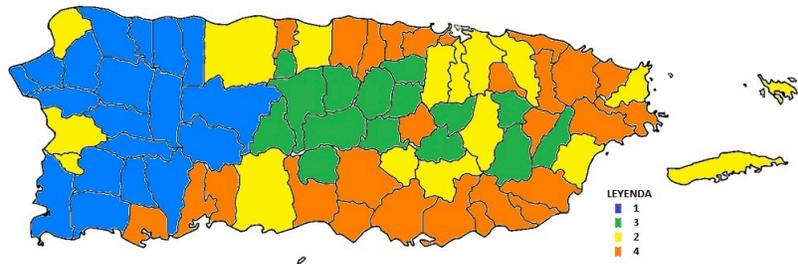


Figura 4.19: Mapa de *clusters* sociales. Los números 1, 3, 2 y 4 en la leyenda corresponden al número del *cluster* al cual pertenece el municipio.

La Figura 4.19 nos muestra que en el área oeste hay una tendencia a tener un índice social alto, característico del primer *cluster*. Le siguen los pueblos del área central pintados en color verde, pertenecientes al *cluster* 3. Notemos que entre los pueblos en color amarillo, están algunas de las ciudades más importantes de Puerto Rico, como por ejemplo: Mayagüez, Ponce, Arecibo y San Juan. Mientras que el *cluster* en color

anaranjado, corresponde mayormente a los municipios del área Sureste, Este y Norte de Puerto Rico.

4.4.2. Clusters económicos

Para el análisis de *clusters* económicos utilizamos los componentes principales económicos en un dendograma y luego vimos su ancho de silueta. La cantidad posible de *clusters* para las variables económicas, según el Cuadro 4.10 fue de 4 o 3 *clusters*, pues tuvo un cambio en el nivel de similaridad de 72.4 a 67.8. Este fue el primer cambio notable en el nivel de similaridad registrado.

Paso	Núm. <i>clusters</i>	Nivel de Similaridad	Distancia
...
68	10	86.9262	1.35104
69	9	86.5154	1.39349
70	8	83.1422	1.74208
71	7	79.5815	2.11003
72	6	77.0528	2.37135
73	5	73.8763	2.69961
74	4	72.4024	2.85192
75	3	67.7796	3.32963
76	2	60.1873	4.11422
77	1	59.9307	4.14074

Cuadro 4.10: Tabla de similaridad para los *clusters* económicos

Notemos que el promedio en los anchos de silueta en la Figura 4.20 es mayor para la cantidad de 4 *clusters*, por lo que en este caso seleccionamos este número de *clusters* para el aspecto económico.

En los intervalos de confianza del Apéndice C.4 para las medias de los componen-

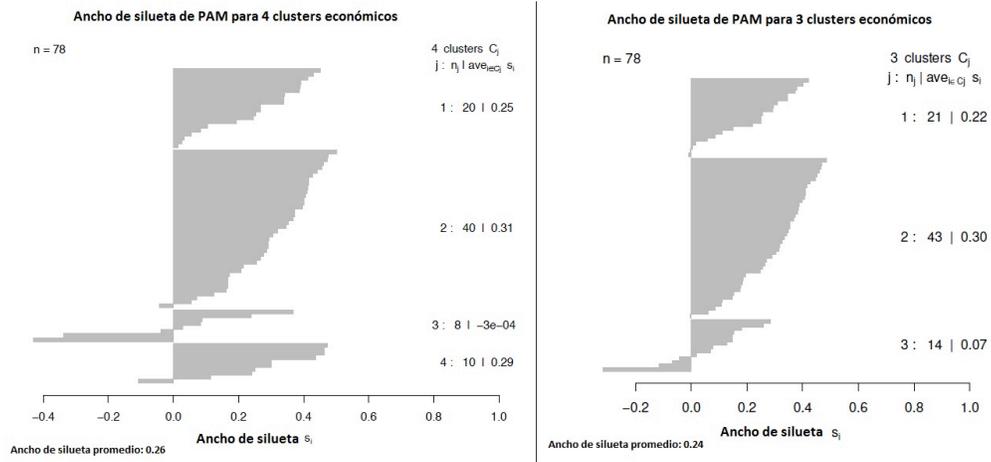


Figura 4.20: Ancho de Silueta en *clusters* económicos

tes principales en el primer *cluster* económico, vemos que tienen valores bajos para la media del primer componente: capacidad de potencial económico, y altos para el segundo componente: marginación socioeconómica. Lo que coloca al *cluster* 1 en una posición de condiciones desfavorables con respecto al aspecto económico. De hecho, en la Figura 4.21, notamos que este *cluster* corresponde a los valores más bajos del índice económico. Le sigue el segundo cluster, con valores relativamente bajos para el intervalo de confianza de los primeros dos componentes principales: capacidad de potencial económico y marginación socioeconómica y un valor alto en el componente: distribución adecuada de los recursos. Esto nos indica que hay una tendencia en este *cluster* a agrupar municipios que distribuyen bien su dinero, aunque no tienen un gran potencial económico. En el *cluster* 3, podemos ver unas mejorías en la capacidad de potencial económico, pero cuentan con mayor marginación socioeconómica. El tercer componente, distribución adecuada de los recursos, tiene un intervalo de confianza para la media bastante amplio en este *cluster*. En el cuarto *cluster*, notamos que hay valores altos

para el intervalo de confianza de la media en el componente de capacidad de potencial económico y un valor bajo para la marginación social. El valor del componente de la distribución adecuada de los recursos está un poco bajo, aunque es más alto que el valor del intervalo de confianza para la media del segundo componente.

En la Figura 4.21 notamos los diagramas de caja del índice económico en cada uno de los *clusters* que se obtuvieron del análisis de los componentes principales económicos.

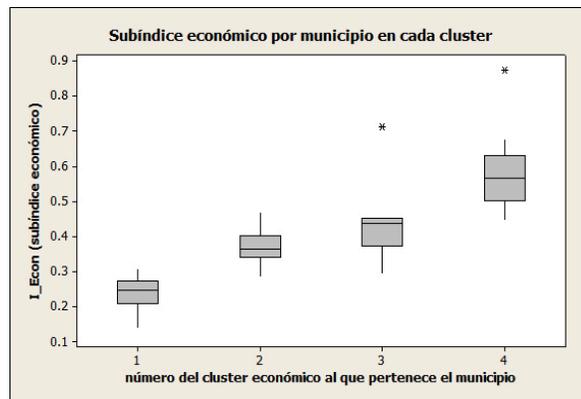


Figura 4.21: Diagramas de caja I_{Ec} en *clusters* económicos

Según el índice económico promedio por *cluster* calculado en la Sección 4.3.3, en orden descendente serían: el cuarto, tercero, segundo y primer *cluster*. En la Figura 4.22 vemos los municipios clasificados en los *clusters* económicos resultantes.

En la Figura 4.22 pintamos los *clusters* económicos en ese mismo orden. El color azul corresponde al *cluster* número cuatro, el verde al tres, el amarillo al dos y el anaranjado al uno. Notemos que el color azul corresponde, mayormente al área metro, y a un pequeño pueblo en el Oeste: Hormigueros. Mientras, que el verde corresponde a algunas de las principales ciudades como San Juan, Mayagüez y Ponce, curiosamente, también Vieques y Culebras están en este *cluster*. La mayoría de los pueblos están



Figura 4.22: Mapa de *clusters* económicos. Los números 4, 3, 2 y 1 en la leyenda corresponden al número del *cluster* al cual pertenece el municipio.

en color amarillo, mientras que hay algunos pueblos, generalmente en el área central, Sureste y Suroeste que representan el *cluster* con los índices económicos bajos.

4.4.3. Clusters educativos

Los *clusters* educativos fueron analizados según los componentes principales educativos en un dendograma, cuyos niveles de similaridad se presentan en el Cuadro 4.11. La cantidad de *clusters* sugeridos para las variables educativas fue de 5 o 6 *clusters*.

El ancho de silueta de la cantidad de 6 *clusters*, presentado en la Figura 4.23, tiene un ancho de silueta para C_4 de -0.01, y no hay mucha diferencia entre los promedios de los anchos de silueta para 5 o 6 *clusters*. Por lo que seleccionamos 5 *clusters* para el aspecto educativo

En la Figura 4.24 se presentan los diagramas de caja de los índices educativos en cada *cluster*. Notemos que el *cluster* 5 tiene el índice más alto.

En la Figura 4.25 se muestran los *clusters* correspondientes al aspecto educativo.

En el Apéndice C.3 presenta los intervalos de confianza para los componentes prin-

Paso	Núm. <i>clusters</i>	Nivel de Similitud	Distancia
...
68	10	83.9051	1.25751
69	9	83.8071	1.26517
70	8	83.2310	1.31017
71	7	83.1799	1.31416
72	6	82.6725	1.35381
73	5	80.1411	1.55159
74	4	80.0583	1.55806
75	3	80.0126	1.56163
76	2	79.9335	1.56781
77	1	77.6368	1.74725

Cuadro 4.11: Tabla de similaridad para los *clusters* educativos

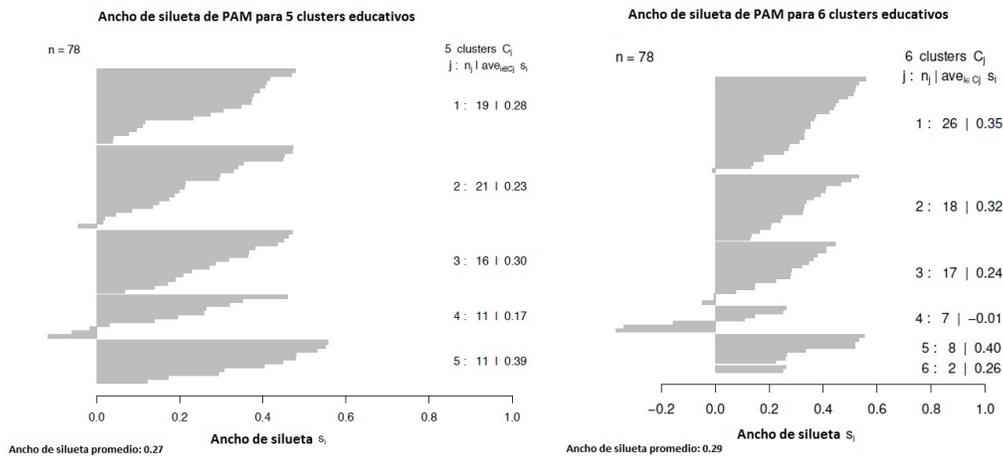


Figura 4.23: Ancho de Silueta en *clusters* educativos

cipales en cada uno de los *clusters* educativos. Notemos que el primer *cluster* tiene intervalos de confianza muy amplios para los tres componentes. Esto debido a que municipios que eran *outliers* en estos componentes fueron clasificados en dicho *cluster*. En el segundo *cluster* hay valores bajos para las medias del componente: acceso a la educación masiva y pobre desempeño escolar. Mientras que en el tercer *cluster* estos

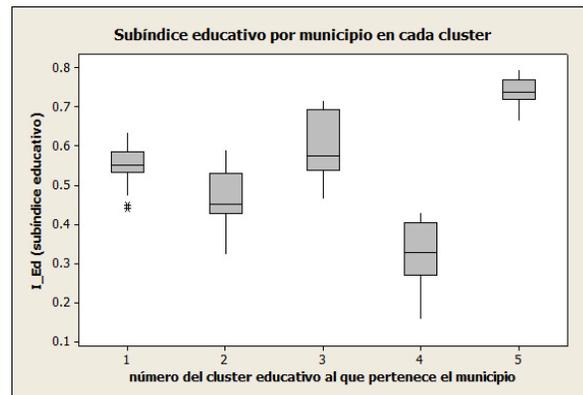


Figura 4.24: Diagramas de caja I_{Ed} en *clusters* educativos



Figura 4.25: Mapa de *clusters* educativos. Los números 5, 3, 1, 2 y 4 en la leyenda corresponden al número del *cluster* al cual pertenece el municipio.

dos componentes tienen valores. En el cuarto *cluster* hay intervalos de confianza con valores altos para la media de los componentes de pobre desempeño escolar y presión del entorno, mientras que en el quinto *cluster* ocurre lo contrario. El cuarto *cluster* corresponde a valores bajos en el índice educativo, como se puede ver en la Figura 4.24. Y el quinto *cluster* reúne los municipios con índices educativos más altos.

En orden del índice educativo en los *clusters* tenemos el quinto, tercero, primero, segundo y cuarto. En la Figura 4.25 los municipios pintados de color azul, representan el quinto *cluster* y tienen una tendencia a estar en el área Noreste. Mientras que los

verdes están en la Costa Oeste, algunos Municipios del área Metro y algunos otros del Sur y Este. Los municipios en color amarillo, corresponden mayormente al área Sur y Suroeste. Mientras que los anaranjados y rojos, se concentran en el área central y noroeste, incluye también algunos pueblos del Sureste y las islas municipio.

Municipio	ICV	I_{Soc}	I_{Ed}	I_{Econ}
Adjuntas	0.501	0.717	0.535	0.252
Aguada	0.552	0.801	0.493	0.361
Aguadilla	0.591	0.679	0.675	0.418
Aguas Buenas	0.458	0.618	0.431	0.325
Aibonito	0.474	0.616	0.447	0.359
Añasco	0.506	0.747	0.418	0.353
Arecibo	0.530	0.657	0.555	0.378
Arroyo	0.440	0.523	0.552	0.247
Barceloneta	0.502	0.514	0.550	0.442
Barranquitas	0.435	0.571	0.381	0.353
Bayamón	0.599	0.493	0.738	0.567
Cabo Rojo	0.581	0.744	0.538	0.459
Caguas	0.621	0.580	0.720	0.564
Camuy	0.511	0.747	0.439	0.348
Canóvanas	0.515	0.430	0.747	0.368
Carolina	0.647	0.495	0.768	0.677
Cataño	0.453	0.322	0.584	0.453
Cayey	0.516	0.567	0.523	0.457
Ceiba	0.552	0.612	0.663	0.382
Ciales	0.389	0.645	0.322	0.200
Cidra	0.542	0.644	0.559	0.425
Coamo	0.486	0.563	0.612	0.285
Comerío	0.391	0.527	0.402	0.243
Corozal	0.505	0.623	0.548	0.344
Culebra	0.394	0.601	0.157	0.424
Dorado	0.623	0.541	0.717	0.613
Fajardo	0.604	0.548	0.797	0.467
Florida	0.414	0.577	0.400	0.265

Cuadro 4.12: Índices para los municipios de Puerto Rico (Adjuntas a Florida)

Guánica	0.414	0.603	0.442	0.197
Guayama	0.456	0.483	0.536	0.349
Guayanilla	0.449	0.545	0.545	0.258
Guaynabo	0.737	0.641	0.697	0.873
Gurabo	0.631	0.634	0.727	0.533
Hatillo	0.527	0.723	0.524	0.334
Hormigueros	0.596	0.706	0.573	0.507
Humacao	0.520	0.533	0.579	0.449
Isabela	0.550	0.740	0.536	0.373
Jayuya	0.436	0.688	0.304	0.315
Juana Díaz	0.464	0.519	0.532	0.341
Juncos	0.542	0.559	0.646	0.422
Lajas	0.487	0.681	0.471	0.308
Lares	0.468	0.777	0.342	0.284
Las Marías	0.477	0.812	0.237	0.383
Las Piedras	0.518	0.585	0.591	0.380
Loíza	0.394	0.134	0.771	0.277
Luquillo	0.507	0.468	0.614	0.441
Manatí	0.458	0.580	0.439	0.355
Maricao	0.373	0.709	0.270	0.139
Maunabo	0.403	0.579	0.423	0.206
Mayagüez	0.537	0.590	0.572	0.450
Moca	0.566	0.820	0.539	0.338
Morovis	0.464	0.614	0.533	0.245
Naguabo	0.427	0.480	0.504	0.297
Naranjito	0.529	0.654	0.575	0.358
Orocovis	0.416	0.692	0.338	0.218

Cuadro 4.13: Índice para los municipios de Puerto Rico (Guánica a Orocovis)

Patillas	0.425	0.593	0.435	0.247
Peñuelas	0.450	0.515	0.523	0.312
Ponce	0.505	0.448	0.636	0.433
Quebradillas	0.499	0.738	0.450	0.309
Rincón	0.624	0.772	0.683	0.419
Río Grande	0.551	0.541	0.713	0.400
Sabana Grande	0.517	0.668	0.576	0.307
Salinas	0.347	0.423	0.449	0.170
San Germán	0.552	0.732	0.571	0.352
San Juan	0.620	0.444	0.705	0.712
San Lorenzo	0.497	0.589	0.501	0.401
San Sebastián	0.502	0.795	0.372	0.339
Santa Isabel	0.450	0.410	0.570	0.369
Toa Alta	0.608	0.629	0.748	0.447
Toa Baja	0.553	0.443	0.731	0.485
Trujillo Alto	0.625	0.570	0.697	0.610
Utuado	0.411	0.638	0.323	0.271
Vega Alta	0.506	0.524	0.593	0.401
Vega Baja	0.464	0.552	0.459	0.381
Vieques	0.292	0.256	0.327	0.293
Villalba	0.486	0.631	0.553	0.274
Yabucoa	0.407	0.521	0.463	0.236
Yauco	0.474	0.652	0.448	0.323

Cuadro 4.14: Índice para los municipios de Puerto Rico (Patillas a Yauco)

Capítulo 5

Conclusión y trabajos futuros

5.1. Conclusiones

Se creó un índice social cuya media fue de 0.59, el mismo presentó un comportamiento homogéneo a través de los municipios de Puerto Rico. Las peores condiciones sociales pertenecen al cuarto *cluster* social, cuyos municipios tienen las siguientes características: baja tendencia a familias tradicionales, mucha criminalidad, bajo Índice de Envejecimiento y baja cantidad de establecimientos dedicados a la salud. Este *cluster* incluye municipios como: Loíza, Cataño y Santa Isabel. Por otro lado, los municipios con las mejores condiciones sociales están en el primer *cluster* e incluye municipios con alto índice social, como: Moca, Las Marías, Aguada y San Sebastián. Generalmente los municipios en este *cluster* están en el área Oeste y se distinguen por tener poca criminalidad, una buena atención de la salud y sistemas de crianza tradicionales.

El índice educativo, por su parte, tiene una media de 0.53 y un comportamien-

to homogéneo para los municipios de Puerto Rico. Los municipios con mayor índice educativo pertenecen al quinto *cluster*, que se distingue por tener una alta tendencia a una educación masiva, bajo aprovechamiento académico, y una presión del entorno baja. Entre los municipios que pertenecen a este *cluster* que a su vez, obtuvieron un alto índice educativo, están los municipios de: Fajardo, Loíza y Carolina. Mientras que los municipios del cuarto *cluster* tienen baja educación masiva, pero un alto aprovechamiento académico. Esto refleja que se debe tener especial cuidado al utilizar este índice, pues tiene un gran peso en el componente: acceso a la educación masiva, y tener dicho acceso no conlleva necesariamente a calidad en la educación. Los municipios con menor índice educativo pertenecen al cuarto *cluster*, entre estos: Culebra, Las Marías, Maricao y Jayuya.

El índice económico tiene una media de 0.37 y mucha variabilidad. Los municipios con mayor índice económico son: Guaynabo y Carolina, quienes a su vez pertenecen al cuarto *cluster*. Este reúne mayormente los municipios del área metro que se caracterizan por una baja tasa de desempleo y bajos niveles de pobreza, además de un alto ingreso per cápita. Mientras que los de menor índice son: Maricao, Salinas, Guánica, Ciales y Maunabo que pertenecen al primer *cluster* y se caracterizan por tener bajo ingreso per cápita, altos niveles de pobreza y alta tasa de desempleo.

Acerca del índice de calidad de vida, vemos un comportamiento homogéneo y una media de 0.501. La mayoría de los municipios fueron categorizados con un índice medio. Geográficamente los municipios con mejores índices de calidad de vida están ubicados en el área metropolitana, quienes fueron los municipios con mayor índice económico. Los siguientes municipios con mayores índices de calidad de vida están ubicados en el Oeste, que a su vez fueron de los mayores índices sociales y pertenecían al segundo

mejor *cluster* educativo. Con respecto a los municipios con bajo índice de calidad de vida, notamos que la mayoría de ellos pertenecen a los dos *clusters* más bajos educativos y todos ellos pertenecen al *cluster* con índice económico y social menor.

En general, se identifican tendencias a un desarrollo económico alto en el área metro, y uno bajo en el área central. Lo que implica que el gobierno puede identificar iniciativas para fomentar la economía en áreas como éstas. En la parte social, los modelos de familias tradicionales resultaron ser beneficiosos para la sociedad. Se recomienda que el gobierno promueva las familias tradicionales y atienda las necesidades de las familias no tradicionales. En adición a esto, encontramos que la criminalidad está asociada a un bajo Índice de Envejecimiento, lo que podría reflejar problemas en la juventud.

5.2. Trabajos futuros

Durante esta investigación, surgieron algunas ideas de trabajos que pueden ser realizados en un futuro. También hubo algunos resultados que nos parecieron interesantes para ser analizados por expertos en las áreas sociales, educativas y económicas. Algunas de las investigaciones que se pueden realizar en un futuro son:

- Calcular los índices para datos actualizados.
- Utilizar variables como la deserción escolar y el ausentismo de maestros en las escuelas para calcular el índice educativo, pues para esta investigación no estuvieron disponibles.
- Analizar desde el punto de vista de las ciencias sociales los resultados del índice social en cada municipio.

- Analizar la situación económica de los municipios por expertos en el área. Uno de los resultados que puede ser interesante es el del municipio de Hormigueros.
- Analizar los aspectos educativos en este trabajo y los resultados obtenidos para mejorar el aprendizaje en los municipios.

Bibliografía

- [1] Edgar Acuna, *Unsupervised classification, clustering. ESMA 6835 Minería de Datos*, Universidad de Puerto Rico, RUM, Mayaguez, Puerto Rico, <http://www.math.uprm.edu/edgar/dm15.ppt>.
- [2] Ramón Álvarez and Andrés Castrillejo, *Un método de alternativo para la vigilancia epidemiológica*, Tech. report, Instituto de Estadística, 2012.
- [3] Rubén Ardila, *Calidad de vida: Una definición integradora*, Revista Latinoamericana de Psicología **35.2** (2003), 161 – 164, <http://redalyc.uaemex.mx/>.
- [4] Earl Babbie, *The Basics of Social Research*, 5th ed., Wadsworth, Cengage Learning, 2011.
- [5] *Carta Circular Núm. 1-2012-2013*, Departamento de Educación del Gobierno de Puerto Rico, 2012.
- [6] *Census QT-H3: Household population and household type by tenure*, U.S. Census Bureau, 2010, <http://factfinder2.census.gov>.
- [7] *Census QT-P12: Household Relationship and Group Quarters Population: 2010*, U.S. Census Bureau, 2010, <http://factfinder2.census.gov>.

- [8] Yovani Correa and Julio Quintana, *Perfil socioeconómico de los municipios de Puerto Rico: aplicación de técnicas estadísticas multivariadas*, Universidad de Puerto Rico RUM, 2011.
- [9] *County Business Patterns*, U.S. Census Bureau, 2010, <http://www.census.gov/econ/cbp/>.
- [10] *Delitos Tipo I informados en Puerto Rico por municipios del 1 de enero al 31 de diciembre de 2010*, Policía del Estado Libre Asociado de Puerto Rico, 2010, <http://www2.pr.gov/agencias/policiapr/Estadisticas/>.
- [11] *Census DP1: 2010 Demographic Profile Data*, U.S. Census Bureau, 2010, <http://factfinder2.census.gov>.
- [12] Luis G. Díaz, *Estadística Multivariada: inferencia y métodos*, 1ra ed., Universidad Nacional de Colombia, Bogotá, Colombia, 2002.
- [13] *Encuesta sobre la comunidad*, U.S. Census Bureau - American Community Survey Office, 2010, <http://www.estadisticas.gobierno.pr>.
- [14] Manuel S. Figueras, *Análisis de conglomerados o cluster*, Universidad de Zaragoza, España, Zaragoza, España, <http://www.5campus.org/leccion/cluster>.
- [15] *Gini Index of Income Inequality, 2006-2010 American Community Survey 5-year estimates*, U.S. Census Bureau, 2010, <http://factfinder2.census.gov>.
- [16] Javier A. Gutierrez, *Guía para diseño, construcción e interpretación de indicadores*, DIRPEN and DANE, Bogotá, Colombia, <http://www.dane.gov.co/files/planificacion/fortalecimiento/cuadernillo>.

- [17] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis, *On clustering validation techniques*, Journal of Intelligent Information Systems **17:2/3** (2001), 107–145, Kluwer Academic Publishers.
- [18] *Human Development Indicators: International Human Development Indicators-Income Gini Coefficient*, Programa de las Naciones Unidas para el Desarrollo - Organización de las Naciones Unidas, 2012, <http://hdrstats.undp.org/en/indicators/67106.html>.
- [19] *Human development report: Human development indicators*, Programa de las Naciones Unidas para el Desarrollo - Organización de las Naciones Unidas, 2004, <http://hdr.undp.org>.
- [20] *Indicadores socioeconómicos*, Oficina del Contralor del Estado Libre Asociado de Puerto Rico, 2012, <http://www.ocpr.gov.pr>.
- [21] *Informe del Perfil Escolar por escuela 2010-2011*, Departamento de Educación del Gobierno de Puerto Rico, 2012, <http://www.de.gobierno.pr/el-perfil-escolar>.
- [22] *LexJuris, Puerto Rico. Ley Núm. 165 del año 2011*, Estado Libre Asociado de Puerto Rico, 2011, <http://www.lexjuris.com/lexlex/Leyes2011/lexl2011165.htm>.
- [23] Anibal Llano and Hernando Moreno, *Metodología para la conformación de la línea de base del Programa BID-PLAN PACIFICO*, Departamento Nacional de Planeación de Colombia, Bogotá, 1995.
- [24] Mario Marazzi, *Inventario de Estadística*, Tech. report, Instituto de Estadísticas de Puerto Rico, Estado Libre Asociado de Puerto Rico, 2012, <http://www.estadisticas.gobierno.pr/iepr/Estadisticas/InventariodeEstadisticas.aspx>.

- [25] *P. de la C. 703 14 de enero de 2009. Texto de Aprobación Final por la Cámara del 22 de octubre de 2010*, Cámara de Representantes de Puerto Rico, 2010, [http://www.senadopr.us/Proyectos del Senado/pc0703-ta.pdf](http://www.senadopr.us/Proyectos%20del%20Senado/pc0703-ta.pdf).
- [26] Teresa Pedroso, *Índice de Desarrollo Humano en Puerto Rico, inicios del siglo XXI*, *Revista de Ciencias Sociales* **19** (2008), 150 – 177, <http://www.docstoc.com/docs/21477226/El-Índice-de-Desarrollo-Humano-en-Puerto-Rico-inicios>.
- [27] Daniel Pena, *Análisis de datos multivariantes*, McGrawHill Interamericana de España S.L., 2002.
- [28] Rigoberto Pérez, Covadonga Caso, Maria J. Río, and Ana J. López, *Introducción a la Estadística Económica*, España: Universidad de Oviedo, 2010, <http://books.google.com.pr>.
- [29] Luis F. Restrepo, Sandra L. Posada, and Ricardo R. Noguera, *Aplicación del análisis por componentes principales en la evaluación de tres variedades de pasto*, *Revista Colombiana de Ciencias Pecuarias* **25.2** (2012), 252–257, Universidad de Antioquia, Colombia.
- [30] Andrés Schuschny and Humberto Soto, *Guía metodológica diseño de indicadores compuestos de desarrollo sostenible*, Comisión Económica para América y el Caribe (CEPAL), Naciones Unidas, Santiago Chile, 2009, <http://www.eclac.cl/publicaciones/xml/7/36127/W255-2.pdf>.
- [31] Ernst Wit and John McClure, *Statistics for microarrays: Design, analysis, and inference*, Wiley & Sons Ltd, England, 2004.

Apéndice A

Gráficas de Análisis Descriptivo

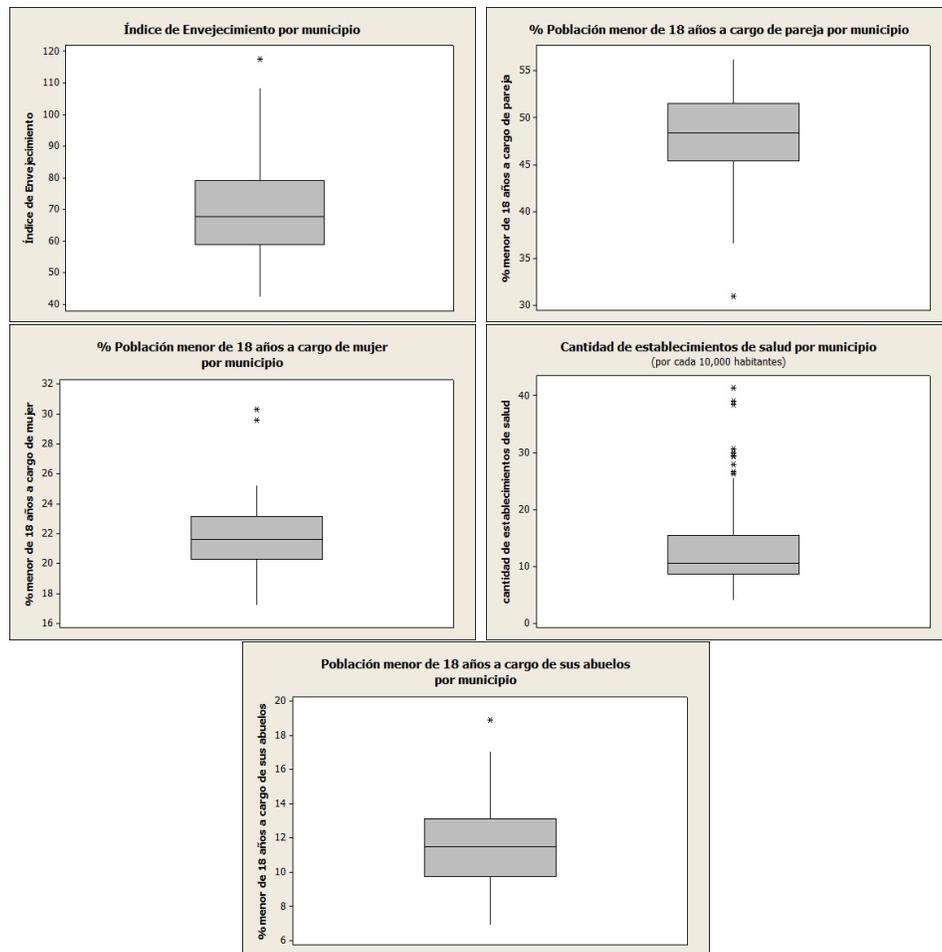


Figura A.1: Diagramas de caja de variables sociales



Figura A.2: Gráficos de puntos de variables sociales. Cada punto representa un municipio.

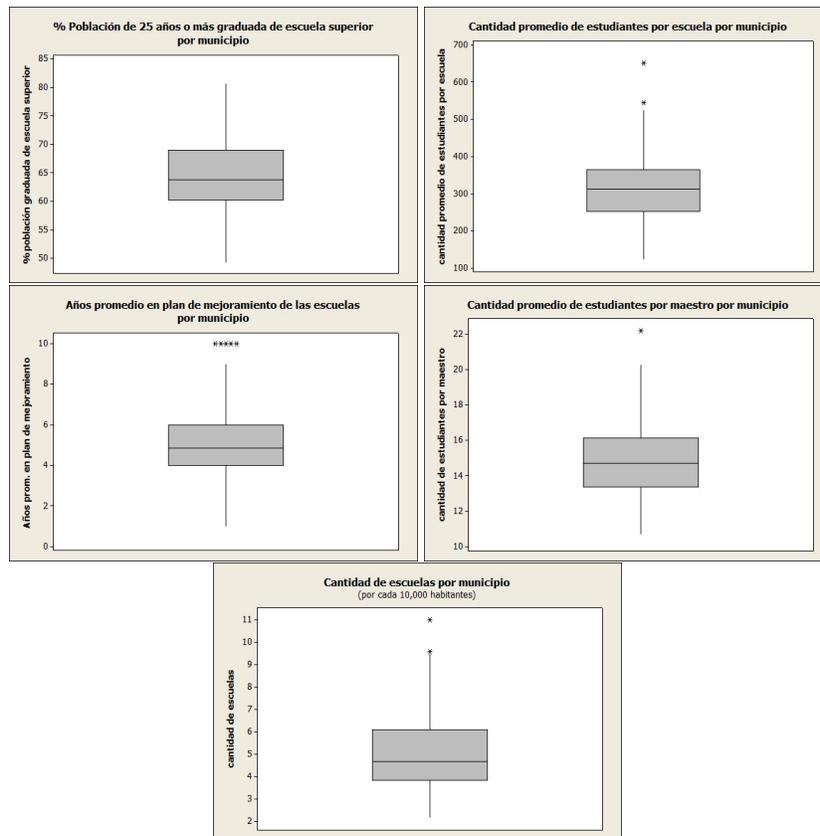


Figura A.3: Diagramas de caja de variables educativas

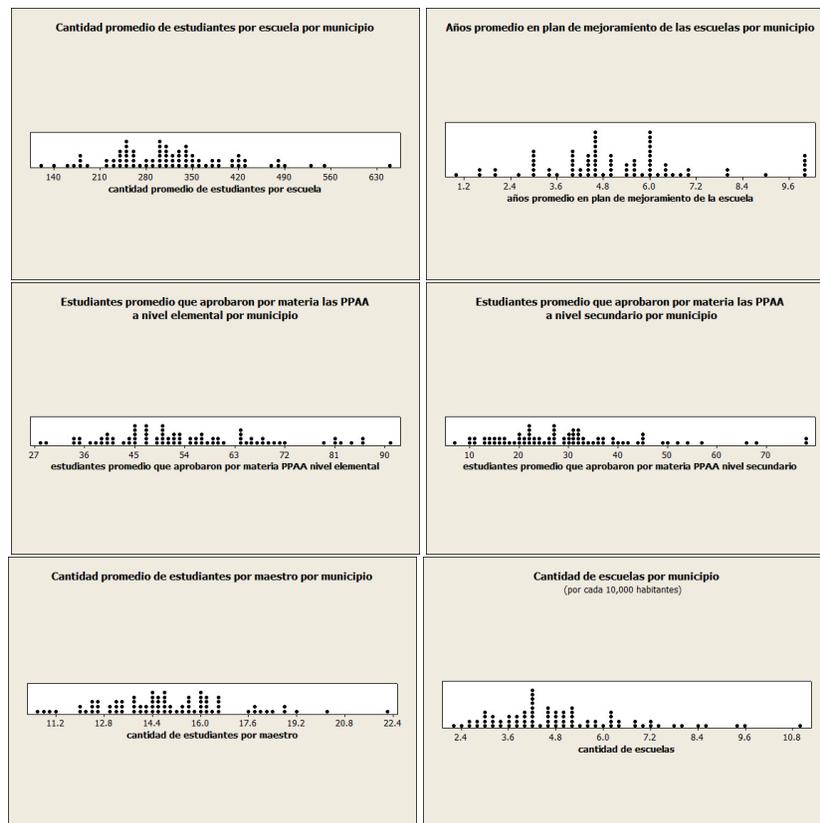


Figura A.4: Gráficos de puntos de variables educativas. Cada punto representa un municipio.

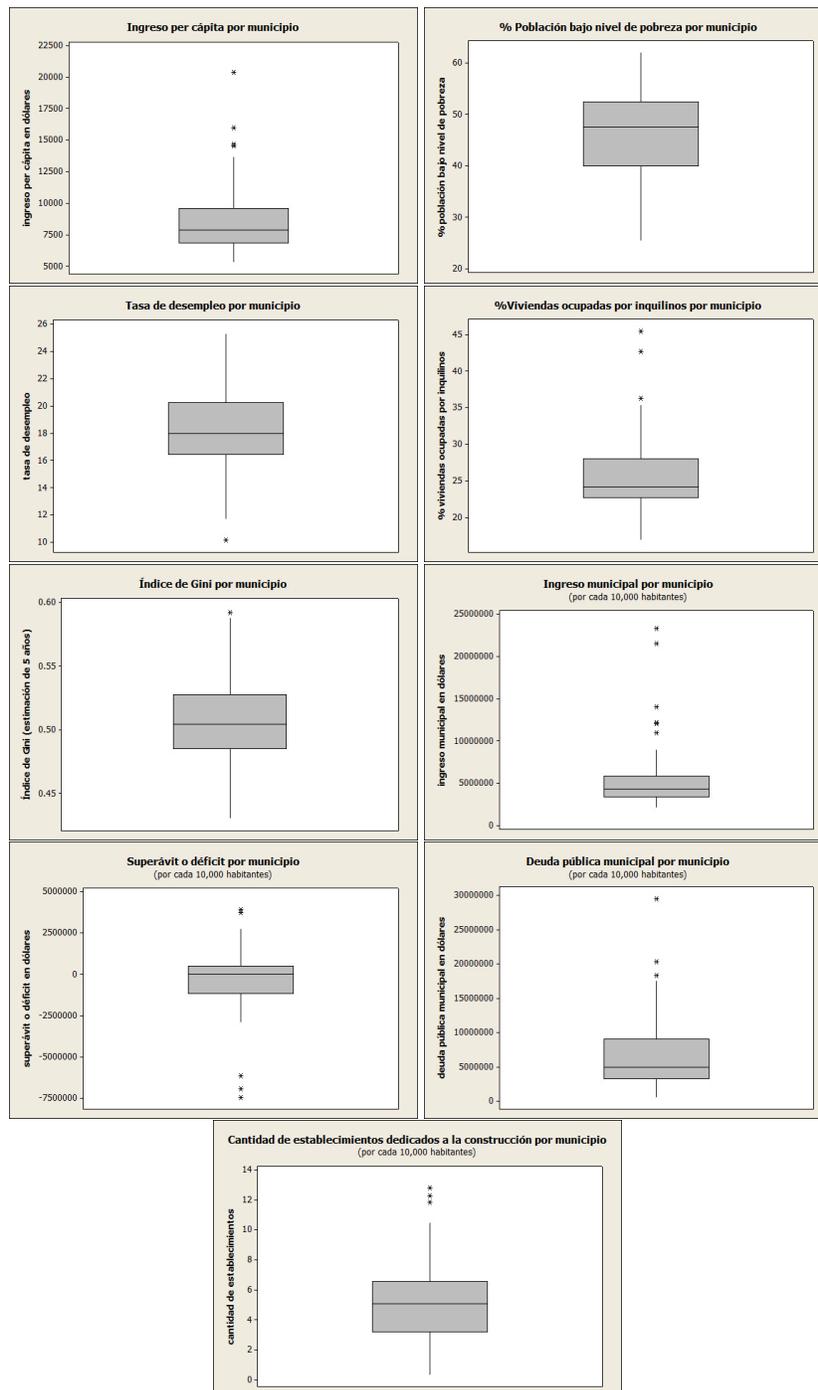


Figura A.5: Diagramas de caja de variables económicas.



Figura A.6: Gráficos de puntos de variables económicas. Cada punto representa un municipio.

Apéndice B

Gráficas relacionadas a índices

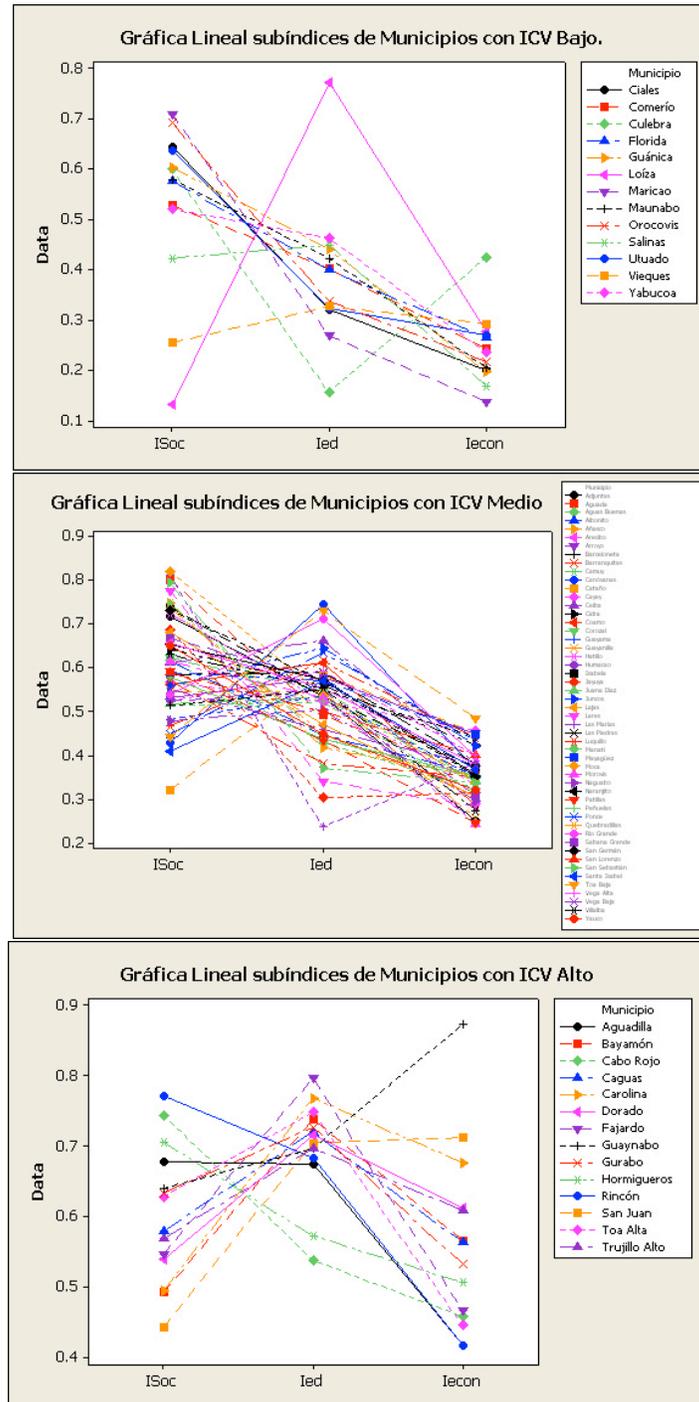


Figura B.1: Gráficas de líneas de los índices I_{Soc} , I_{Ed} , I_{Ec} en cada municipio

Apéndice C

Gráficas de análisis de *clusters*

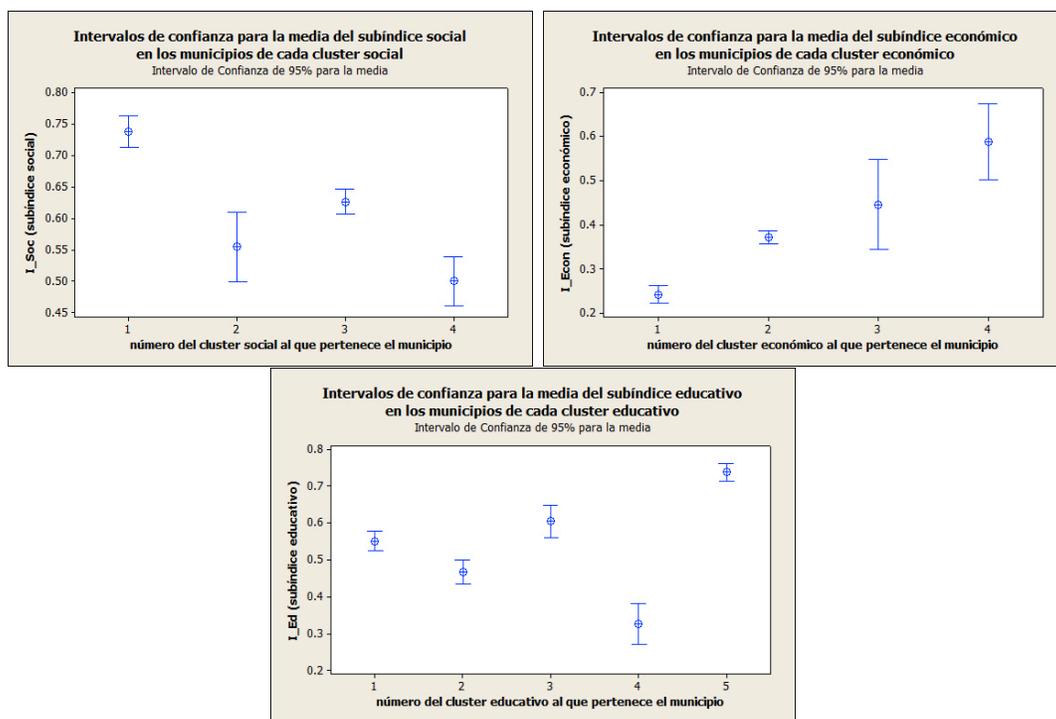


Figura C.1: Intervalos de confianza para la media de los índices de los municipios en sus *clusters*

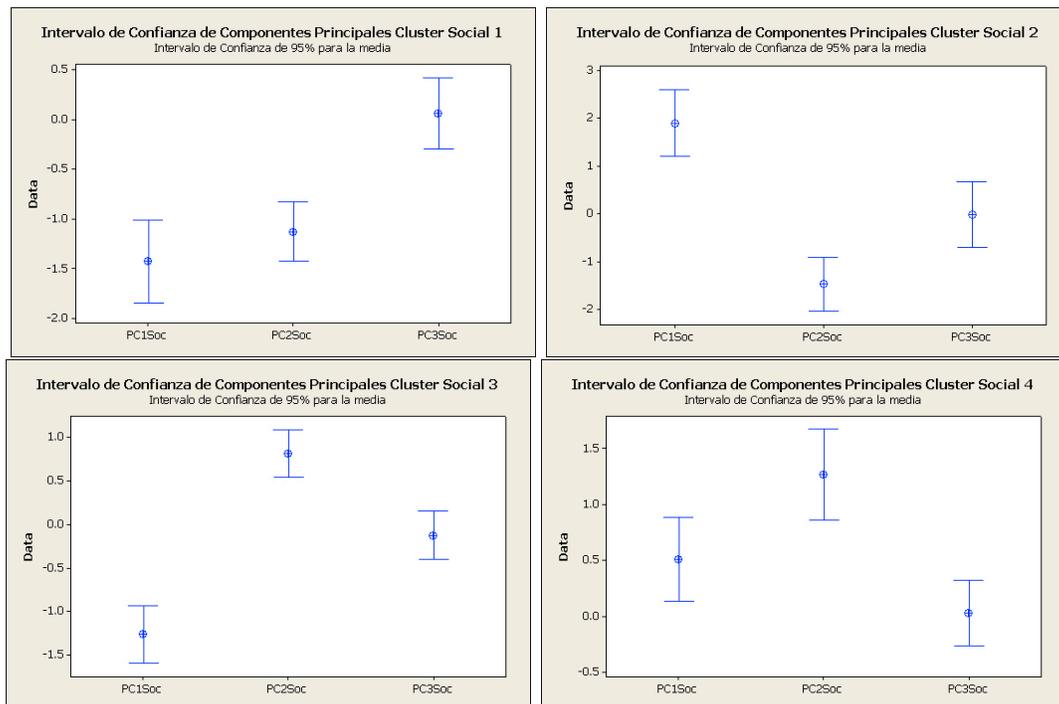


Figura C.2: Intervalos de confianza para la media de PC sociales

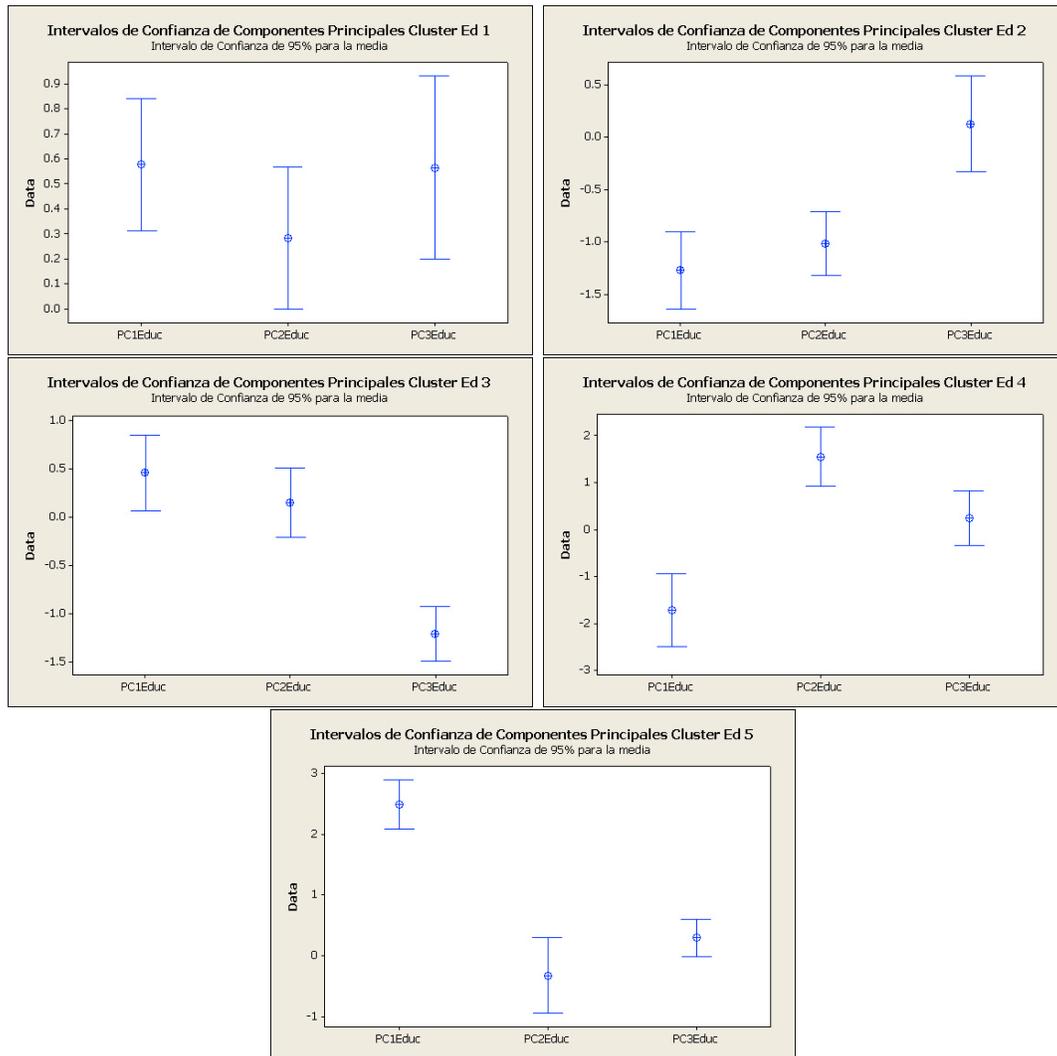


Figura C.3: Intervalos de confianza para la media de PC educativos

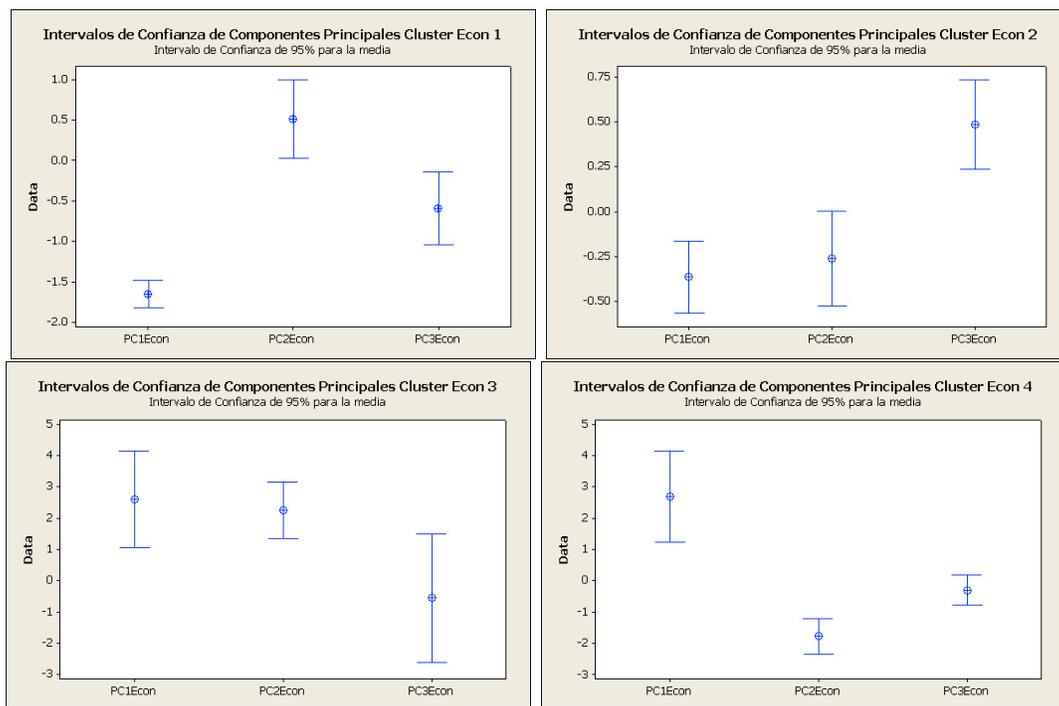


Figura C.4: Intervalos de confianza para la media de PC económicos

Apéndice D

Fuentes de las variables utilizadas

Índice de fuentes de las variables:

- CENSO DP1: Base de datos del Censo 2010 de Estados Unidos. Accedido desde <http://factfinder2.census.gov/>.
- CENSO QT-H3: Base de datos del Censo 2010 de Estados Unidos. Accedido desde <http://factfinder2.census.gov/>.
- CENSO QT-P12: Base de datos del Censo 2010 de Estados Unidos. Accedido desde <http://factfinder2.census.gov/>.
- County Business Patterns: Base de datos de la oficina del Censo. Accedido desde <http://www.census.gov/econ/cbp/index.html>
- Departamento de Policía: Datos del Departamento de Policía de Puerto Rico. Datos correspondientes al año 2010. Accedido desde <http://www2.pr.gov/agencias/policiapr/Estadisticas>

- Contralor: Datos del Departamento de Contralor de Puerto Rico para el año 2010. Accedido desde <http://www.ocpr.gov.pr/map/otros.htm>
- DEPR: Datos del Perfil Escolar por escuela del Departamento de Educación de Puerto Rico. Año escolar 2010-2011, también se usó la base de datos del año escolar 2011-2012, que incluye datos del 2010-2011.
- Encuesta sobre la comunidad: Datos obtenidos para estimar el Índice de Gini para muestras en los municipios calculado por el Puerto Rico Community Survey de la Oficina del Censo. Esta estimación se realiza para 5 años.