



Universidad Popular Autónoma del Estado de Puebla
Centro Interdisciplinario de Posgrados
Investigación y Consultoría
Departamento Económico-Administrativo
Doctorado en Planeación Estratégica y Dirección de Tecnología

Título de la investigación

Análisis geográfico y socioeconómico del agua en México:
Evaluación de políticas y estrategias basadas
en simulación e imputación de datos

Tesis que para obtener el Grado de Doctor
en Planeación Estratégica y Dirección de Tecnología

Presenta

Miguel Ángel Torres González



UPAEP – Secretaría General

Dirección General de Apoyos Académicos

Dirección del Centro de Recursos para el Aprendizaje y la Investigación.

Biblioteca Central - **Karol Wojtyla**

Tesis Digitales Restricciones de uso:

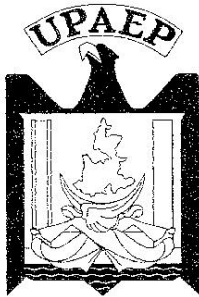
DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de textos, imágenes, gráficas, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente de donde la obtuvo mencionando el autor o autores involucrados en el documento.

Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



Universidad Popular Autónoma del Estado de Puebla

Centro Interdisciplinario de Posgrados

Investigación y Consultoría

Departamento de Ingeniería

Doctorado en Planeación Estratégica

y Dirección de Tecnología

Se aprueba la Tesis:

Análisis geográfico y socioeconómico del agua en México:

Evaluación de políticas y estrategias basadas

en simulación e imputación de datos

que presenta:

Miguel Ángel Torres González

Comité Asesor

Dr. Zeus Salvador Hernández Veleros

Director de Tesis

Dr. José Pablo Nuño de la Parra

Asesor

Dra. Beatriz Rico González

Asesora

Puebla, México.

30 de noviembre de 2012

DEDICATORIA

Dedico este proyecto de investigación a mi familia, ya que sin su apoyo, no hubiera sido posible realizarlo, especialmente a mi esposa Lupita, a mis hijas Mariana y Michelle, a mis padres Herminio y Guadalupe, a mis hermanos Julieta, Refugio, Sonia, Jaime y Lorena quienes siempre apoyaron los esfuerzos para terminar este trabajo.

ÍNDICE GENERAL

RESUMEN

ABSTRACT

INTRODUCCIÓN i

CAPÍTULO 1. PROPÓSITO Y ORGANIZACIÓN 1

1.1 Planteamiento del problema. 1

1.2 Propósito de la investigación. 10

1.3 Objetivo general. 10

1.4 Objetivos específicos. 11

1.5 Justificación de la investigación. 12

1.6 Alcances y limitaciones. 13

1.6.1 Alcances. 13

1.6.2 Limitaciones. 14

1.7 Organización del estudio. 14

CAPÍTULO 2. MARCO TEÓRICO 16

2.1 Modelos econométricos de elección múltiple. 16

2.1.1 Modelo de regresión logístico multinomial. 16

2.2	Simulación.	17
2.2.1	Introducción.	17
2.2.2	Simulación como técnica para solucionar problemas.	18
2.2.3	Modelos de simulación como herramienta de investigación.	19
2.2.4	Modelos de simulación como herramienta de entrenamiento y aprendizaje.	19
2.2.5	Modelos de simulación como herramienta de soporte a la decisión y Negociación.	20
2.3	Sistemas de información geográfica.	21
2.3.1.	Introducción a los Sistemas de Información Geográfica (SIG).	21
2.4	Componentes principales.	24
2.4.1.	El análisis de componentes principales.	24
2.4.2.	Los componentes principales.	25
2.4.2.1.	Formulación matemática del problema de componentes principales.	27
2.4.2.2.	Los resultados de los componentes principales y su Interpretación.	31
2.5	El modelo de imputación .	34
2.5.1.	El proceso de imputación de datos.	34
CAPÍTULO 3. METODOLOGÍA		37
3.1	Hipótesis.	37

3.2	Diseño de la investigación.	37
3.2.1.	Definición de variables.	39
3.3	Selección de la muestra.	40
3.4.	Recolección de datos.	40
3.4.1	Preparación de datos.	41
3.5	Aplicación del método.	42
3.5.1	El modelo de componentes principales.	42
3.5.1.1	Construcción de los indicadores del índice de marginación del agua y alcantarillado.	42
3.5.1.2	Porcentaje de viviendas sin servicio de agua potable.	43
3.5.1.3	Porcentaje de viviendas que se abastecen del servicio de agua por río o pozo.	43
3.5.1.4	Porcentaje de viviendas que cuentan con servicio de agua potable de vez en cuando.	44
3.5.1.5	Porcentaje de viviendas que cuentan con servicio de agua potable por medio de hidrante.	44
3.5.1.6	Porcentaje de viviendas que no cuentan con drenaje.	45
3.5.1.7	El índice de marginación del agua y alcantarillado.	45
3.5.1.8	El grado de marginación del agua y alcantarillado.	49
3.5.2	La georeferenciación del índice de marginación del agua y alcantarillado.	50
3.5.3	Aplicación del modelo multilogístico y simulación.	59

3.5.3.1 Aplicación del modelo multilogístico.	62
3.5.3.2 Escenarios para la simulación.	64
3.5.4 Aplicación al Estado de Oaxaca.	67
3.5.4.1 Aplicación del modelo de imputación de datos.	67
3.5.4.2 Análisis descriptivo de los ingresos imputados.	74
3.5.4.3 Aplicación de la metodología a los datos imputados.	76
 CAPITULO 4. ANÁLISIS DE RESULTADOS	 82
4.1 Conclusiones relativas a los objetivos específicos.	82
4.2 Conclusiones relativas al objetivo general.	83
4.3 Aportaciones originales.	84
4.4 Límites del modelo planteado.	85
 CAPITULO 5. RECOMENDACIONES	
5.1 Recomendaciones a los Organismos gestores del agua potable y alcantarillado.	86
5.2 Recomendaciones relativas a los resultados obtenidos	88
5.3 Recomendaciones generales.	89
 REFERENCIAS	
 ANEXOS	

ÍNDICE DE FIGURAS

Figura 1. Estructura de capas y estructuras de información de un sistema de información geográfico.	22
Figura 2. Porcentaje de viviendas sin servicio de agua potable.	43
Figura 3. Porcentaje de viviendas que se abastecen del servicio de agua por río o pozo.	43
Figura 4. Porcentaje de viviendas que cuentan con servicio de agua potable de vez en cuando.	44
Figura 5. Porcentaje de viviendas que cuentan con servicio de agua potable por medio de hidrante.	44
Figura 6. Porcentaje de viviendas que no cuentan con drenaje.	45
Figura 7. Estandarización de los indicadores.	46
Figura 8. Combinaciones lineales de las variables estandarizadas.	46
Figura 9. Índice de marginación del agua por Municipio.	48
Figura 10. Municipios de la República según grado de marginación del agua y alcantarillado.	53
Figura 11. Estados con grado de marginación del agua y alcantarillado Muy alto.	54
Figura 12. Estados con grado de marginación del agua muy alta.	55
Figura 13. Estados con grado de marginación del agua alto.	55
Figura 14. Estados con grado de marginación medio alto.	56

Figura 15. Estados con grado de marginación del agua medio baja .	57
Figura 16. Estados con grado de marginación del agua bajo I.	58
Figura 17. Estados con grado de marginación del agua bajo II.	59

ÍNDICE DE GRÁFICAS

Gráfica 1. Ingresos imputados por tipo de servicio, controlados por sexo del jefe de familia y zona.	75
Gráfica 2. Ingresos imputados por tipo de servicio, controlados por sexo del jefe de familia y zona.	76

ÍNDICE DE TABLAS

Tabla 1. Definición de variables.	40
Tabla 2. Estadísticas principales del análisis de componentes principales .	47
Tabla 3. Coeficientes de correlación entre el índice de marginación del agua y los cinco indicadores.	48
Tabla 4. Coeficientes del índice de marginación del agua y alcantarillado .	49
Tabla 5 Estratificación del índice de marginación del agua y alcantarillado.	50
Tabla 6. Número de Municipios por Estado conforme el índice de marginación del agua y alcantarillado.	52
Tabla 7. Evolución de los servicios de agua potable.	60
Tabla 8. Valores de la variable dependiente .	61
Tabla 9. Variables utilizadas en el modelo multilogístico.	61
Tabla 10. Modelos generados por el algoritmo multilogístico.	63
Tabla 11. Escenarios contruidos.	
Tabla 12 Resultado de la simulación parametrizada según los escenarios definidos en el tabla 11.	64
Tabla 13. Estimación por cambio de grupo.	66
Tabla 14. Análisis por cambio de percentiles en el ingreso mensual.	67
Tabla 15. Lista de variables comparables.	68
Tabla 16. Estadística descriptiva de los ingresos imputados.	74
Tabla 17. Escenarios contruidos para la aplicación del Estado de Oaxaca.	76

Tabla 18. Resultados de la simulación conforme los escenarios propuestos en la tabla 17.	78
Tabla 19 Análisis por cambio de percentiles del ingreso mensual per cápita imputado.	79
Tabla 20. Análisis por cambio de grupo.	80

RESUMEN

Dentro de los elementos fundamentales para el crecimiento y desarrollo social, el agua emerge como un factor determinante en este proceso. Dada la importancia que representa el vital líquido para la sustentabilidad económica y social, se justifica la aplicación de técnicas multivariadas de análisis que permitan la elaboración de políticas y estrategias que optimicen su gestión.

Los modelos de componentes principales, econométricos de elección múltiple, de imputación de datos, de simulación y de georeferenciación, son utilizados para el análisis e inferencia de la información. En esta investigación, se aplicaron a datos del Censo de Población y Vivienda 2010 y a la Encuesta Nacional de Ingresos y Gastos en los Hogares del mismo año, con el objetivo de construir el *índice de marginación del agua y alcantarillado* que permitió la clasificación de los Municipios de la República y su identificación geográfica, para detectar las zonas más afectadas por esta deficiencia del servicio. Mediante el modelo econométrico combinado con técnicas de simulación, se determinaron los factores socioeconómicos que tienen mayor influencia en la existencia del equipamiento de servicios de agua potable y alcantarillado en los hogares.

Se aplicó la metodología propuesta para el Estado de Oaxaca, por presentar el mayor número de Municipios con grado de marginación alta y muy alta. Esta prueba requirió de la imputación de los ingresos proporcionados por la ENIGH 2010 para la

Entidad a cada hogar del Censo de Población y Vivienda 2010 con la finalidad de resolver el problema de falta de representatividad de la Entidad en la muestra.

Se concluye, que la combinación de las técnicas propuestas, ofrecen visiones alternativas y realistas del problema que facilitan la planeación estratégica en el sector, las cuales serían difícilmente alcanzables por otra vía.

SUMMARY

Among key elements for growth and social development, water emerges as a key factor in this process. Given importance that represents vital liquid for economic and social sustainability, justifies application of multivariate techniques of analysis for development of policies and strategies to optimize their management.

The main component models, econometric multiple choice data allocation, simulation and georeferencing, are used for analysis and information inference. In this research, were applied to data from the Census of Population and Housing 2010 and the National Household Income and Household Expenditure in 2010, with the goal of building marginality index and sewage water was used to classify municipalities of the Republic and its geographical identification, to identify most affected areas by this deficiency service. . Using econometric model combined with simulation techniques, socioeconomic factors were determined to have a greater influence on equipment existence of drinking water and sewage in homes.

The methodology proposed for Oaxaca state, to present as many municipalities with high levels of marginalization. This test required the imputation of income provided by the ENIGH 2010 for the Bank, to every household in the Census of Population and Housing 2010, to solve lack problem of representation of the entity in the sample.

We conclude that combination of the proposed techniques, and realistic offer alternative views of the problem, facilitating strategic planning in the sector, which would be difficult to achieve by other means.

INTRODUCCIÓN

La concentración de la población y la actividad económica han creado zonas de alta escasez de agua potable, no solo en las regiones de baja precipitación pluvial, sino también en zonas donde esto no se percibía (CONAGUA, 2010).

Al considerar el agua como uno de los elementos fundamentales para el crecimiento y desarrollo social, se hace necesario realizar estudios que propicien y fortalezcan el establecimiento de políticas públicas y planes estratégicos para su adecuada gestión.

Esta investigación inicia con la construcción del *índice de marginación del agua y alcantarillado* a partir de datos del Censo de Población y Vivienda 2010, utilizando para ello, la aplicación del modelo de componentes principales para clasificar los Municipios de la Republica Mexicana en función de su marginación en servicios de agua potable y alcantarillado, los resultados indicaron que de las 2,456 Municipalidades, 472, tienen un grado de marginación bajo, 865 medio bajo, 689 medio alto, 335 alto y 95 muy alto. La georeferenciación de éste indicador, permitió identificar que los Estados al norte de la República como Baja California Norte, Baja California Sur, Chihuahua, Sonora, Coahuila, Nuevo León, y Durango, presentan poca marginación en cuanto al servicio de abastecimiento de agua potable y alcantarillado, en contraste, al sur del País, en Estados como Oaxaca, Guerrero, Chiapas y Veracruz el abastecimiento de servicios de suministro de agua potable y alcantarillado en algunas localidades es deficiente en extremo, destacando en este

rubro Oaxaca, del cual de los 570 Municipios que lo conforman, 72 se encuentran clasificados en grado de marginación muy alta y 169 en alta, lo cual representa el 42% de la Entidad.

Derivado de la problemática descrita, se volvió imperativo el conocer los factores determinantes para la existencia de equipamiento de servicios de agua potable y alcantarillado en los hogares. Para el logro de este objetivo, se aplicó un modelo econométrico multilogístico combinado con técnicas de simulación a los datos de la ENIGH 2010, encuesta que presenta una gran precisión en variables socioeconómicas de ingreso y gasto, además de que sus datos tienen representatividad a nivel nacional.

Los resultados obtenidos, indicaron que aunque existe una correlación directa entre los ingresos y el equipamiento de los servicios de agua potable y alcantarillado en los hogares, variables socioeconómicas como la presencia de analfabetismo en los miembros del hogar, el grado máximo de estudios del jefe de familia y la presencia del servicio de energía eléctrica influyen en la existencia de estos servicios.

El caso del Estado de Oaxaca, motivó a la aplicación de la técnica propuesta en sus Municipios a fin de conocer los factores que determinan el abastecimiento del servicio de agua potable y alcantarillado. La principal problemática encontrada para la aplicación del modelo, es que los datos para Oaxaca en la ENIGH 2010 no son representativos para la encuesta, de hecho, la muestra, no permite una desagregación

de información a nivel estatal o municipal con un grado de precisión estadísticamente confiable en todos los Estados y, por ser una muestra aleatoria de la población, la encuesta no incluye información sobre todos los Municipios del país. La principal causa de esta deficiencia es simplemente de orden presupuestal, ya que la captación de la información detallada de ingresos y gastos en los hogares requiere de cuestionarios muy detallados con personal especializado y en algunos casos de visitas continuas.

Por otro lado, los datos del Censo de Población y Vivienda 2010, cuentan con un grado de precisión confiable y son estadísticamente representativos en todas las Entidades, sin embargo, las preguntas incluidas en éste sobre ingresos, son muy restringidas, lo cual hace impreciso este indicador.

En respuesta a esta problemática, se utilizó la metodología propuesta por Elbers, Lanjouw y Lanjouw (2003), la cual consiste en especificar al ingreso como función de las características del hogar, la vivienda, los miembros del hogar y otros indicadores, para después imputar un ingreso con menor grado de error a cada hogar en el censo con base en una predicción estadística.

Una vez realizada la imputación de los ingresos de la ENIGH a las viviendas del Censo de Población y Vivienda, ambos para el año 2010, se aplicó el modelo econométrico multilogístico y técnicas de simulación de Monte Carlo para

determinar los factores que influyen en el abastecimiento del servicio de agua potable y alcantarillado en los hogares del Estado de Oaxaca.

Los resultados encontrados a nivel Estado, fueron consistentes con los Nacionales, lo cual permite proponer estrategias a los gobiernos como el abatimiento del analfabetismo, la educación continua y la generación de empleos mejor remunerados, lo cual traerá como efecto colateral una disminución de marginación de agua y alcantarillado y por ende, la sustentabilidad del desarrollo de las Entidades.

CAPÍTULO 1

PROPÓSITO Y ORGANIZACIÓN

1.1 Planteamiento del problema

La escasez de agua dulce es uno de los principales problemas ambientales que tiene la humanidad. Se considera que el 60% de la población del planeta sufre escasez, y lo más preocupante es que actualmente se estima que 1,000 millones de personas no tienen acceso al agua potable. Para los países desarrollados, el problema del agua afecta, sobre todo, la conservación de la naturaleza y las posibilidades de crecimiento económico, mientras que en las naciones en desarrollo, además, la falta de agua potable es la causante directa de enfermedades evitables como la diarrea y el cólera que originan un alto porcentaje en la muerte de niños menores de 5 años (WHO-UNICEF, 2004).

La toma de decisiones asociadas con la asignación y planificación de los recursos hídricos son situaciones complejas que requieren técnicas multidisciplinarias para evaluar sus efectos en un contexto social, económico y medioambiental (Manoli, Katsiardi, Arampatzis & Assimacopoulos, 2005)

Según Manoli, et al. (2005), la gestión del agua para consumo humano debe realizarse como una evaluación sistemática con un horizonte de planificación a largo plazo, simulando los efectos acumulativos de la demandada a través del tiempo con el objeto de prevenir los posibles cambios en futuros e inciertos.

De acuerdo con M. Thobani (2007), históricamente, los procesos de planificación de la gestión de agua partían de la proyección de la población a la que se necesitaba satisfacer, estimando un uso per cápita de agua, y simplemente multiplicando una proyección por la otra para obtener una estimación del agua futura. A partir de esa estimación, el objetivo de los gestores consistía en identificar aquellas fuentes de oferta disponible en la región para añadirlas apropiadamente a la oferta ya existente.

Hoy en día y sobre todo a partir de la aparición de restricciones ecológicas, financieras y políticas en las décadas de los 80 y 90, se considera que esta metodología sufre de importantes carencias.

Se considera que el problema de la gestión del agua cuenta con un amplio abanico de factores que juegan un influyente e importante papel en la planificación de la gestión de recursos de agua en general y específicamente en las aguas urbanas y de uso doméstico (Wunderlin, Díaz, Amé, Pesce, Hued, & Bistoni, 2001).

Como consecuencia, surge la necesidad de técnicas y herramientas que capten ese grupo de factores para tenerlos en cuenta en la toma de decisiones (Arrojo, 2003).

Hoy los gestores del recurso agua, se plantean intentar entender, modificar y controlar la demanda, alternativa generalmente ignorada, y dejar de considerarla como un valor exógeno e inmodificable (Beard, 1995).

Actualmente, las familias se basan en una variedad de fuentes de abastecimiento de agua con diferentes características y niveles de servicios (precio, distancia, calidad y fiabilidad). Para muchos hogares de los países en vías de desarrollo, el agua es un bien de origen heterogéneo, que no suele ser el caso en los países industrializados (MU, Whittington, & Briscoe, 1990).

La calidad del agua es uno de los factores más relevantes que influyen directamente en la salud humana y de todo ser vivo (Kazi, Arain, Jamali, Jalbani, Afridi, Sarfraz, Baig & Shah, 2009). Por lo tanto, es imperativo prevenir y controlar la contaminación de ríos y mantos acuíferos que son fuente de abastecimiento para el consumo de agua humano.

La caracterización de la distribución espacial, la variación, calidad y fuente del suministro de agua, puede proporcionar una mejor comprensión de los parámetros de calidad del agua y ayudar a los políticos a establecer las prioridades para la gestión sostenible del agua (Huang, Wang, Liping, Zhiqing & Jiaping, 2010).

El análisis de componentes principales es una herramienta eficiente para reducir la dimensión de los datos y la agrupación de las variables de acuerdo a sus características comunes y su aplicación al sector agua potable puede ayudar a establecer las prioridades para la gestión sostenible del recurso. (Vialle, Sablayrolles, Lovera, Jacob & Huau, 2011).

Shrestha & Kazama (2007), demostraron que el análisis de componentes principales ofrece reducción de datos y pone de manifiesto los parámetros más significativos resultantes de su aplicación a conjuntos de datos referentes al abasto del agua.

El abastecimiento del agua está caracterizado por un gran número de variables y parámetros que representan una distribución de agua particular en localidades específicas como el tiempo y tipo de servicio. Los métodos multivariantes como el análisis de conglomerados, el análisis de componentes principales, el análisis factorial y el análisis discriminante, se justifican para la evaluación de la calidad de las aguas y el servicio (Zeng & Rasmussen, 2005).

La obtención de agua de fuentes fuera del hogar propicia gastos de recaudación, mismos, que deben tenerse en cuenta para evaluar el impacto en la economía familiar. Según resultados de la Encuesta Nacional del Uso del Tiempo (2002), la falta de servicio de agua potable entubada en los hogares, provoca un aumento en las labores domésticas hasta por tres horas.

De acuerdo con Whittington et. al (1990), los investigadores han utilizado cuatro estrategias principales para obtener la información necesaria y modelar eficientemente el comportamiento de los requerimientos de agua en los hogares de los países subdesarrollados. 1) encuestas a las viviendas para complementar los datos existentes de fuentes públicas y/o privadas. 2) levantamiento de censos directos en los hogares buscando conocer su reacción en situaciones hipotéticas de uso del agua.

3) análisis del incremento de la plusvalía de la vivienda al agregar servicios de agua potable. 4) métodos experimentales (incluyendo ensayos controlados aleatorios) para probar cómo se comportan los hogares en respuesta a diferentes intervenciones de abastecimiento de agua (Kremer, 2007).

En este contexto, satélites de observación de la Tierra y técnicas avanzadas de sensores remotos han sido propuestos como herramientas para reducir la incertidumbre en cuanto a las expectativas de crecimiento urbano, el cual figura en los planes municipales de zonas urbanas. Sin embargo, también se han encontrado útiles para la detección, cuantificación y planificación de las áreas verdes urbanas, que son considerados como elementos importantes para satisfacer la demanda de agua. (Beamonte , 2010).

El análisis exploratorio de los datos espaciales (ESDA) consiste en un conjunto de técnicas utilizadas para describir y visualizar distribuciones espaciales, detectar patrones de asociación espacial y aglomeraciones en torno a un atributo, así como sugerir regímenes espaciales u otras formas de heterogeneidad (Haining, 1990).

Este tipo de análisis requiere contar con una medida de asociación lineal para medir grados de semejanza entre las observaciones en función de sus atributos, para tal efecto, en este trabajo de investigación se construyó el *índice de marginación del agua y alcantarillado*, indicador que permitió visualizar la clasificación de los

Municipios en función del abastecimiento y tipo de servicio del agua en un mapa utilizando técnicas de georeferenciación.

Para Grafton, Landry, Libecap, and O'Brien. (2010), la eficiencia en el consumo de agua se lograría si la administración de ésta funcionara dentro de un mercado competitivo, sin embargo esto no sucede en la realidad. Tanto en el caso mexicano como en el de otros países, se observa que los organismos operadores de la distribución de agua en su mayoría son entidades descentralizadas del sector público, es en este contexto que el análisis se dificulta y adquiere mayor relevancia demostrar que aun con este tipo de esquemas se pueden aplicar incentivos costo-efectivos que permitan el uso eficiente del líquido durante periodos de escasez de corto y largo plazo en la mayoría de las Entidades y Municipios de México.

Conforme con las estimaciones del Consejo Nacional de Población (2008), entre 2007 y 2030 la población en México se incrementará en casi 14.9 millones de personas. Además, aproximadamente el 82% de la población total se asentará en localidades urbanas.

La República Mexicana se puede dividir en dos grandes zonas: la zona norte, centro y noroeste, donde se concentra el 77% de la población, se genera el 87% del PIB, pero únicamente ocurre el 31% del agua renovable; y la zona sur y sureste, donde habita el 23% de la población, se genera el 13% del PIB y ocurre el 69% del agua renovable (CNA, 2008).

Al año 2030 en algunas de las Regiones del país, la disponibilidad natural media de agua alcanzará niveles cercanos o incluso inferiores a los 1,000 m³/hab/año, es decir una condición calificada como una condición grave de escasez. (CONAPO, 2008).

El constante crecimiento poblacional y la disponibilidad de agua futura en México, destaca la importancia de introducir estrategias de conservación en la demanda de agua en los Estados y Municipios, considerando que la cobertura del servicio difícilmente puede seguir el ritmo de aumento de consumo de agua debido al alto costo de las infraestructuras (Dube y Van der Zaag, 2003).

El cálculo del bienestar en los hogares, es una herramienta indispensable para la planeación de las políticas públicas e implementación de estrategias que permitirán el buen desempeño de un país. En este contexto, los ingresos captados por los miembros del hogar se consideran como fundamentales para la construcción de indicadores que contribuyan a estos objetivos.

En México, la principal fuente para conocer la captación de los ingresos, es proporcionada por Censos y Conteos de población realizados por el INEGI. Los instrumentos de captación de la información referente a los ingresos en los censos no están diseñados para representar los ingresos con precisión, lo cual es resultado de una subestimación del nivel de bienestar. Por otro lado, los conteos y encuestas, permiten integrar con mayor precisión los ingresos familiares, sin embargo, por las características de los cuestionarios aplicados para este fin, no es posible contar con

estos datos con representatividad en todos los Estados de la República y más difícil, a niveles Municipal o por Localidad.

Esta problemática, dio pie a las primeras teorías sobre la imputación de datos, las cuales inician a partir de los años cincuenta por Hansen, Hurwitz y Madow (1953), los cuales utilizan modelos de regresión para mejorar estimadores en áreas pequeñas.

Moore (1952) y Hartley (1958) utilizan una técnica simplificando el método de máxima verosimilitud con el objeto de anexar información a bases de datos incompletas.

En los noventa, Angrist y Krueger (1992) relacionan información del censo de población de los Estados Unidos de 1960 con el censo de población de 1980 para medir la relación entre los años de escolaridad y la edad en la que una persona se incorpora a la enseñanza formal. Por otro lado, Arellano y Meghir (1992) relacionan la encuesta del gasto de las familias y la encuesta de la fuerza laboral del Reino Unido, ambos para el año 1983. Construyeron un modelo de optimización intertemporal de oferta del trabajo para las mujeres con variables existentes en ambas fuentes. Similarmente, Alderman y otros (2002) combinaron la encuesta de ingreso y gasto de hogares de Sudáfrica de 1995, con el censo de población de 1996 para calcular indicadores desagregados de bienestar.

Elbers, Lanjouw y Lanjouw (2003) toman como base el método de estimación de áreas pequeñas y proponen una extensión de esta metodología calculando la población con funciones no lineales, siendo necesario considerar medidas semi-paramétricas y no paramétricas. La técnica se ha aplicado a diferentes países, entre ellos: Tailandia, Cambodia, Sudáfrica, Brasil, Nicaragua, Madagascar, Guatemala y Ecuador entre otros.

En México, Hernández (1996) estima la oferta laboral y la pobreza en las zonas urbanas de México. Utiliza información de la ENIGH (Encuesta Nacional de Ingresos y Gastos de los Hogares) para imputar ingresos de los hogares a los registros de la ENEU (Encuesta Nacional de Empleo Urbano). En primer lugar identifica las variables comunes para los hogares urbanos de dichas bases de datos y después estimó un modelo Tobit del ingreso no laboral con el fin de capturar los parámetros estimados de dicho ingreso y predecir el ingreso no laboral en la ENEU.

1.2 Propósito de la investigación

El propósito de esta investigación es la utilización de diferentes técnicas estadísticas que permitan identificar y analizar los factores sociales, económicos, demográficos, culturales, climáticos y territoriales que ejercen una influencia importante sobre la existencia del servicio de agua potable y alcantarillado en los hogares y determinar estrategias de conservación que ayuden a gestionar el recurso y plantear una metodología para integrar y sensibilizar las variables determinantes en el equipamiento de estos servicios en el hogar y construir escenarios para la sensibilización de políticas y estrategias encaminadas a una mejor gestión del recurso.

1.3 Objetivo general

Diseñar una metodología basada en el análisis de aspectos económicos, geográficos y técnicos que ejerzan influencia sobre la existencia de agua urbana y alcantarillado, utilizando como vehículo de análisis la simulación, técnicas de georeferenciación, modelos econométricos de elección múltiple, de imputación de datos y análisis factorial.

1.4 Objetivos específicos.

- Realizar un diagnóstico de las condiciones actuales del servicio de agua potable en el País.
- Describir y analizar los factores socioeconómicos y demográficos que afectan la existencia de agua y alcantarillado en los hogares.
- Probar la metodología propuesta en un Estado con marginación de servicios de agua potable.
- Evaluar el impacto de las estrategias y políticas de solución propuestas mediante la simulación.

1.5 Justificación de la investigación

La gestión doméstica del agua representa un caso paradigmático que presenta un comportamiento complejo. Mientras que su administración se ha basado históricamente en políticas del lado de la oferta, centradas en compensar el abastecimiento de la demanda, que depende de un sistema socio-económico y tecnológico percibido como una condición de contorno exógena, actualmente, se está produciendo un cambio de paradigma en la concepción de su gestión.

La gestión del recurso agua potable se presenta como una oportunidad para hacer políticas y estrategias de cambio. Este canje de paradigma, exige también una evolución en las herramientas de ayuda a la toma de decisión.

Muchas de las técnicas de estimación se han basado en sistemas de predicción sin teoría, en mero ajuste predictivo pero sin proporcionar conocimiento de la estructura, y sin ella, se dificulta tomar decisiones. Si se pretende entender los mecanismos y las implicaciones que tienen lugar en las dimensiones sociales y de cambio tecnológico, es necesario representarlas explícitamente en los modelos.

Es prioritario representar la dimensión humana y cierto nivel de cognición. Este hecho se ve reforzado por la importancia social que está cobrando el agua como recurso y que es fácilmente detectable en los últimos años.

La investigación propuesta se justifica por la utilidad que representa el abastecimiento de agua urbana y alcantarillado en nuestro medio, el no considerar ciertas dinámicas que están articuladas a los consumos de agua y alcantarillado que demanda la población, resulta en una ineficiencia en su administración. Estas dinámicas incluyen factores sociales, demográficos, económicos y territoriales que de una u otra forma influyen en su demanda. Por lo tanto, es necesario explorar y analizar dichos elementos tanto para estimación futura del servicio como para una adecuada gestión del recurso en los diferentes Estados y Municipios.

1.6 Alcances y limitaciones

1.6.1 Alcances.

- Para la construcción de los indicadores, se utilizó la ENIGH (Encuesta Nacional de Ingresos y Gastos en los Hogares) y el Censo de Población y vivienda, ambos para el año 2010.
- Se realizó un análisis de la marginación del agua y alcantarillado a nivel Nacional.
- Se aplicó la metodología propuesta al Municipio de Oaxaca.
- En la definición del modelo, se incluyeron variables socioeconómicas y culturales de los hogares.
- Se utilizó el modelo Entidad-Relación para la construcción de las bases de datos.
- Se diseñaron programas y rutinas en Stata 11.1 para la ejecución de los modelos econométrico, de simulación y de componentes principales

- Se definieron coberturas y se estandarizaron las proyecciones del marco geoestadístico Nacional y Municipal, se calcularon centroides y se relacionaron con base de datos generada de la estratificación del índice de marginación del agua y alcantarillado utilizando ArcGis Ver. 9.0.
- Se realizó la imputación de datos utilizando el software libre PovMap.

1.6.2 Limitaciones.

- La metodología propuesta no se aplica a todos los Municipios de la República Mexicana, por lo tanto es un estudio de un solo caso.
- Solo se imputaron los ingresos mensuales per cápita del Estado de Oaxaca reportados en la ENIGH a los registros de los hogares del Censo de Población y Vivienda.
- El estudio fue realizado utilizando el Censo de población y Vivienda 2010 y la ENIGH para este mismo año.

1.7 Organización del estudio

El primer capítulo trata de la descripción y planteamiento del problema, los objetivos y alcances del estudio de investigación.

El segundo capítulo nos proporciona la teoría de los modelos econométricos de elección múltiple, de imputación de datos, técnicas de simulación, componentes principales, y georeferenciación, que fueron utilizados para la determinación de las variables socioeconómicas que influyen en el equipamiento del recurso agua y

alcantarillado en los hogares y poder crear escenarios de prueba para las diferentes estrategias y políticas propuestas encaminadas a la eficiencia en el manejo del recurso.

El tercer capítulo se plantea el alcance y el tipo de investigación, se presenta el diseño y las características del mismo, los resultados a priori del estudio de investigación y se realiza el análisis de los mismos.

En el cuarto capítulo se revisa el cumplimiento de los objetivos planteados, la comprobación de las hipótesis y se plantea la metodología utilizada.

En el quinto capítulo se proporcionan recomendaciones hacia los Organismos operadores del servicio de agua potable y alcantarillado y sugerencias en función a los resultados obtenidos.

CAPÍTULO 2.

MARCO TEÓRICO

2.1 Modelos econométricos de elección múltiple.

2.1.1 Modelo de regresión logístico multinomial.

Los modelos de elección múltiple pueden plantearse como un problema de decisión. Por ejemplo, supóngase que cada individuo u hogar de una muestra puede elegir una de tres alternativas posibles.

Modelo logístico. Denotemos por Y_{ij} una variable que toma el valor 1 si el individuo i escoge la opción j , mientras que toma el valor 0 si el individuo escoge otra de las opciones. Sea P_{ij} la probabilidad de que la variable Y_{ij} tome el valor 1 y supongamos que las probabilidades P_{i1} , P_{i2} , P_{i3} dependen linealmente de un vector de características x_i , de modo que:

$$P_{i1} = x_i' \beta_1; \quad P_{i2} = x_i' \beta_2; \quad P_{i3} = x_i' \beta_3$$

Donde β_1 , β_2 y β_3 son vectores de dimensión k . En este contexto, se puede plantear el siguiente modelo:

$$Y_{ij} = x_i' \beta_j + u_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, 3$$

Donde N es el número de individuos en la muestra. Como $P_{i1} + P_{i2} + P_{i3} = 1$, se tiene $\beta_{j1} + \beta_{j2} + \beta_{j3} = 0$ para los coeficientes de la variable j -ésima en los tres modelos, excepto para el término independiente, para el que $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Estos modelos son probit o logit multinomiales (Novales, 2008).

2.2 Simulación.

2.2.1 Introducción

La simulación permite resolver cuestiones complejas para las cuales los métodos cuantitativos no son eficaces o no pueden aplicarse. Robert Shannon (1988, p.18) lo define como “el proceso de diseñar y desarrollar un modelo computarizado de un sistema o proceso, y conducir experimentos con este modelo con el propósito de entender el comportamiento del sistema o evaluar estrategias con las cuales se puede operar sobre él”. Esta definición muestra las grandes posibilidades que brinda esta herramienta para desarrollar la creatividad y construir el conocimiento, como así también las nuevas necesidades que la misma plantea. La experimentación con un modelo conlleva normalmente la necesidad de generar datos e investigar parámetros a efectos de analizar el comportamiento global del mismo o la sensibilidad a los distintos parámetros. Tales procesos no son, en general, simples; tienen un costo. Por ejemplo, el gran número de variables que se pueden manejar exige un mayor cuidado y profundidad en el estudio de la propagación de errores. Así entonces el control del método y del software utilizado en la resolución de un problema adquiere un peso crucial a la hora de evitar la obtención de conclusiones erróneas.

2.2.2 Simulación como técnica para solucionar problemas

Simulación, es una forma de realizar experimentos en la computadora, la cual ayuda a las empresas a realizar la simulación de un proyecto para ver si valdrá la pena desarrollarlo en la empresa.

Según Acoff (1971, p.150) “la simulación hace posible estudiar y experimentar con las complejas interacciones que ocurren en el interior de un sistema dado, ya que sea en una empresa, industria, economía o un subsistema de cualquiera de ellas”.

A través de la simulación se puede estudiar los efectos de ciertos cambios informativos, de organización y ambientales, en la operación de un sistema, al hacer alteraciones en el modelo y observar los efectos de éstas en el comportamiento del problema.

La observación detallada del sistema que se está simulando, conduce a un mejor entendimiento del mismo y proporciona sugerencias para mejorarlo.

La experiencia que se adquiere al diseñar un modelo de simulación en una computadora, puede ser más valiosa que la simulación en sí misma. El conocimiento que se obtiene al diseñar un estudio de simulación sugiere, frecuentemente, cambios en el sistema en cuestión. Los efectos de estos cambios pueden probarse, entonces, a través de la simulación, antes de implantarlos en el sistema real.

Para Barnes (1971, p.55) “la secuencia de los eventos puede ser muy importante, pues la información acerca de los valores esperados y de los momentos, puede ser suficiente para describir el proceso”. En estos casos los métodos de Monte Carlo pueden constituir la única forma satisfactoria de obtener la información requerida.

2.2.3 Modelos de simulación como herramientas de investigación

Los sistemas que engloban la gestión de recursos naturales se consideran sistemas complejos puesto que suelen integrar heterogeneidad, influencia espacial e integración con procesos autónomos. Además con frecuencia suelen ser procesos lentos. Por todo ello no resulta fácil realizar experimentos directamente sobre los sistemas reales de estudio. Como argumentan Barreteau (2001), la necesidad de replica y de control de los parámetros, y la cantidad de tiempo necesario para realizar los experimentos justifica la utilización de metodologías indirectas como la simulación.

2.2.4 Modelos de simulación como herramientas de entrenamiento y aprendizaje

Los modelos de simulación en general (Axelrod, 1997) pueden ser utilizados también para propósitos de enseñanza y entrenamiento en áreas de recursos naturales o en cualquier otra.

La simulación proporciona herramientas para poder jugar con el conjunto de reglas conocidas en un determinado contexto ambiental y poder explorar las consecuencias del comportamiento del gestor, de sus políticas y su sensibilidad, o simplemente de

las condiciones iniciales o de contorno que se impongan. Así se puede llegar a conseguir un mejor entendimiento de los comportamientos complejos de los sistemas con recursos. La utilización de estos modelos supone desplazarse del famoso *learning by doing* al *learning by simulation*. Si se asume que el aprendizaje en un sistema virtual puede ser útil en el aprendizaje sobre el sistema real quizá sea posible limitar parcialmente el costo de los métodos de ensayo y error (Barreteau et al., 2001)

2.2.5 Modelos de simulación como herramientas de soporte a la decisión y negociación

Otro de los usos de la simulación es como herramienta de ayuda a la decisión y negociación. Lo que se entiende por decisión en el marco de la gestión de un recurso emerge frecuentemente de la interacción de un pequeño número de stakeholders. En este contexto los sistemas de simulación se conciben como un objeto de mediación y soporte entre los grupos de decisión para una mejor negociación (Bousquet, Barreteau, Le Page, Mullon, & Weber, 1999), o bien para inducir a los propios stakeholders a acordar representaciones comunes del recurso y, por tanto, utilizar el sistema como herramienta de facilitación de la comunicación, concentrando los procesos de selección en tiempo y espacio.

La utilización de modelos de simulación como herramientas de discusión, especialmente los modelos basados en agentes, generalmente más descriptivos y fácilmente comprensibles, pueden prevenir ciertas ideas erróneas entre los

stakeholders y comportamientos de mala fe, favoreciendo procesos de empatía y elaboraciones conjuntas de los modelos. Conviene tener presente que casos en los que los modelos utilizados no son públicos, y en consecuencia sus hipótesis no son consensuales, pueden provocar procesos de negociación fallidos (Reitsma, Zigurs, Lewis, Wilson & Sloane, 1996).

2.3 Sistemas de información geográfica.

2.3.1 Introducción a los Sistemas de Información Geográfica (SIG)

El estudio de fenómenos que incorporan de manera relevante el concepto de espacio se puede ver favorecido a través de la inclusión de tecnologías computacionales para el análisis y modelado de espacios geográficos o los Sistemas de Información Geográficos, (SIG)

Entre las múltiples definiciones de SIG, Star y Estes (1990) lo describen como un sistema de información diseñado para trabajar con datos referenciados por coordenadas espaciales o geográficas. En otras palabras, un SIG es tanto un sistema de base de datos con capacidades específicas para datos referenciados espacialmente, como un conjunto de operaciones para trabajar con los datos propiamente.

La base de datos de un SIG está formada a su vez por dos tipos de información. Por un lado están los datos sobre la localización espacial, datos espaciales de los objetos incluidos en la base de datos de estudio, lo que se conoce como la base de datos

geográfica. Por otro, el SIG contiene datos sobre la información temática o los atributos de los objetos geográficos incluidos, la base de datos alfanumérica (Clarke, 1999).

En un SIG, los datos que forman la base de datos espacial se representan mediante capas, cada una sobre una temática similar, que representan un determinado aspecto geográfico del entorno bajo estudio (figura 1). Cada una de estas capas puede contener datos de atributos formando las bases de datos alfanuméricas. Estos datos contienen descripciones sobre cada entidad representada en la capa y suelen almacenarse de forma tabular de tal manera que mantengan vínculos con los objetos gráficos.

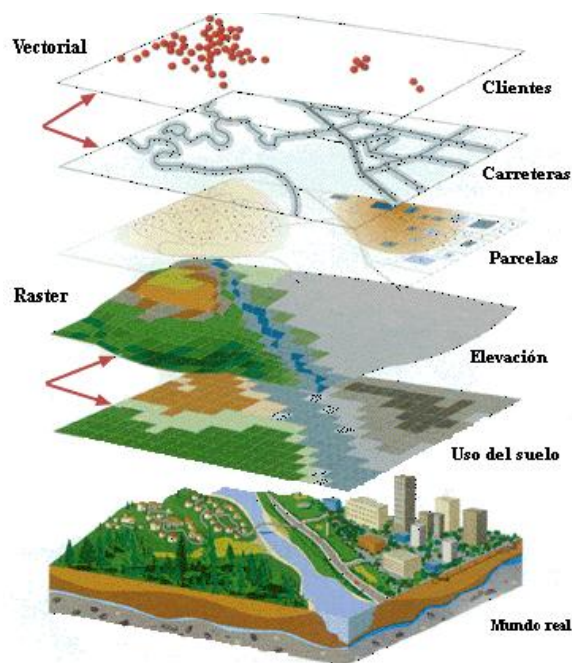


Figura 1. Estructura de capas y estructuras de información de un sistema de información geográfico. Imagen retocada a partir de original en <http://cier.uchicago.edu/gis/gis.htm> [Fecha de acceso:01 de noviembre de 2011]

La información de las capas se puede almacenar utilizando dos formatos, el formato raster y las estructuras de datos vectoriales. El modelo de SIG raster o matricial compartimenta el espacio en celdas regulares donde cada una de ellas representa un único valor. Matemáticamente se puede definir como un conjunto de celdas localizadas en coordenadas contiguas, implementadas en una matriz 2D. Cada celdilla, también llamada elemento de imagen, elemento de matriz o píxel, es referenciada por índices de línea y columna y contiene un número representando el tipo o valor del atributo de la celda. Cuanto mayor sean las dimensiones de las celdas (resolución) menor es la precisión o detalle en la representación del espacio geográfico.

En el caso del modelo de SIG vectorial, el interés de las representaciones se centra en la precisión de localización de los elementos sobre el espacio. La representación vectorial de un objeto es un intento de representar de la forma más exactamente posible, procurando definir de manera precisa todas las posiciones, longitudes y dimensiones de las entidades geográficas. Consiste normalmente en listas de coordenadas 2D que delimitan regiones temáticas mediante puntos, líneas y polígonos.

Las ventajas que presenta la representación vectorial frente a la representación raster se deben principalmente a la representación matemática más precisa y más compacta de elementos discretos, como por ejemplo la representación de las relaciones

topológicas de forma adecuada. Además permite asociar de forma sencilla atributos a elementos gráficos, proceso mucho más complejo en la representación raster.

La representación raster, no obstante, tiene también mucha importancia puesto que, pese a la pérdida de resolución, permite representar fenómenos con variación continua en el espacio y por tanto mecánicas y algoritmos mucho más sencillos en problemas de modelado y simulación.

2.4. Componentes principales.

2.4.1. El análisis de componentes principales

El análisis de componentes principales tiene este objetivo: dadas n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales. Por ejemplo, con variables con alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20 por 100 de las originales) expliquen la mayor parte (más del 80 por 100 de la variabilidad original) (Aluja, 1999).

De acuerdo con Anderson (1958), la técnica de componentes principales es debida a Hotelling (1933), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson (1901). Su utilidad es doble:

1. Permite representar óptimamente en un espacio de dimensión pequeña, observaciones de un espacio general ***p-dimensional***. En este sentido, componentes principales es el primer paso para identificar las posibles variables latentes, o no observadas que generan los datos.

2. Permite transformar las variables originales, en general correlacionadas, en nuevas variables incorrelacionadas, facilitando la interpretación de los datos.

2.4.2. Los componentes principales.

El análisis en componentes principales es una técnica de reducción de la dimensionalidad. Su objetivo es explicar la mayor parte de la variabilidad total de un conjunto de unas variables cuantitativas con el menor número de componentes o factores comunes posible (Anderson & Willis, 2003).

El análisis en componentes principales es un método geométrico de carácter descriptivo. El objetivo es descubrir la estructura subyacente en un conjunto de ***n*** individuos estudiados bajo una serie de ***p*** variables cuantitativas.

Imaginemos que tenemos ***n*** individuos medidos bajo una sola variable, es fácil describir a estos individuos representándolos sobre una recta. Si fuesen dos las variables representaríamos a dichos individuos sobre un plano. Incluso si las variables fuesen tres, el recurso a la representación gráfica de lo que se suele llamar ***la nube de puntos*** (el conjunto de los individuos) seguiría siendo posible. Estas

representaciones gráficas nos permiten observar geoméricamente la estructura interna de la nube de puntos.

Supongamos que a pesar de que la dimensión del conjunto de variables sea superior a dos nosotros pretendemos representar nuestros individuos, nuestra nube de puntos sobre un plano. En este caso es claro que nuestra proyección sobre el plano reflejará una nube de puntos deformada, donde las distancias entre los puntos ya no son las originales. Conscientes de ello, buscaremos aquel plano en el cual la deformación sea mínima, en otras palabras, aquel en el que la nube se extienda al máximo

El análisis en componentes principales es un método que permite transformar un conjunto de variables originales en otro conjunto de variables llamado conjunto de componentes principales. Estas componentes principales son combinación lineal de las variables originales y se caracterizan por estar incorreladas entre sí. En principio se generarán p componentes principales, es decir, tantas como variables originales.

Si se toman las variables originales y se calcula su matriz de correlaciones se puede observar que, habitualmente, existe un alto grado de correlación entre algunas de las variables. Esto nos lleva a pensar que, quizá, podríamos trabajar sobre un conjunto de variables incorrelacionadas de menor dimensión. Así, si las variables originales están muy correlacionadas entre sí, es esperable que su información se pueda expresar a través de unas pocas componentes principales. Si por el contrario las

variables originales están muy poco correlacionadas entre sí, el número de componentes principales será similar al de las variables originales.

2.4.2.1 Formulación matemática del problema de componentes principales.

En todo problema de análisis en componentes principales se suele trabajar con las variables originales tipificadas para evitar problemas de escala. Sea Y la matriz de variables originales tipificadas. Recordemos que pretendemos obtener un nuevo conjunto de variables que son combinación lineal de las variables originales, señalemos también que la varianza será una ***medida de la información*** que contiene cada variable (Cooley & Lohnes, 1971), con estas hipótesis el problema del análisis en componentes principales puede ser descrito de la siguiente forma:

Toda combinación lineal c de las variables originales puede expresarse de la siguiente manera:

$$c = Yv$$

Donde v es el vector que nos permite obtener la combinación lineal.

Pues bien, ***la primera componente principal es la combinación lineal de variables originales de varianza máxima***. Por tanto, buscamos v_1 de norma una tal que la varianza de la primera componente principal c_1 sea máxima.

Como la varianza de las componentes principales c se escribe como:

$$S_c^2 = v^t V_y v$$

Podemos resumir el planteamiento del problema de la siguiente manera:

Buscar v tal que $v^t V_y v$ sea máxima

Sujeto a $\|v\| = 1$

Es decir, se trata de resolver el siguiente problema:

$$\text{Max } v^t V_y v$$

Sujeto a $v^t v = 1$.

Para ello escribimos el siguiente lagrangiano:

$$L = v^t V_y v - \lambda(v^t v - 1).$$

Ahora basta con derivar respecto a v e igualar a cero:

$$\frac{\partial L}{\partial v} = 2V_y v - 2\lambda v = 0 \Rightarrow V_y v = \lambda v \Rightarrow (V_y - \lambda I)v = 0$$

De donde se concluye que v es vector propio de la matriz de varianzas- covarianzas de los datos originales tipificados. Como hemos impuesto la condición de que la varianza sea máxima, escogeremos el vector propio con mayor valor propio asociado.

Es decir, la primera componente principal c_1 se obtiene haciendo: $c_1 = Yv_1$ donde v_1 es el vector propio de la matriz de varianzas-covarianzas con mayor valor propio asociado.

Una vez obtenida la primera componente principal vemos cómo se obtiene el resto de componentes principales.

La segunda componente principal sería aquella combinación lineal de variables originales de varianza máxima y ortogonal a c_1 . Para obtener ésta y las siguientes componentes principales no es sin embargo necesario retomar el problema de maximización desde el principio.

Toda matriz de varianzas-covarianzas es simétrica y semidefinida positiva, y tiene, por tanto, p vectores propios ortogonales dos a dos y sus valores propios asociados son todos positivos o nulos.

Los vectores propios de la matriz V_y asociados a los valores propios escritos en forma decreciente son, por tanto, los vectores buscados (los factores principales). Estos

vectores nos permiten calcular las componentes principales a través de la expresión $c = Yv$. La varianza de cada componente principal viene dada por los valores propios.

El número de valores propios no nulos nos da la dimensión del espacio de las componentes principales.

Puesto que la varianza es una medida de la información, veamos cuál es la varianza de las componentes principales. Habíamos dicho que

$$S_c^2 = v^t V_y v.$$

Así, la varianza de la componente h -ésima será:

$$S_{ch}^2 = v_h^t V_y v_h = \lambda_h.$$

Por tanto, tendría sentido definir la medida de la información recogida por cada componente como el cociente entre la variabilidad componente y la varianza total, esto es:

$$\frac{S_{ch}^2}{\text{Varianza total}} = \frac{S_{ch}^2}{\text{Traza } V_y}$$

Pero hemos visto que traza $V_y = \sum_{h=1}^p \lambda_h$, así se puede definir de la proporción de la variabilidad explicada por la componente h-ésima como:

$$\frac{\lambda_h}{\sum_1^p \lambda_h}$$

Además, como en nuestro caso las variables están tipificadas se tiene: Traza $(V_y) = p$, y de esta forma, la anterior proporción quedaría de la siguiente manera:

$$\frac{\lambda_p}{p}$$

2.4.2.2 Los resultados de los componentes principales y su interpretación.

La primera cuestión es la de cuántas componentes tomar. Por un lado, parece lógico tomar aquellas componentes cuyo valor propio asociado sea mayor que uno, es decir aquellas componentes que aportan más información que cualquiera de las variables originales tipificadas. Sin embargo, hay que tener en cuenta algunas otras consideraciones como que las componentes que agrupan una gran cantidad de variables originales tienen una gran importancia. La experiencia y un buen conocimiento del contexto en el que se enmarca el estudio son casi siempre determinantes a la hora de la interpretación de los ejes (Cuadras & Fortiana, 2000).

En los valores propios se observa que la suma de éstos es igual al número total de variables iniciales p . esto es así porque la suma de los valores propios no es más que la traza de la matriz de varianza de las componentes principales, es decir, es la

inercia total I expresada en función de las componentes principales. Ahora bien, puesto que las variables iniciales estaban tipificadas, tenemos que la inercia total I en función de las variables originales vale $I = p$.

En resumen:

$$I = \sum_{l=1}^p \lambda_l = p.$$

Las coordenadas finales de los individuos en las componentes principales se obtienen haciendo $c_i = Yv_i$ $i = 1, \dots, p$, donde los v_i son los vectores propios de la matriz de varianzas de las variables originales tipificadas con valores propios asociados escritos en orden decreciente.

Finalmente la interpretación de las componentes principales. Es ésta una de las fases más complejas del análisis en componentes principales. En general, la interpretación de los ejes factoriales (de los componentes principales) se hace a través del estudio de las correlaciones entre las componentes principales y las variables originales.

Se demuestra que estos coeficientes de correlación se obtienen a través de la siguiente expresión:

$$r_{kj}^* = c_{kj} \sqrt{\lambda_k}$$

Donde r_{kj}^* es el coeficiente de correlación lineal entre la k – *ésima* componente principal (c_k) y la j -ésima variable inicial tipificada (Y^j), y c_{kj} es la j -ésima coordenada de la k – *ésima* componente principal.

En efecto, tomando $M = D$ se tiene que:

$$Cov(Y^j, c_k) = \langle Y^j, c_k \rangle_D = Y_j^t D c_k.$$

Sea $Y_j^t = \delta_j^t Y_t$ donde $\delta_j^t = (0, \dots, 0.1_j, 0, \dots, 0)$. Podemos escribir entonces:

$$Cov(Y^j, c_k) = Y_j^t D c_k = \delta_j^t Y^t D c_k = \delta_j^t Y^t D Y v_k = \delta_j^t V_y v_k = \delta_j^t \lambda_k v_k = \lambda_k \delta_j^t v_k = \lambda_k c_{kj}.$$

Y en consecuencia:

$$r_{kj}^* = \frac{Cov(Y^j, c_k)}{\sqrt{Var(Y^j)Var(c_k)}} = \frac{c_{kj}\lambda_k}{\sqrt{\lambda_k}} = c_{kj}\sqrt{\lambda_k}$$

Se trata de observar, por tanto, el círculo de correlaciones. Aquellas variables originales cuya correlación con una componente principal dada esté muy próxima a 1 en valor absoluto, serán las que más contribuyan a la explicación de dicha componente principal.

Puede también ser de interés en algunas ocasiones representar el conjunto de individuos sobre el plano factorial y estudiar los individuos que más contribuyen a la

explicación de un determinado factor. Esto último no tiene sentido más que cuando los individuos tienen un interés en sí mismos.

Además, una vez calculados los coeficientes c_{kj} , se pueden obtener las **puntuaciones factoriales**, es decir, los valores de las componentes correspondientes a cada observación:

$$C_{hl} = C_{h1}X_{1l} + c_{h2}X_{2l} + \cdots + c_{hp}X_{pl} \quad h = 1, 2, \dots, p \quad i = 1, 2, \dots, n.$$

Si una componente se divide por desviación típica se obtiene una componente tipificada. De esta forma, las puntuaciones factoriales tipificadas de la componente h -ésima serán:

$$\frac{c_{hl}}{\sqrt{\lambda_h}}$$

2.5 El modelo de imputación.

2.5.1 El proceso de imputación de datos.

El proceso de imputación supone que si un indicador de bienestar W depende de la variable de interés (ingreso per-cápita del hogar y_h) a partir de una encuesta de hogares y el censo de población y vivienda se puede obtener la distribución conjunta de y_h y una serie de variables independientes x_h . El método consiste en tomar a la encuesta de hogares como una muestra aleatoria de la población representada en el censo, de manera que las variables independientes x_h , se encuentren disponibles en ambas fuentes. De esta manera se construyen estimaciones puntuales y errores de

predicción del vector y_h . En primera instancia, se obtiene un modelo de predicción. Así, la variable y_{ch} , es el ingreso del hogar h que reside en la comunidad c , se estima linealmente como:

$$\ln y_{ch} = E[\ln y_{ch} | X_{ch}] + u_{ch} \quad (1)$$

en donde el vector de errores se distribuye como $u \sim \Gamma(0, \Sigma)$. El vector β carece de toda interpretación económica. El vector β podría estar sesgado dado que algunas de las variables consideradas pueden ser endógenas. Por otro lado, si éstas contribuyen a reducir el error de predicción, pueden permanecer en la modelación de los ingresos del hogar. El error se puede desagregar en:

$$u_{ch} = \eta_c + \varepsilon_{ch} \quad (2)$$

En donde η_c corresponde al error de la comunidad c y ε_{ch} corresponde al error intrínseco del hogar h que vive en la comunidad c .

Las especificaciones en (1) y (2) forman conjuntamente un *modelo lineal jerárquico*, Wooldridge (2003). Es necesario asumir que ambos componentes en (2), el correspondiente a la comunidad y el que se refiere al hogar, no están correlacionados entre sí. Si una mayor proporción del error total corresponde al factor de localidad, las estimaciones serán de menor precisión y no se ganará demasiado por agregar más hogares en la localidad. La forma alternativa de reducir el componente inherente a la

localidad consiste en agregar a la estimación de los ingresos, variables que correspondan a ésta y que no estén relacionadas con otras localidades, variables que no solo expliquen la condición de ciertos niveles de ingreso por el hecho mismo de pertenecer a dicha localidad, sino también que logren capturar la heterogeneidad entre las localidades.

Esta metodología se concentra en la minimización del valor de los dos componentes del error total de predicción en (2). Adicionalmente, el modelo permite corregir por heteroscedasticidad de los errores inherentes al hogar (idiosincráticos).

CAPÍTULO 3

METODOLOGÍA

3.1 Hipótesis.

Ht_{E1}: Existen municipios con servicio de agua potable y alcantarillado deficiente.

Ht_{E2}: Se conocen los factores determinantes en el abastecimiento del servicio de agua potable por medio de red de distribución, río, pozo, hidrante o servicios particulares.

La metodología elegida para llevar a cabo la presente investigación es la propuesta por los autores Hernández (2010) que consideran los siguientes pasos:

- Diseño de la investigación
- Selección de la muestra
- Recolección de datos
- Análisis de resultados y conclusiones

En el caso particular de la investigación realizada, no se realizará selección de la muestra ni recolección de datos, toda vez que la información se tomo de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2006, 2008 y 2010, así como del Censo de Población y Vivienda 2010.

3.2 Diseño de la investigación

Es cuantitativo, dado que “los estudios correlacionales miden el grado de relación entre dos o más variables” (ibídem, p. 121) donde se evaluarán los comportamientos

de los participantes dentro de las situaciones diseñadas mediante ambientes simulados.

Es una investigación transeccional, dado que la información fue recolectada en un solo momento y en un tiempo único, cuyo propósito es describir variables y analizar su incidencia e interrelación en un momento dado (ibídem, p. 151).

Dentro de sus alcances (Danhke, 1989 citado por Hernández et al., 2010, p.117), se considera un estudio descriptivo pues “busca especificar propiedades, las características y los perfiles importantes de personas, grupos comunidades o cualquier otro fenómeno que se someta al análisis”

El diseño de la investigación propuesta es por su naturaleza experimental, Hernández et al, (2010, p188) lo definen como “un estudio en el que se manipulan intencionalmente una o más variables independientes (supuestas causas-antecedentes) para analizar las consecuencias que la manipulación tiene sobre una o más variables dependientes (supuestos efectos consecuentes), dentro de una situación de control para el investigador” es considerado experimento puro de acuerdo a tres características que han de descubrirse según los mismos autores:

- La manipulación intencional de una o más variables independientes
- La manipulación o variación de una variable independiente puede realizarse en dos o más grados. El nivel mínimo de manipulación es de dos: presencia-ausencia de la variable independiente

- Ha de cumplirse con el control de la validez interna de la situación experimental.

Corresponde a la tipología de diseño experimental como preexperimental y exploratorio “se efectúan, normalmente, cuando el objetivo es examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se han abordado antes” Sierra (2001, p. 144).

Es un experimento puro por ser “un estudio de investigación en una situación realista en la que una o más variables independientes son manipuladas por el experimentador en condiciones tan cuidadosamente controladas como lo permite la situación” (Kerlinger, 2002 citado por Hernández et al., p.254)

3.2.1 Definición de variables.

Las variables dependientes son el *abastecimiento del servicio de agua potable y la marginación de los servicios de agua potable y alcantarillado*. Estas variables se evaluaron numéricamente (Sierra, 2001, p.106) por lo cual corresponden a las categorías de cuantitativas, no categóricas, continuas.

La definición de las variables que intervinieron en el experimento se muestra en el tabla 1.

Tabla 1. Definición de variables

No.	Variable	Codificación
1	Viviendas con servicio por red hidráulica	tred
2	Viviendas con abastecimiento por río o pozo.	trío_pozo
3	Viviendas con dotación de servicio de vez en cuando	tdot_dvez
4	Viviendas con abastecimiento por medio de hidrante	thidrante
5	Viviendas sin servicio de drenaje	tsin_drenaje
6	Ingreso mensual del hogar	ing_mes
7	Gasto en energía eléctrica	gas_luz
8	Gasto en agua	gas_agua
9	Total de residentes en el hogar	tot_resi
10	Presencia de analfabetismo en el hogar	analf_hog
11	Zona habitacional	zona (0 = rural, 1 urbana)
12	Sistema de alcantarillado en el hogar	Condrenaje
13	Energía eléctrica en el hogar	Conluz
14	Índice de marginación	Conapo
15	Grado máximo de estudios del jefe de familia	grado_max

Fuente: Elaboración propia.

3.3 Selección de la muestra

El presente estudio de investigación considera fuentes de información secundaria para sus inferencias y deducciones por lo tanto la muestra utilizada es la totalidad de los datos de la Encuesta Nacional de Ingreso Gasto en los Hogares y del Censo de Población y Vivienda, ambos para el año 2010.

3.3 Recolección de datos

Al tratarse de información de fuentes secundarias, como lo es el INEGI, se adoptarán como validos los instrumentos que se utilizaron en su momento para realizar el levantamiento de la muestra, así como la información resultante recuperada gratuitamente del portal del INEGI en la dirección electrónica:

<http://www.inegi.org.mx/est/contenidos/Proyectos/Encuestas/Hogares/regulares/Enigh/Enigh2010/tradicional/default.aspx>.

3.4.1 Preparación de datos.

Los datos fueron procesados mediante el software estadístico Stata versión 11.1, la simulación de los parámetros resultantes y prueba de políticas se efectuó utilizando la rutina Clarify de Stata (Ver anexo A).

Para la explotación de las base de datos de la ENIGH y el Censo de Población y Vivienda 2010, se aplicó el modelo relacional, con esto, se lograron consultas eficientes y sin redundancia (ver anexos E y F respectivamente).

La estructura fundamental del modelo relacional es la relación, es decir, una tabla bidimensional constituida por filas y columnas. Las relaciones representan las entidades que se consideran interesantes en la base de datos.

Cada instancia de la entidad encontrará sitio en una fila de la relación, mientras que las columnas de la relación representan las propiedades de la entidad.

El modelo relacional evitó la duplicidad de campos, y redujo la utilización de campos vacíos, teniendo como resultado que las consultas fueran más rápidas y de una manera más eficiente.

Al agrupar la información en tablas y por tema, se realizaron consultas más pequeñas sin tener que utilizar un gran número de condicionantes.

3.5 Aplicación del método

3.5.1 El modelo de componentes principales.

El crecimiento demográfico es afectado por procesos múltiples de orden social, político, económico y cultural. Este efecto exige respuestas multisectoriales, orientadas a dar solución a este fenómeno con el objeto de elevar la calidad de vida de los habitantes.

La construcción del *índice de marginación del agua y alcantarillado* tuvo como finalidad medir la intensidad y severidad de la carencia del servicio en los municipios de la República Mexicana e identificarlos geográficamente.

3.2.2 Construcción de los indicadores del índice de marginación del agua y alcantarillado.

El cálculo del índice de marginación del agua fue realizado en base a los indicadores censales proporcionados por el Censo de Población y Vivienda 2010, debido a que cuenta con la cobertura, grado de desagregación y actualidad de los datos necesarios para la construcción del índice a escala Municipal.

La obtención de los indicadores que conforman el índice de marginación del agua siguieron el procedimiento descrito a continuación.

3.5.2.1. Porcentaje de viviendas sin servicio de agua potable ($p_{sin_red_{i1}}$).

Para el cálculo de este indicador se obtuvo primero el porcentaje de viviendas con servicio de agua potable y posteriormente se uso el complemento para obtener el porcentaje de viviendas sin servicio de agua potable.

$$p_{sin_red_{i1}} = 100 - \left[\frac{t_{red_i} * 100}{t_{viv_i}} \right] \quad \text{Figura 2.}$$

donde:

t_{red_i} : es el total de viviendas con servicio de agua potable por municipio i .

t_{viv_i} : es el total de viviendas en el municipio i .

3.3.1.2. Porcentaje de viviendas que se abastecen del servicio de agua por río o pozo ($p_{rio_pozo_{i2}}$).

Este indicador se obtiene al dividir el total de viviendas que abastecen en el municipio i del servicio de agua potable por medio de río o pozo entre la suma de todas la viviendas en el municipio i .

$$p_{rio_pozo_{i2}} = \left[\frac{trío_pozo_i * 100}{t_{viv_i}} \right] \quad \text{Figura 3.}$$

Donde:

$trío_pozo_i$: es el total de viviendas con abastecimiento de agua potable por medio de río o pozo en el municipio i .

t_{viv_i} : es el total de viviendas en el municipio i .

3.5.2.3 Porcentaje de viviendas que cuentan con servicio de agua potable de vez en cuando ($p_dot_devez_{i3}$).

Este indicador se obtuvo al dividir el total de viviendas que tienen servicio de agua potable de vez en cuando en el municipio i entre la suma de todas las viviendas en el municipio i .

$$p_dot_devez_{i3} = \left[\frac{tdot_devez_i}{t_viv_i} * 100 \right] \quad \text{Figura 4.}$$

Donde:

$tdot_devez_i$: es el total de viviendas con abastecimiento de agua potable de vez en cuando en el municipio i .

t_viv_i : es el total de viviendas en el municipio i .

3.5.2.4 Porcentaje de viviendas que cuentan con servicio de agua potable por medio de hidrante ($p_hidrante_{i4}$).

Para construir este indicador se dividió el total de viviendas que se abastecen del servicio de agua potable por medio de hidrante en el municipio i entre la suma de todas las viviendas en el municipio i .

$$p_hidrante_{i4} = \left[\frac{thidrante_i}{t_viv_i} * 100 \right] \quad \text{Figura 5.}$$

Donde:

$thidrante_i$: es el total de viviendas con abastecimiento de agua potable por medio de hidrante en el municipio i .

t_viv_i : es el total de viviendas en el municipio i .

3.5.2.5 Porcentaje de viviendas que no cuentan con drenaje (*psindrenaje_{i5}*).

El cálculo del indicador se obtiene al dividir el total de viviendas que no cuentan con drenaje en el municipio *i* entre la suma de todas la viviendas en el municipio *i*.

$$p\sin_drenaje_{i5} = \left[\frac{t\sin_drenaje_i}{t_viv_i} * 100 \right]$$

Figura 6.

Donde:

t_{sin_drenaje_i}: es el total de viviendas con abastecimiento de agua potable por medio de hidrante en el municipio *i*.

t_{viv_i}: es el total de viviendas en el municipio *i*.

3.5.3 El índice de marginación del aguay alcantarillado.

Una vez calculados los cinco indicadores del rezago del servicio de agua potable y alcantarillado, se procedió a construir una medida resumen que condensara la información obtenida utilizando el análisis de componentes principales. Este método estadístico, permite simplificar la información proporcionada por un número determinado de variables en un número menor de variables o "componentes".

Para el cálculo de los componentes principales se puede utilizar la matriz de covarianzas o la matriz de correlaciones. la primera se emplea cuando las variables originales tienen aproximadamente la misma varianza, de forma que el cálculo de los componentes se hace en términos de las variables originales. La segunda se utiliza cuando las escalas de medición de las variables difieren o sus varianzas son

notablemente distintas, en este caso, los componentes principales se obtienen de las variables originales estandarizadas (Harman, 1976). Se siguió la segunda opción para obtener el *índice de marginación del agua y alcantarillado*.

Para la estandarización de los indicadores, se utilizó la siguiente identidad:

$$Z_{ij} = \left[\frac{I_{ij} - I_j}{ds_j} \right] \quad \text{Figura 7.}$$

Donde:

Z_{ij} : Es el indicador estandarizado,

I_{ij} : Es el indicador j de la unidad de análisis i ,

I_j : Es el promedio aritmético de los valores del indicador j , y

ds_j : Es la desviación estándar insesgada del indicador j .

La técnica de componentes principales, transforma el espacio de las variables estandarizadas en uno nuevo, es decir, encontrar Y_k ($k=1,2,\dots,m$) tales que sean combinaciones lineales de las variables estandarizadas.

En términos matriciales, puede esto puede expresarse como:

$$Y=ZA \quad \text{Figura 8.}$$

En esta identidad, Z es la matriz de datos estandarizados, A es la matriz de coeficientes que transforman el espacio definido por los valores Z en uno ortonormal. Y representa las nuevas variables transformadas, las cuales se conocen como componentes principales.

Para el cálculo de componentes principales se utilizó el paquete estadístico Stata versión 11.1, considerando los indicadores estandarizados construidos.

Las estadísticas iniciales del análisis de componentes principales se presentan en el tabla 2, en el que se mostró la varianza atribuible a cada componente en la columna *valor propio* y su importancia relativa en el porcentaje de *varianza explicada*.

El criterio para determinar el número de componentes es considerar el porcentaje de varianza explicada por cada componente. Se extrajo solo un componente, pues explica el 44.59 por ciento del total de la varianza.

Tabla 2. Estadísticas principales del análisis de componentes principales.

Componente	Valor propio	% de varianza explicada	% acumulado
1	2.22946	0.44590	0.44590
2	0.98749	0.19750	0.64340
3	0.91168	0.18230	0.82570
4	0.55247	0.11050	0.93620
5	0.31891	0.06380	1.00000

Fuente: Estimación propia con base en el Censo de Población y Vivienda 2010

La tabla 3, muestra los coeficientes de correlación de cada uno de los indicadores con el componente uno, al que denominamos *índice de marginación del agua y alcantarillado*. Se puede observar que los indicadores muestran una alta correlación con respecto al índice de marginación del agua, tres de ellos presentan una correlación superior a 0.73, el porcentaje de viviendas con servicio de vez en cuando indica 0.24 y el porcentaje de viviendas con servicio por medio de hidrante, 0.40.

Tabla 3.- Coeficientes de correlación entre el índice de marginación del agua y los cinco indicadores.

Dimensión	Indicador	Coefficiente
Vivienda	Porcentaje de viviendas sin servicio de agua potable	0.88210
	Porcentaje de viviendas con servicio por medio de río o pozo	0.73570
	Porcentaje de viviendas con servicio de vez en cuando	0.24150
	Porcentaje de viviendas con servicio por medio de hidrante	0.40840
	Porcentaje de viviendas sin drenaje	0.82760

Fuente: Estimación propia con base en el Censo de Población y Vivienda 2010

El tabla 4, proporciona los coeficientes derivados del análisis de componentes principales para la construcción del *índice de marginación del agua y alcantarillado*. Dicho índice se obtuvo al substituir los coeficientes en la ecuación:

$$IMA_j = \sum C_i Z_{ij}$$

Figura 9.

Donde:

IMA_j : es el índice de marginación del agua para el Municipio j .

C_i : es el valor del coeficiente para el indicador i , y

Z_{ij} : es el valor del estandarizado del indicador i para el Municipio j .

Tabla 4.- Coeficientes del índice de marginación del agua.

Dimensión	Indicador	Coeficiente
Vivienda	Porcentaje de viviendas sin servicio de agua potable	0.59080
	Porcentaje de viviendas con servicio por medio de río o pozo	0.49270
	Porcentaje de viviendas con servicio de vez en cuando	0.16170
	Porcentaje de viviendas con servicio por medio de hidrante	0.27350
	Porcentaje de viviendas sin drenaje	0.55430

Fuente: Estimación propia con base en el Censo de Población y Vivienda 2010

3.5.4 El grado de marginación del agua y alcantarillado.

Para la obtención del ***grado de marginación del agua***, se formaron 6 grupos de acuerdo a la desviación del *índice de marginación del agua y alcantarillado* por subíndice con respecto a la media, siendo los criterios los siguientes:

1. Grado de marginación del agua muy alto: Aquellos Municipios cuyo índice de marginación del agua está dos o más desviaciones estándar por encima de la media.
2. Grado de marginación del agua alto: Aquellos Municipios cuyo índice de marginación del agua está entre una y dos desviaciones estándar por encima de la media.
3. Grado de marginación del agua medio-alto: Aquellos Municipios cuyo índice de marginación del agua está por encima de la media y hasta una desviación estándar.
4. Grado de marginación del agua medio-bajo: Aquellos Municipios cuyo índice de marginación del agua está por debajo de la media y hasta una desviación estándar.

5. Grado de marginación del agua bajo: Aquellos Municipios cuyo índice de marginación del agua está entre una y dos desviaciones estándar por debajo de la media.
6. Grado de marginación del agua muy bajo: Aquellos Municipios cuyo índice de marginación del agua está dos o más desviaciones estándar por debajo de la media.

De esta manera, conforme el tabla 5, los Municipios tienen muy alto, alto, medio alto, medio bajo, bajo y muy bajo índice de marginación del agua. La aplicación de ésta técnica de estratificación muestra que ningún Municipio tiene grado de marginación del agua muy bajo, 19.22 por ciento el de bajo, 35.22 por ciento medio bajo, 28.05 por ciento medio alto, 13.64 por ciento alto y 3.87 por ciento muy alto.

Tabla 5.- Estratificación del índice de marginación del agua.

Grado de marginación	Intervalo	Municipios	Porcentaje	Acumulado
Muy bajo	-2.986274001 , -2.257032000	0	0.00%	0.00%
Bajo	-2.257032000 , -1.493137001	472	19.22%	19.22%
Medio bajo	-1.493137001 , -0.000000001	865	35.22%	54.44%
Medio alto	-0.000000001 , 1.493136999	689	28.05%	82.49%
Alto	1.493136999 , 2.986273999	335	13.64%	96.13%
Muy alto	2.986273999 , 5.727122000	95	3.87%	100.00%
Total		2,456		

Fuente: Estimación propia con base en el Censo de Población y Vivienda 2010

3.5.2 La Georeferenciación del índice de marginación del agua y alcantarillado.

El tabla 6, muestra los Municipios de los Estados de la República Mexicana conforme el grado de marginación del agua. Salta a la vista que los estados que

muestran un mayor número de Municipios con problemas de marginación del agua, se encuentran al sur de la República, mientras que al norte, el servicio de agua potable y alcantarillado mejora considerablemente.

Los Municipios que presentan municipios con muy alto grado de marginación del agua son Oaxaca, Guerrero y Veracruz, con 72, 9 y 6, respectivamente.

Por otro lado, Aguascalientes, Baja California Norte, Baja California Sur, Colima, Quintana Roo y Jalisco, presentan un buen manejo de sus recursos hídricos, al ubicar la mayoría de sus Municipios en el estrato del grado de marginación de medio bajo a bajo.

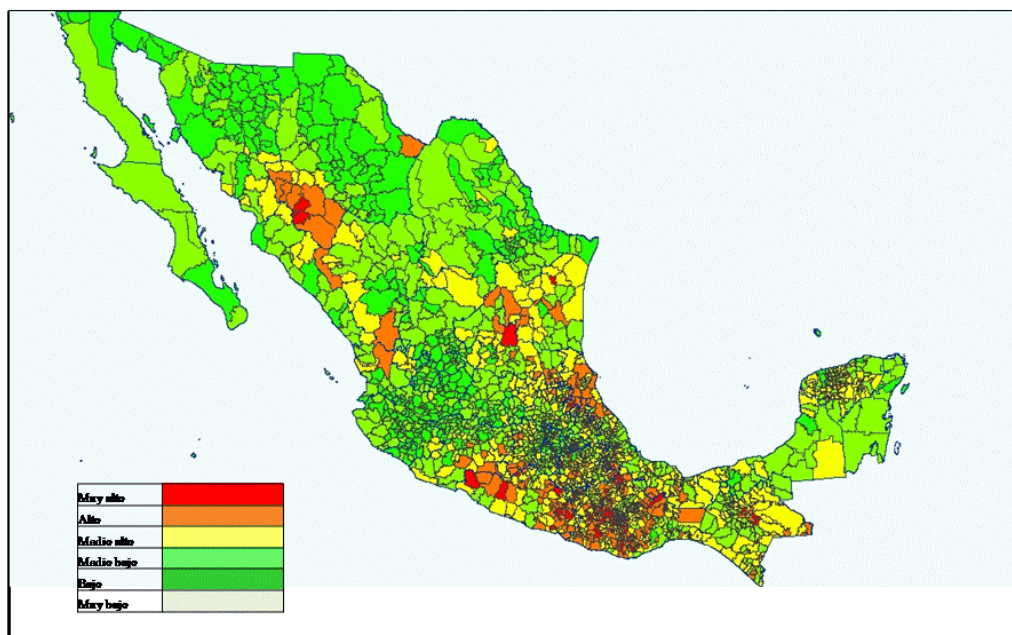
Tabla 6. Número de municipios por estado conforme el índice de marginación del agua.

Estado	Muy alto	Alto	Medio alto	Medio bajo	Bajo	Muy bajo	Total
Aguascalientes	0	0	0	1	10	0	11
Baja California	0	0	0	1	4	0	5
Baja California Sur	0	0	0	4	1	0	5
Campeche	0	0	4	6	1	0	11
Coahuila de Zaragoza	0	0	3	20	15	0	38
Colima	0	0	0	2	8	0	10
Chiapas	3	13	51	51	0	0	118
Chihuahua	2	8	5	16	36	0	67
Distrito Federal	0	0	0	1	15	0	16
Durango	0	2	10	19	8	0	39
Guanajuato	0	0	5	25	16	0	46
Guerrero	9	29	30	13	0	0	81
Hidalgo	0	7	25	43	9	0	84
Jalisco	0	0	7	32	86	0	125
Estado de México	0	6	27	61	31	0	125
Michoacán de Ocampo	0	6	18	55	34	0	113
Morelos	0	1	9	20	3	0	33
Nayarit	0	1	4	8	7	0	20
Nuevo León	0	1	5	16	29	0	51
Oaxaca	72	169	191	124	14	0	570
Puebla	0	35	92	84	6	0	217
Querétaro de Arteaga	0	0	8	6	4	0	18
Quintana Roo	0	0	0	7	2	0	9
San Luis Potosí	2	12	21	20	3	0	58
Sinaloa	0	0	3	12	3	0	18
Sonora	0	0	6	19	47	0	72
Tabasco	0	0	5	9	3	0	17
Tamaulipas	1	4	14	13	11	0	43
Tlaxcala	0	0	1	39	20	0	60
Veracruz de Ignacio de la Llave	6	39	84	73	10	0	212
Yucatán	0	2	58	41	5	0	106
Zacatecas	0	0	3	24	31	0	58
Total	95	335	689	865	472	0	2,456

Fuente: Estimación propia con base en el Censo de Población y Vivienda 2010

La figura 10, muestra la georeferenciación del grado de marginación del agua. Se observó, de manera general que en los estados que se ubican al norte, el grado de marginación es más bajo, en contraste, los municipios de los estados que se encuentran al sur, tienen un grado de marginación del agua considerablemente más alto.

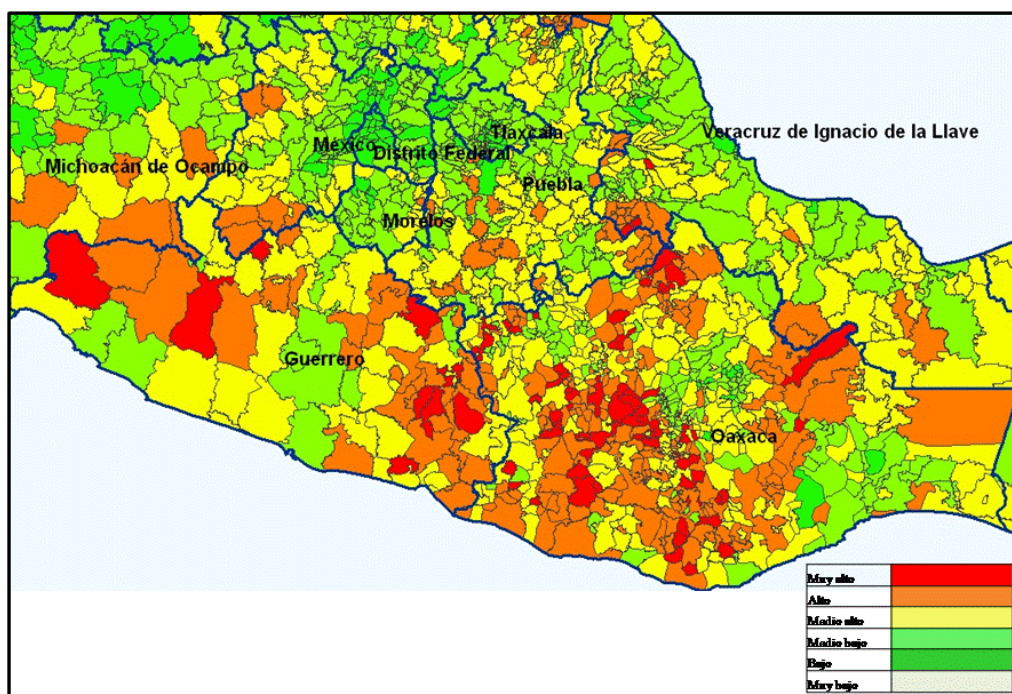
Figura 10: Municipios de la República según grado de marginación del agua.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

El Estado con mayor número de Municipios con deficiencias en los servicios de agua potable y alcantarillado es Oaxaca, al sumar 432 de sus Municipios con grado de marginación del agua de medio alto a muy alto, lo cual representa el 75.78 por ciento del total y 138 se ubican entre medio bajo y bajo grado de marginación del agua que representan el 24.21 por ciento, seguido de Guerrero con 9 municipios clasificados como muy alto y 29 como alto grado de marginación del agua. Esta situación, la podemos apreciar en la figura 11.

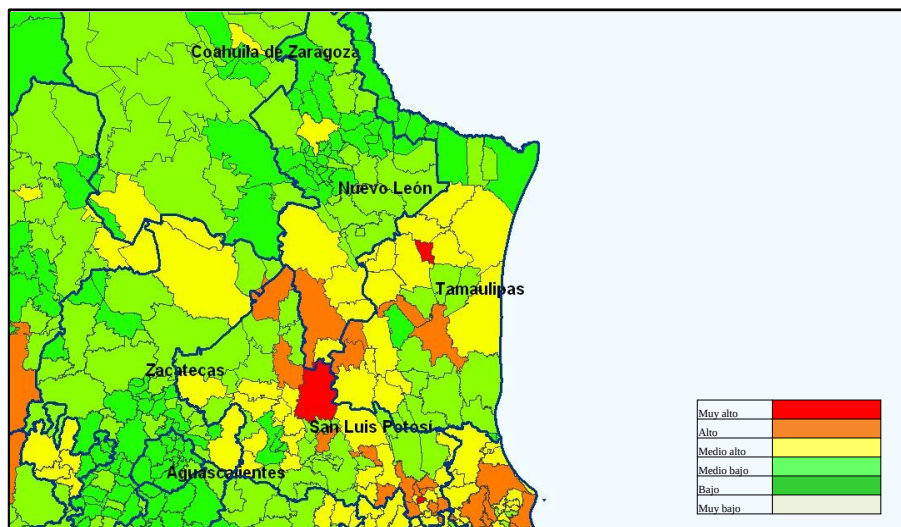
Figura 11: Estados con grado de marginación del agua muy alto.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010.

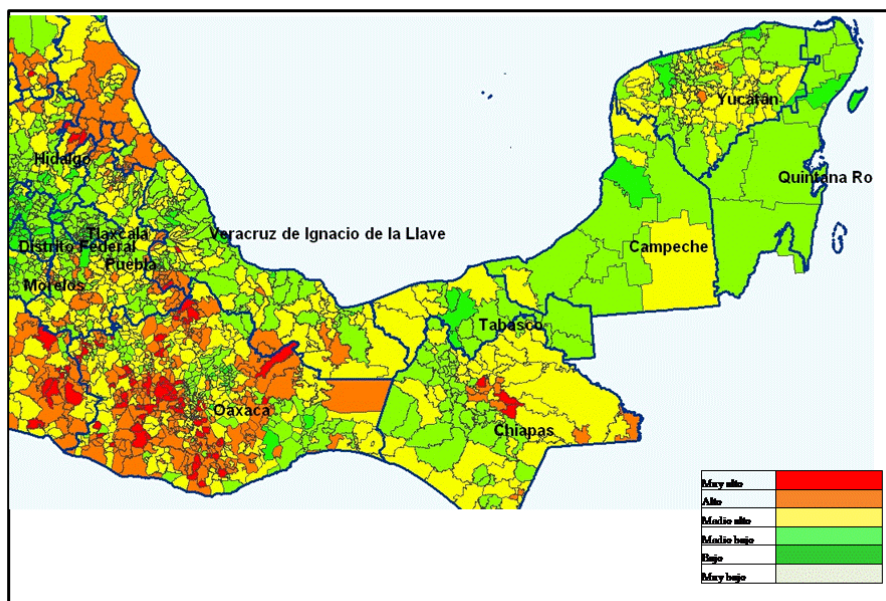
La figura 12, indica que San Luis y Tamaulipas presentan 12 y 4 Municipio con problemas de agua y alcantarillado alto. Por otro lado, la figura 13, indica que Oaxaca y Veracruz, participan en el estrato, con 169 y 39 municipios en este rango.

Figura 12: Estados con grado de marginación del agua muy alto.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

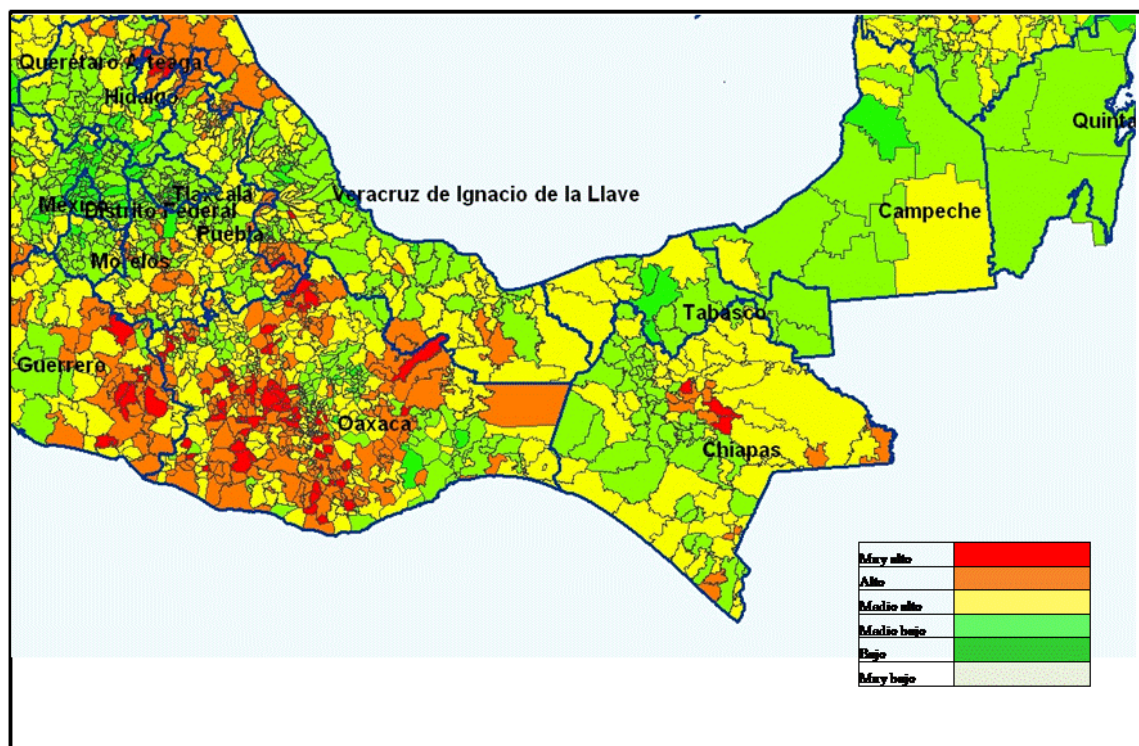
Figura 13: Estados con grado de marginación del agua alto.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

Los Municipios que se encuentran clasificados con grado de marginación medio alto se encuentran en los Estados de Oaxaca, Puebla y Veracruz, con 191, 92 y 84 respectivamente (ver figura 14).

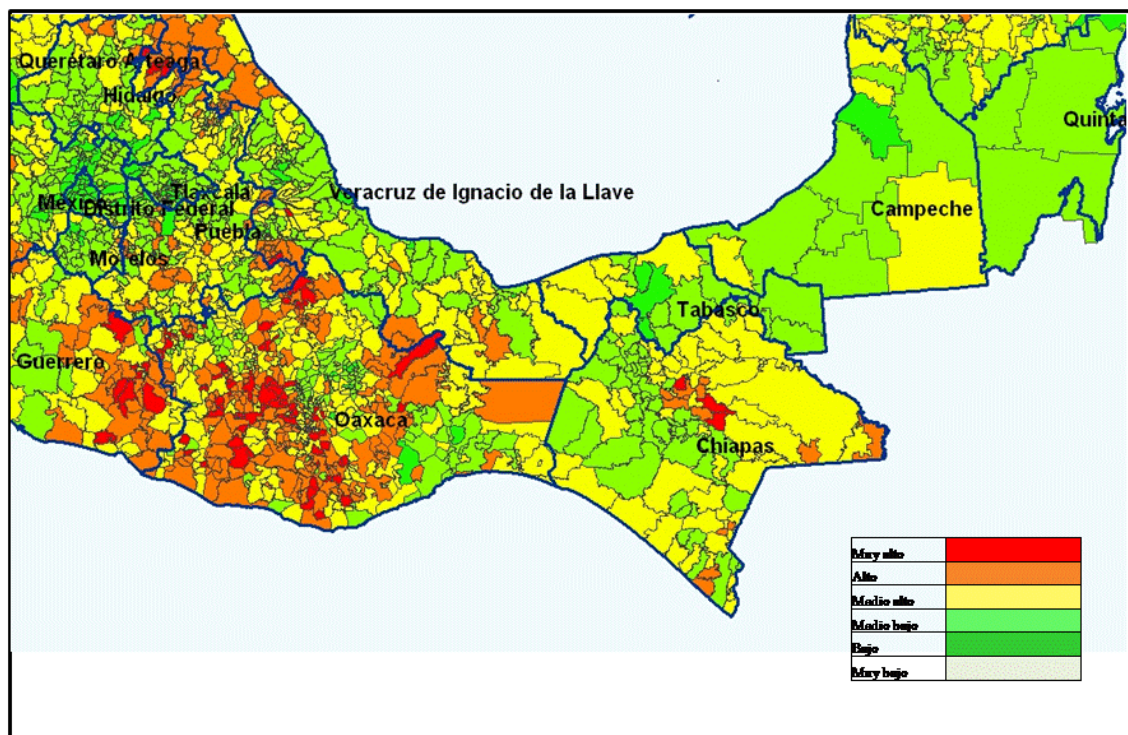
Figura 14: Estados con grado de marginación del agua medio alto.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

Según la figura 15, en el rango medio-bajo, se encuentran los Estados de Oaxaca con 124 Municipios, Puebla con 84 y Veracruz con 73.

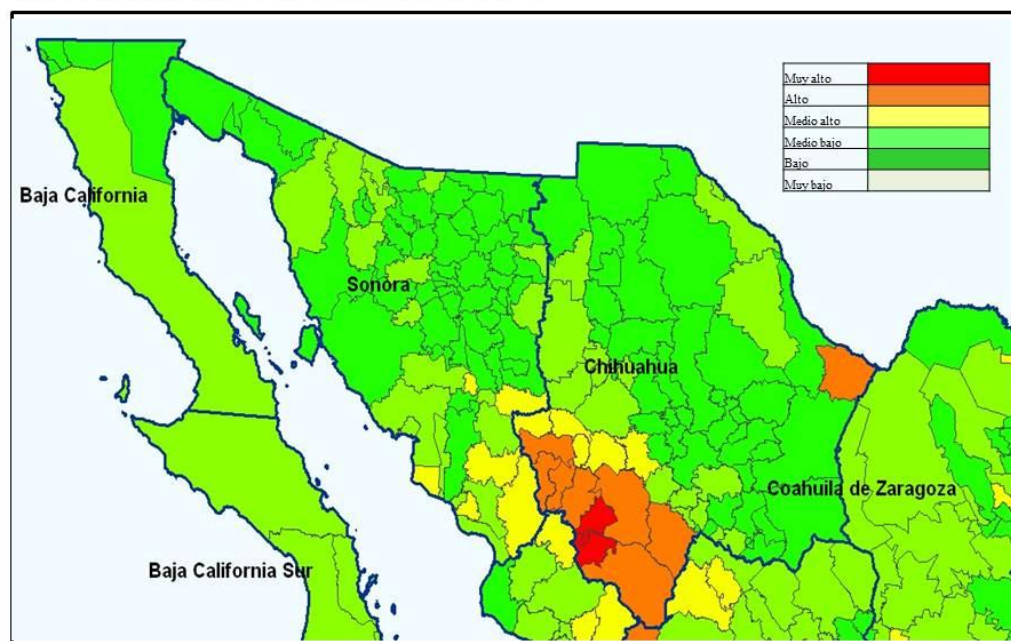
Figura 15: Estados con grado de marginación del agua medio bajo.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

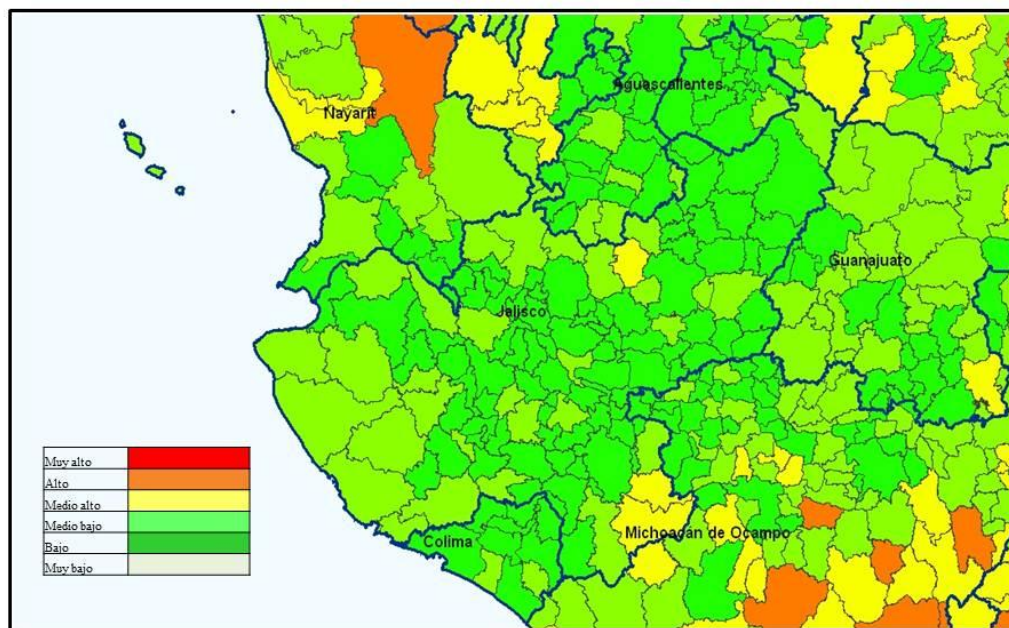
Los Estados que presentan una grado de marginación del agua bajo son Jalisco con 86 de sus 125 Municipios, seguido de Sonora con 47 de un total de 72. Por lo tanto, se consideró que estas entidades son las que tuvieron mejor manejo de los servicios de agua y alcantarillado (ver figuras 16 y 17).

Figura 16 Estados con grado de marginación del agua bajo I.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

Figura 17: Estados con grado de marginación del agua bajo II.



Fuente: Elaboración propia con datos del Censo de Población y Vivienda 2010

3.5.3 Aplicación del modelo multilogístico y simulación.

El tabla 7, muestra la evolución que han tenido los diferentes servicios de agua potable de las variables de estudio según la ENIGH para los años 2006, 2008 y 2010, de acuerdo con la clasificación propuesta.

Se puede apreciar que del 2006 al 2008, los hogares sin servicio, tuvieron una mejora al presentar una incremento negativo de -2.96 por ciento, en contraste, el abastecimiento de agua por río o pozo, se incremento en un 37.43 por ciento. La provisión por medio de pipa e hidrante, bajaron, presentando incrementos negativos de -0.06 y -80.47 por ciento. Un punto de análisis importante se presenta al observar que el servicio por medio de red de agua potable, disminuyó en un -0.20 por ciento,

mientras que hubo un incremento en los hogares del 0.72 por ciento, estos indicadores permiten inferir que la red de distribución del servicio de agua potable, no creció en el mismo tenor que la población.

Para el periodo 2008-2010, se observa un incremento del 1.72 por ciento de hogares sin servicio de agua potable mientras que disminuyó de manera significativa el abastecimiento por agua de río o pozo y pipa, en -31.96 y -0.29 por ciento, respectivamente. Así mismo, los indicadores de suministro de agua potable por medio de hidrante y red de agua potable mejoraron al presentar incremento en su cobertura Nacional del 51.54 y 12.63 por ciento correspondientemente. De manera general, los hogares en México, presentaron un aumento de 8.76 por ciento.

Tabla 7.- Evolución de los servicios de agua potable

Tipos de abastecimiento de agua en los hogares	Hogares			Incremento	
	2006	2008	2010	2006-2008	2008-2010
Sin servicio	568,643	551,801	561,294	-2.96%	1.72%
Río o pozo.	1,562,918	2,147,985	1,461,451	37.43%	-31.96%
Pipa	394,351	394,126	392,981	-0.06%	-0.29%
Hidrante	408,417	79,767	120,875	-80.47%	51.54%
Red de de agua potable	23,606,998	23,558,915	26,537,731	-0.20%	12.64%
Total	26,541,327	26,732,594	29,074,332	0.72%	8.76%

Fuente: Elaboración propia a partir de datos de la ENIGH 2010

Para la aplicación del modelo multilogístico, se realizó una clasificación de los diferentes tipos de servicio que se muestra en el tabla 8.

El abastecimiento del servicio de agua potable en los hogares es excluyente, esto quiere decir, que por hogar, solo se cuenta un tipo de fuente de suministro de agua

para consumo humano, motivo por el cual, no se consideran las demás posibles combinaciones de abasto de agua en el estudio.

Tabla 8. Valores de la variable dependiente

Combinación	Red	Hidrante	Pipa	Río/pozo
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
5	0	1	0	0
9	1	0	0	0

Fuente: Elaboración propia a partir de datos de la ENIGH 2010.

En este apartado del estudio, la variable dependiente es el comportamiento observado de la cobertura del servicio de agua potable, y las variables independientes definidas, se observan en el tabla 9.

Tabla 9, Variables utilizadas en el modelo multilogístico

No.	Variable	Codificación
1	Ingreso mensual del hogar	ing_mes
2	Gasto en energía eléctrica	gas_luz
3	Gasto en agua	gas_agua
4	Total de residentes en el hogar	tot_resi
5	Presencia de analfabetismo en el hogar	analf_hog
6	Zona habitacional	zona (0 = rural, 1 urbana)
7	Sistema de alcantarillado en el hogar	Condrenaje
8	Energía eléctrica en el hogar	Conluz
9	Índice de marginación	Conapo
10	Grado máximo de estudios del jefe de familia	grado_max

Fuente: Elaboración propia.

3.5.3.1 Aplicación del modelo multilogístico.

Los diferentes modelos de regresión resultados de la aplicación del modelo multilogístico se muestran en el tabla 10. Se tomo como base de comparación la variable dependiente Red (abastecimiento de agua potable por red), los resultados indican que solo en el caso de la variable dependiente “sin servicio”, las variables explicativas condrenaje y conluz, no son significativas para el modelo, en los demás casos, todas contribuyen a la descripción de los diferentes modelos con un nivel de significancia del 5%.

Tabla 10. Modelos generados por el algoritmo multilogístico.

Abastecimiento	Variables	Coef.	Std. Err.	z	P> z	[95% Conf.	Interval]
Sin servicio	ing_mes	0.000	0.000	-81.210	0.000	0.000	0.000
	gas_agua	0.003	0.000	42.370	0.000	0.003	0.003
	gas_luz	0.010	0.000	164.050	0.000	0.010	0.011
	tot_resi	-0.311	0.002	-162.750	0.000	-0.315	-0.307
	analf_hog	0.386	0.006	60.460	0.000	0.373	0.398
	zona	1.642	0.008	212.840	0.000	1.627	1.657
	condrenaje	-92.247	72096.820	0.000	0.999	-141399.400	141214.900
	conluz	-69.973	34561.120	0.000	0.998	-67808.520	67668.580
	conapo	-0.038	0.002	-15.180	0.000	-0.043	-0.033
	grado_max	0.219	0.001	259.840	0.000	0.217	0.221
	_cons	-1.117	0.012	-95.140	0.000	-1.140	-1.094
Río/pozo	ing_mes	0.000	0.000	-82.420	0.000	0.000	0.000
	gas_agua	0.001	0.000	121.830	0.000	0.001	0.001
	gas_luz	-0.001	0.000	-125.400	0.000	-0.001	-0.001
	tot_resi	0.044	0.000	93.390	0.000	0.043	0.045
	analf_hog	0.170	0.002	83.120	0.000	0.166	0.175
	zona	-0.742	0.003	-247.100	0.000	-0.748	-0.736
	condrenaje	-1.113	0.002	-493.500	0.000	-1.117	-1.109
	conluz	-0.580	0.004	-138.970	0.000	-0.588	-0.572
	conapo	-0.408	0.001	-479.730	0.000	-0.410	-0.407
	grado_max	-0.039	0.000	-174.800	0.000	-0.039	-0.038
	_cons	0.458	0.005	89.040	0.000	0.448	0.468
Pipa	ing_mes	0.000	0.000	-51.020	0.000	0.000	0.000
	gas_agua	0.001	0.000	23.010	0.000	0.000	0.001
	gas_luz	0.000	0.000	-42.970	0.000	0.000	0.000
	tot_resi	0.066	0.001	77.210	0.000	0.064	0.067
	analf_hog	0.260	0.004	71.130	0.000	0.253	0.267
	zona	0.737	0.005	162.990	0.000	0.728	0.746
	condrenaje	-2.322	0.004	-557.730	0.000	-2.330	-2.314
	conluz	-0.852	0.007	-125.740	0.000	-0.865	-0.839
	conapo	0.007	0.002	4.260	0.000	0.004	0.010
	grado_max	-0.041	0.000	-108.380	0.000	-0.042	-0.040
	_cons	-2.259	0.009	-244.130	0.000	-2.277	-2.241
Hidrante	ing_mes	0.000	0.000	-31.140	0.000	0.000	0.000
	gas_agua	0.002	0.000	77.730	0.000	0.001	0.002
	gas_luz	-0.002	0.000	-74.120	0.000	-0.002	-0.002
	tot_resi	0.046	0.002	30.590	0.000	0.043	0.049
	analf_hog	0.189	0.007	28.960	0.000	0.176	0.201
	zona	-1.004	0.009	-118.080	0.000	-1.021	-0.987
	condrenaje	-0.675	0.007	-96.810	0.000	-0.688	-0.661
	conluz	-1.040	0.011	-94.390	0.000	-1.061	-1.018
	conapo	-0.014	0.003	-5.320	0.000	-0.019	-0.009
	grado_max	-0.034	0.001	-48.880	0.000	-0.035	-0.032
	_cons	-2.905	0.015	-194.520	0.000	-2.935	-2.876

Fuente: Elaboración propia a partir de la ENIGH 2010.

3.5.3.2 Escenarios para la simulación.

En el tabla 11, se muestran los diferentes escenarios con los que se analizaron las probabilidades de equipamiento de servicio de agua potable en las viviendas, donde el valor referido como *media*, se refiere a los valores promedio de cada variable.

Para las variables *analf_hog*, *condrenaje*, *conluz* y *zona*, 0 = inexistencia y 1 = existencia.

En el caso de *conapo*, *gas_agua*, *gas_luz*, *grado_max*, *ing_mes* y *tot_resi*, sus valores cambiaran conforme a su rango mínimo y máximo.

Tabla 11. Escenarios contruidos.

Variable	Primer escenario	Segundo escenario	Tercer escenario	Cuarto escenario	Quinto escenario
<i>analf_hog</i>	1	1	1	1	1
<i>conapo</i>	1	1	1	1	1
<i>condrenaje</i>	0	0	0	1	1
<i>conluz</i>	0	1	1	1	1
<i>gas_agua</i>	0	0	0	0	0
<i>gas_luz</i>	0	Media	Media	Media	Media
<i>grado_max</i>	0	0	0	Media	Media
<i>ing_mes</i>	Media	Media	Media	Media	Media
<i>tot_resi</i>	Media	Media	Media	Media	Media
<i>zona</i>	0	0	1	0	1

Fuente: Elaboración propia con datos de la ENIGH 2010

Tabla 12. Resultado de la simulación parametrizada según los escenarios definidos en el tabla 11.

Tipos de abastecimiento de agua en los hogares	Primer escenario	Segundo escenario	Tercer escenario	Cuarto escenario	Quinto escenario
Sin servicio	2.67%	0.21%	0.21%	0.00%	0.00%
Río o pozo.	49.77%	35.47%	19.93%	9.93%	4.99%
Pipa	5.95%	3.59%	8.86%	0.29%	0.64%
Hidrante	2.55%	0.92%	0.40%	0.43%	0.17%
Red de de agua potable	39.06%	59.81%	70.60%	89.35%	94.21%

Fuente: Elaboración propia a partir de datos de la ENIGH 2010

Los diferentes escenarios proporcionados por la rutina *estsimp* de *stata*, se muestran en el tabla 12, el primero de ellos indica una probabilidad del 49.77% de abastecerse por medio de río o pozo, en contraste con un 39.06% de obtener el servicio por medio de la red de agua potable, considerando que el hogar ésta en una zona rural, que no se cuenta con servicios de drenaje y energía eléctrica, así como ingresos y número de habitantes en el hogar promedio.

Para el segundo escenario, analizamos las posibilidades de contar con el servicio de agua potable en el hogar, considerando que se cuenta con servicios de energía eléctrica. En la zona rural, disminuye la probabilidad con respecto al primer y escenario de abastecerse por medio de río o pozo y no contar con agua en 14.29% y 2.46%, respectivamente, también observamos un aumento considerable en la posibilidad de contar con servicio de red pública (59.81%). El cambio de zona rural a urbana (tercer escenario), mejora aun más estos indicadores.

El cuarto y quinto escenarios, muestran condiciones muy favorables al incluir servicios públicos de drenaje y energía eléctrica al modelo en el abastecimiento por red (89.35% para la zona rural y 94.21% para la zona urbana). En ambos casos, es nula la probabilidad de no contar con el servicio y muy baja de abastecerse con cualquiera de la demás fuentes.

Conforme a los resultados obtenidos en los diferentes escenarios simulados, el contar con servicios de energía eléctrica y drenaje tiene un impacto significativo en el equipamiento de servicio de agua potable en el hogar.

En el tabla 13, se plasman las probabilidades por cambio de grupo derivados de la obtención de estos servicios, el abastecerse por medio de la red de distribución aumenta en un 40.78% con la obtención de servicio de energía eléctrica y drenaje, en el otro extremo, este incremento de los servicios disminuye la probabilidad de no contar con agua potable y abastecerse por medio de río o pozo en un 2.67%. y 31.05% respectivamente.

Tabla 13. Estimación por
cambio de grupo

Servicio	Energía eléctrica y drenaje
Sin servicio	-2.67%
Río/pozo	-31.05%
Pipa	-5.44%
Hidrante	-1.61%
Red	40.78%

Fuente: Elaboración propia con datos de la ENIGH 2010

Dada la relevancia de la variable independiente *ingreso mensual*, en el tabla 14, se muestran los cambios de grupo por percentiles.

Podemos apreciar que mientras se incrementa el ingreso, aumenta la probabilidad de abastecerse por medio de la red de agua potable y disminuye la probabilidad de abastecerse por medio los demás servicios.

Tabla 14. Análisis por cambio de percentiles en el ingreso mensual

Servicio	10 a 20	20 a 30	30 a 40	40 a 50	50 a 60	60 a 70	70 a 80	80 a 90	90 a 100
Sin servicio	-2.93%	-2.96%	-2.99%	-3.02%	-3.05%	-3.08%	-3.11%	-3.14%	-3.17%
Río/pozo	-30.51%	-30.68%	-30.82%	-31.18%	-31.81%	-32.86%	-33.19%	-33.52%	-33.85%
Pipa	-5.56%	-5.56%	-5.56%	-5.56%	-5.55%	-5.52%	-5.57%	-5.63%	-5.68%
Hidrante	-1.63%	-1.64%	-1.64%	-1.66%	-1.69%	-1.73%	-1.75%	-1.76%	-1.78%
Red	42.51%	42.65%	42.57%	42.69%	42.82%	43.04%	43.47%	43.90%	44.34%

Elaboración propia con datos de la ENIGH 2010

Las estimaciones de los cambios de grupo se hicieron con simulaciones Monte Carlo con base en los resultados de la regresión multilogística para la que se corrieron 10,000 muestras aleatorias a partir de la distribución normal con media y varianza .

Con estos resultados, se fijaron valores para la media de cada uno de los grupos y, posteriormente, se efectuaron simulaciones para cada muestra aleatoria, considerando un nivel de confianza del 95% (Ver anexo A).

3.5.4 Aplicación al Estado de Oaxaca.

La estratificación y mapeo del grado de marginación del agua potable y alcantarillado, indicó que los Municipios del Estado de Oaxaca son los que presentan la mayor deficiencia en estos servicios.

3.5.4.1 Aplicación del modelo de imputación de datos

Con el objeto de conocer más a detalle la situación del Estado, se propuso la aplicación del modelo multilogístico y simulación, para tratar de determinar los parámetros que permiten u obstaculizan el equipamiento en los hogares del servicio de agua potable y alcantarillado. Sin embargo, la muestra de la ENIGH 2010 para Oaxaca no es representativa. En atención a esta problemática se realizó el proceso de imputación de los ingresos de la ENIGH 2010 al Censo de Población de ese mismo

año propuesto por Elbers et al. (2003), dado que este variable constituye uno de los factores determinantes en la existencia de los servicios en los hogares.

La primera etapa del proceso de imputación es la selección de variables comunes entre el Censo y la ENIGH, para una mayor precisión, es deseable que el levantamiento de ambas fuentes sea del mismo año.

Las variables seleccionadas se sometieron a una comparación estadística de sus distribuciones. Se requirió una comparación de sus medias muestrales con respecto a la media poblacional con base a pruebas de significancia estadística. Las variables aceptadas como iguales a la población, fueron consideradas para el modelo.

Para el caso de Oaxaca, la lista de variables sociodemográficas y de vivienda estadísticamente comparables entre la ENIGH 2010 y el Censo de Población y Vivienda 2010 se muestran en la tabla 15:

Tabla 15. Lista de variables comparables.

No.	Variable	Codificación
1	Presencia de analfabetismo en el hogar	Analf_hog
2	Sistema de alcantarillado en el hogar	ConDrenaje
3	Energía eléctrica en el hogar	ConLuz
4	Disponibilidad del agua en el hogar	Dis_agua
5	Edad del jefe de familia	Edad_jefe
6	Grado máximo de estudios del jefe de familia	Grado_jefe
7	Grado máximo de estudios de cualquier miembro de la familia	Grado_max
8	Sexo del jefe de familia	Sexo_jefe
9	Total de residentes en el hogar	Tot_resi
10	Ingreso per cápita del hogar	Ing_pp

Fuente: Elaboración propia.

A partir de las variables seleccionadas se construyó el conjunto de variables explicativas para los Municipios del Estado de Oaxaca. El modelo final, indica una R^2 de 0.345, la cual se considera aceptable tomando como base los resultados en los documentos consultados. Por ejemplo, Elbers y otros (2004) presentan un poder explicativo de 0.30 en sus estimaciones para Madagascar, Duran Benavides y Noguera (2001) presentan R^2 que va desde 0.31 a 0.63 para las 7 regiones de Nicaragua.

En la segunda etapa, se realizó una transformación logarítmica al ingreso per cápita de la ENIGH 2010, una vez seleccionadas las variables explicativas, se realizó la estimación de un modelo de regresión por mínimos cuadrados generalizados, considerando como variable dependiente los ingresos per-cápita del hogar y como variables independientes, las variables comunes entre la ENIGH y el Censo de Población y Vivienda, ambos para el año 2010. Los parámetros estimados de esa regresión se utilizaron para predecir el logaritmo del ingreso per cápita para cada hogar en el Censo. Esto se realizó con base en las mismas variables independientes consideradas en el modelo de ingresos estimado en la encuesta, pero ahora seleccionadas en el Censo.

La tercera etapa comienza con un modelo de asociación del ingreso per cápita del hogar para un hogar h en la ubicación c , en donde las variables explicativas son un conjunto de características observables, como se menciona en el modelo (1). En suma, se estiman diez variantes del modelo (1).

Las ubicaciones corresponden a los clusters de la encuesta como se definen en un esquema de muestreo típico de dos etapas, y el logaritmo del ingreso per cápita del hogar se modela empíricamente como:

$$Ln y_{ch} = x_{ch}\beta + u_{ch} \quad (3)$$

que se deriva del modelo teórico (1). El vector de errores u se distribuye como $F(0, \Sigma)$. El modelo se estima por mínimos cuadrados generalizados usando los datos del hogar de la ENIGH. Para poder estimar el modelo de regresión, se obtiene la matriz de varianzas y covarianzas.

El error se modela como se muestra en la especificación (2). Esta estructura de errores permite modelar la auto-correlación espacial, o “efecto de localidad” para hogares de la misma área geográfica, así como la heteroscedasticidad en el componente idiosincrático del error. Ambos componentes son independientes entre sí.

Para calcular Σ se construyó un modelo de *ingreso* estimado mediante regresión lineal y controlando según el diseño muestral de la encuesta, por las unidades primarias de muestreo y por el estrato. La variable dependiente del modelo es el logaritmo del ingreso neto total per cápita y las variables independientes son aquellas variables comparables entre las distintas fuentes de información. La selección de las variables explicativas se realizó conforme el nivel de significancia,

eliminando aquellas que no son significativas o que tienen poca contribución a la R^2 ajustada.

Construido el mejor modelo utilizando (3), se aplica una prueba F , considerando el factor de expansión. Si no se acepta la hipótesis nula de que los parámetros son conjuntamente iguales a cero, entonces se estima el modelo con factores de expansión. En caso de no rechazar la hipótesis nula, la estimación se realiza sin considerar el factor de expansión.

Los errores totales son estimados mediante los residuales del modelo u_{ch} , que al descomponerse en sus dos elementos no correlacionados entre si, se obtuvo:

$$\hat{u}_{ch} = \hat{\eta}_c + \varepsilon_{ch} \quad (4)$$

Los componentes de ubicación estimados del error, dados por $\hat{\eta}_c$, son la media de residual de cada cluster. Las estimaciones del componente de hogar ε_{ch} son los errores totales menos el componente de ubicación. También se estimaron las varianzas correspondientes y la matriz de varianzas-covarianzas.

Con la especificación del modelo (3), se incluyen variables a nivel cluster para minimizar el efecto a nivel localidad y disminuir los errores estándar de la estimación.

Los parámetros estimados junto con las características observables de cada hogar del Censo de Población y Vivienda son utilizados para simular los errores.

Para cada simulación r se utiliza un conjunto de parámetros de la primera etapa. Se obtuvieron un conjunto de coeficientes α y β , de las distribuciones normales multivariadas descritas por los estimadores puntuales de la primera etapa y sus matrices de varianza-covarianza asociadas. El valor simulado de la varianza del error del componente del Municipio, σ^2_{η} , se asume con una distribución gamma con varianza $V(\sigma^2_{\eta})$.

Se obtuvieron términos de error simulados y sus distribuciones correspondientes, se simuló el ingreso para cada hogar y_{ch} , que es utilizado para estimar los resultados del modelo multilogístico y la simulación.

La simulación se realizó con una frecuencia de cien repeticiones para obtener nuevos coeficientes de $\alpha, \beta, \sigma^2, \eta$, así como términos de error para cada una.

En resumen, en las estimaciones existen dos componentes principales que conforman al error de la estimación:

En primer lugar el conocido como “error del modelo” en Elbers, et al. (2003), el cual está asociado a la reducción del error a nivel cluster que depende de la disponibilidad, significancia y nivel de predicción de las variables a nivel Municipal sobre los ingresos del hogar.

El segundo componente, llamado “error idiosincrático”, se asocia con la representación de la heterogeneidad de los ingresos de los hogares en el modelo especificado. Esto es, se encuentra asociado con la desviación de los valores predichos con respecto a los valores reales observados. El error idiosincrático aumenta conforme el número de hogares en una región geográfica disminuye.

3.5.4.2 Análisis descriptivo de los ingresos imputados

En el tabla 16, podemos observar que el promedio de los ingresos imputados es de \$1,710.60, el valor mínimo encontrado fue de \$0.58, el máximo, de \$4,342.65, con una desviación estándar de \$854.84, esto implicó un rango de los ingresos imputados es de \$4,342.07. Considerando una desviación estándar, los hogares tuvieron ingresos desde \$855.76 y hasta \$2,565.44

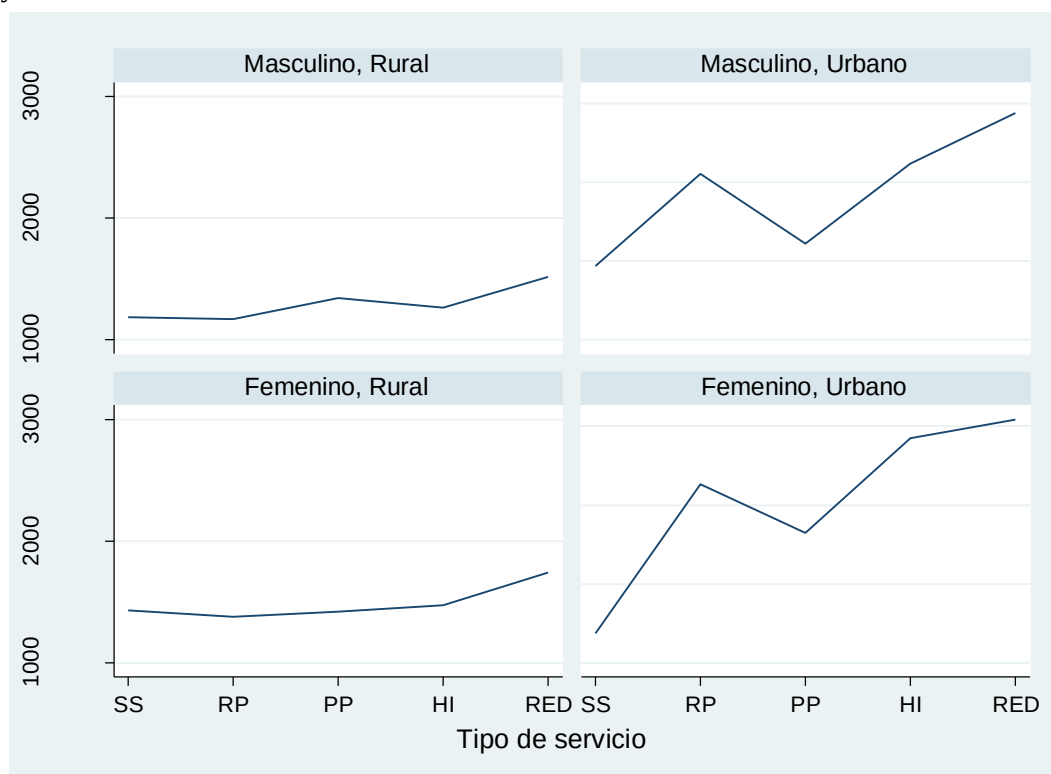
Tabla 16. Estadística descriptiva de los ingresos imputados

Variable	No. De observaciones	Media	Desviación estándar	Mínimo	Máximo
ing_pp_imp	936,588	1,710.60	854.84	0.58	4,342.65

Fuente: Elaboración propia a partir de los ingresos per cápita imputados

La gráfica 1, presenta los ingresos imputados por tipo de servicio de agua potable, controlados por sexo y región. Salta a la vista que los ingresos son una variable de gran influencia, el gráfico indicó que los hogares con menor ingreso, tienen menos posibilidades de abastecerse de agua potable por alguno de los medios existentes, cabe resaltar la influencia de la región en este contexto, ya que observamos que la zona rural se va más afectada en este rubro, por otro lado un ingreso mayor propicia una tendencia al mejoramiento en el abastecimiento en cualquiera de las zonas.

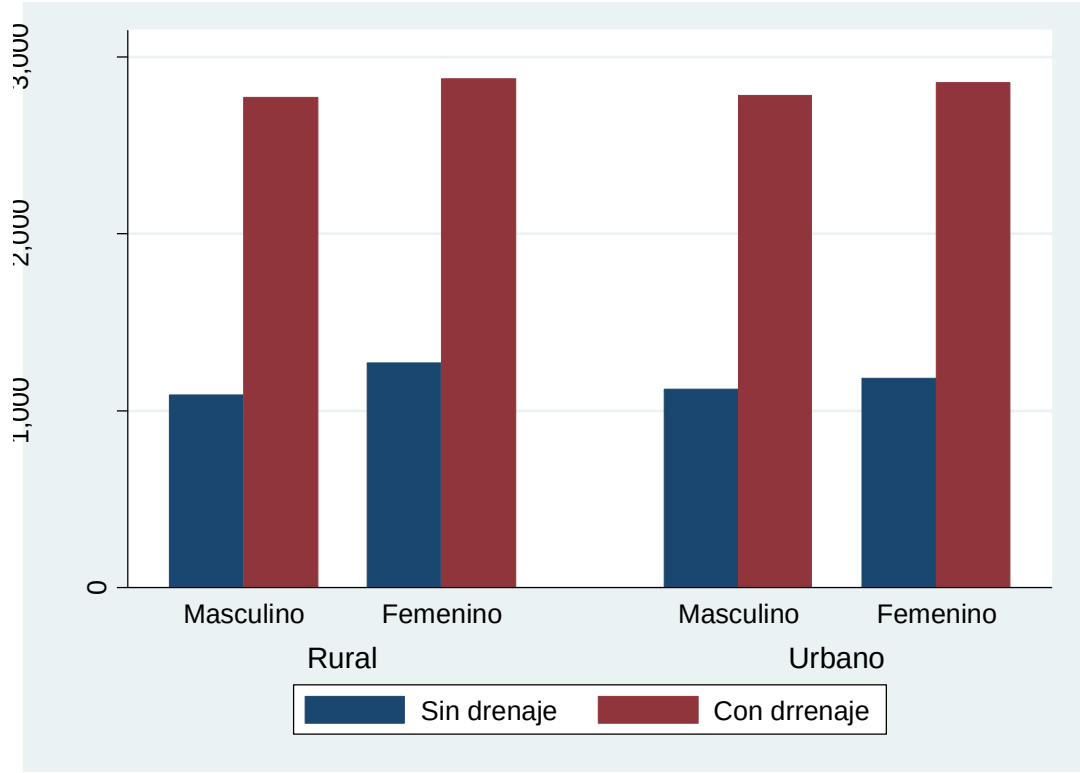
Gráfica. 1. Ingresos imputados por tipo de servicio, controlados por sexo del jefe de familia y zona



Fuente: Elaboración propia a partir de los ingresos per cápita imputados

En la gráfica 2, observamos que la región no es determinante en que los hogares cuenten con el servicio de drenaje, sin embargo, se presenta una mayor frecuencia a favor de contar con el servicio cuando el jefe de familia es del sexo femenino. Nuevamente podemos apreciar la gran influencia que ejercen los ingresos en la obtención del servicio, esto es, mayor ingreso, mayores posibilidades de contar con drenaje en el hogar.

Gráfica. 2. Ingresos imputados por tipo de servicio, controlados por sexo del jefe de familia y zona



Fuente: Elaboración propia a partir de los ingresos per cápita imputados

3.5.4.3 Aplicación de la metodología a los datos imputados.

Para el análisis y aplicación del modelo, se proponen los diferentes escenarios que se muestran en el tabla 17.

Tabla 17. Escenarios contruidos para la aplicación de Estado de Oaxaca.

Variable	Primer escenario	Segundo escenario	Tercer escenario	Cuarto escenario	Quinto escenario
analf_hog	1	1	1	0	0
conapo	1	1	1	3	5
condrenaje	0	0	0	1	1
conluz	0	0	1	1	1
gas_agua	0	0	0	0	1
gas_luz	0	0	Media	1	Media
grado_max	0	0	0	Media	Media
ing_mes	Décimo percentil	Media	Media	Media	Media
tot_resi	Media	Media	Media	Media	Media
zona	0	0	0	1	1

Fuente: Elaboración propia con datos de la ENIGH 2010

Los resultados de la sensibilización del modelo propuestos en el tabla 15 se muestran en el tabla 18. En primer lugar, se consideró analfabetismo en el hogar, un grado de marginación socioeconómica alto, residentes promedio, ingresos mensuales en el décimo percentil, zona rural, si servicios de drenaje, sin energía eléctrica y con el grado mínimo de estudios del jefe de familia. Con estas características el resultado de la simulación indica resultados muy pesimistas, pues que existe un 30.13 por ciento de probabilidad de no contar con el servicio, 34.31 por ciento de abastecerse por medio de río o pozo y 34.16 por ciento de tener servicio por medio de la red pública de agua potable.

La siguiente prueba consistió en solo cambiar el valor de la variable ingreso mensual a su valor promedio. El resultado fue una disminución en la probabilidad de no tener servicio, abastecimiento por río o pozo, pipa e hidrante, de 18.85, 29.59, 2.25 y 0 por ciento respectivamente. En cuanto a la variable dependiente red de agua potable, hubo un aumento favorable hasta 49.31 por ciento. Lo anterior, corrobora la importancia que tienen los ingresos en el equipamiento de servicio de agua potable en las viviendas.

El tercer escenario consistió en suponer los hogares con energía eléctrica y gastos en este rubro promedio. El resultado fue que no existe probabilidad de no contar con servicio de agua potable e hidrante, disminución a 18.76 por ciento de abastecerse por río o pozo y 1.40 por ciento por pipa. Así mismo, la probabilidad de tener

servicio por medio de red, aumento a 74.71 por ciento. Esto indica que existe una correlación directa de contar con servicio de agua potable y energía eléctrica.

En cuanto al cuarto escenario, se cambio la variable de precariedad social conapo a su valor promedio (3), equipamiento de drenaje, gastos en energía eléctrica, nivel de estudios del jefe de familia promedio y zona urbana. Para estas condiciones, no existe posibilidad de no contar con servicio, abastecimiento por medio de pipa e hidrante, una disminución hasta 7.71 por ciento del abastecimiento por río o pozo y un aumento significativo de 91.43 por ciento de tener servicio de agua potable por red.

El quinto escenario consistió en cambiar la variable conapo al nivel 5 (muy baja precariedad), gastos en agua y gasto en luz promedio. las variables sin servicio, abastecimiento por pipa e hidrante se mantuvieron en 0 por ciento, mientras que disminuye la probabilidad de abastecerse por río o pozo y aumentó hasta 97.76 por ciento la probabilidad de contar con servicio por medio de red.

Tabla 18. Resultados de la simulación conforme los escenarios propuestos en la tabla 17.

Tipos de abastecimiento de agua en los hogares	Primer escenario	Segundo escenario	Tercer escenario	Cuarto escenario	Quinto escenario
Sin servicio	30.13%	18.85%	0.00%	0.00%	0.00%
Río o pozo.	34.31%	29.59%	18.76%	7.17%	2.24%
Pipa	6.54%	2.25%	1.40%	0.00%	0.00%
Hidrante	1.40%	0.00%	0.00%	0.00%	0.00%
Red de de agua potable	34.16%	49.31%	74.71%	91.43%	97.76%

Fuente: Elaboración propia a partir de datos de la ENIGH 2010

Considerando la importancia de los ingresos per cápita en el hogar, se calcularon los percentiles del ingreso mensual y se hicieron cambio de grupo por cada 10 percentiles (ver tabla 19). Se apreció que las variables dependientes sin servicio, abastecimiento por pipa e hidrante se mantuvieron prácticamente en cero, mientras que la probabilidad de tener servicio por medio de río o pozo, disminuyo en cada escenario. En contraste, el contar con servicio de agua potable por red, se incremento.

Tabla 19. Análisis por cambio de percentiles del ingreso mensual per cápita imputado.

Tipos de abastecimiento de agua en los hogares	10 a 20	20 a 30	30 a 40	40 a 50	50 a 60	60 a 70	70 a 80	80 a 90	90 a 100
Sin servicio	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Río o pozo.	-0.430%	-0.542%	-0.435%	-0.703%	-0.711%	-0.839%	-1.245%	-1.626%	-2.468%
Pipa	0.000%	0.001%	0.001%	0.001%	0.001%	0.001%	0.002%	0.004%	0.025%
Hidrante	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Red de de agua potable	0.430%	0.541%	0.434%	0.702%	0.710%	0.837%	1.243%	1.622%	2.443%

Fuente: Elaboración propia a partir de datos de la ENIGH 2010

El tabla 20, muestra las probabilidades por cambio de grupo de las variables independientes. La variable analfabetismo, indica que no existe probabilidad de no tener servicio y abastecerse por medio de hidrante, disminución de tener servicio de río o pozo y pipa. El servicio por medio de red muestra una probabilidad de 2.60 por ciento a favor. Contar con drenaje en la vivienda, disminuye las probabilidades de no tener servicio, de tener servicio por medio de río o pozo y pipa. Mientras que no se cuenta con ninguna probabilidad de abastecerse por medio de hidrante y una probabilidad del 10.48 por ciento de tener servicio por medio de red.

El cambio de la variable energía eléctrica arroja una probabilidad negativa de abastecerse por medio de río o pozo y una probabilidad del 6.73 por ciento de abastecerse por medio de red.

La variable grado máximo de educación del jefe de familia (educación), indica que si el jefe de familia tiene el grado máximo en educación, no hay probabilidad de no tener servicio, y de abastecerse por medio de hidrante. Las variables dependientes río o pozo y pipa, muestran probabilidades negativas, mientras que el contar con servicio de la red pública muestra una probabilidad positiva.

El cambio de la variable residentes del hogar de su valor mínimo al máximo, incrementa la probabilidad de abastecerse por medio de río o pozo y tiene un efecto negativo para contar con servicio por medio de la red pública.

El cambio de zona de rural a urbana, muestra que no existe probabilidad de abastecerse por medio de hidrante y de no contar con servicio. Del mismo modo, se muestra una probabilidad negativa de tener servicio por medio de río o pozo (-10.14 por ciento) y de -1.14 por ciento de abastecerse por medio de pipa. En contraste, el abastecimiento por medio de red tiene una probabilidad del 12.23 por ciento.

Tabla 20. Análisis por cambio de grupos.

Tipos de abastecimiento de agua en los hogares	Analfabetismo	Drenaje	Energía eléctrica	Educación	Residentes en el hogar	Zona
Sin servicio	0.000%	-0.043%	0.000%	0.000%	0.000%	0.000%
Río o pozo.	-2.564%	-8.828%	-6.751%	-0.431%	0.178%	-10.145%
Pipa	-0.037%	-1.618%	0.018%	-0.047%	0.003%	-1.141%
Hidrante	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Red de de agua potable	2.601%	10.489%	6.734%	0.479%	-0.182%	12.231%

Fuente: Elaboración propia a partir de datos de la ENIGH 2010

CAPÍTULO 4

ANÁLISIS DE RESULTADOS

4.1 Conclusiones relativas a los objetivos específicos

- Realizar un diagnóstico de la marginación del servicio de agua potable y alcantarillado en el País.

Se construyó el *índice de marginación del agua y alcantarillado* a partir de datos del Censo de Población y Vivienda 2010, aplicando para ello, la técnica de componentes principales, el objeto principal del indicador resultante fue realizar el una clasificación efectiva de los Municipios de la República Mexicana a partir de una menor dimensión de las variables que influyen directamente en la carencia de estos servicios

Utilizando técnicas de georeferenciación, el índice, permitió visualizar ésta problemática que aqueja a la población visualmente utilizando cartografía digital, lo cual proporcionó una visión general de esta problemática a nivel Nacional.

- Describir y analizar los factores socioeconómicos y demográficos que afectan la existencia de agua y alcantarillado en los hogares.

Para alcanzar este objetivo, se utilizó un modelo multilogístico combinado con técnicas de simulación. El resultado del tratamiento de la información, permitió determinar que aunque el nivel de ingresos influye de manera determinante en contar

con el servicio de agua potable y alcantarillado, existen variables socioeconómicas como el nivel de estudios del jefe de familia, el total de residentes en el domicilio y la ubicación geográfica, también tienen un gran peso en el equipamiento de los hogares con estos servicios.

- Probar la metodología propuesta en un Estado con marginación de servicios de agua potable y alcantarillado.

Se aplicó la metodología al Estado de Oaxaca, ya que de acuerdo al *índice de marginación del agua*, 72 de sus municipios tienen un grado de marginación muy alto y 169 alto, los cuales representan el 42.28% de un total de 570. En contraste 138 tienen un grado de marginación del agua y alcantarillado de medio bajo a bajo, 24.21%. Estas condiciones, son las más críticas a nivel Nacional, lo cual justifica la aplicación de la metodología propuesta para su análisis particular.

Previo a la aplicación del modelo multilogístico y la simulación, se realizó la imputación de los ingresos de la ENIGH 2010 a los datos del CENSO para este mismo año. Esto encaminado a resolver el problema encontrado entre la precisión de los ingresos presentados en el Censo y la falta de representatividad de la encuesta.

- Evaluar el impacto de las estrategias y políticas de solución propuestas mediante la simulación.

Se usó la rutina `estsim` del módulo `clarify` de `stata` versión 11.1, con la cual se realizó una simulación de 10,000 repeticiones de los parámetros de la regresión multilogística y con esto, estar en posibilidades de probar con diferentes escenarios las políticas y estrategias encaminadas al diagnóstico y mejoramiento de los servicios de agua potable en los Municipios más afectados a nivel Nacional y para el Estado de Oaxaca.

4.2 Conclusiones relativas al objetivo general

- Diseñar una metodología basada en el análisis de aspectos económicos, geográficos y técnicos que ejercen influencia sobre la existencia de agua urbana y alcantarillado, utilizando como vehículo de análisis la simulación, técnicas de georeferenciación, modelos econométricos de elección múltiple, de imputación de datos y análisis factorial.

Los resultados obtenidos en esta investigación, evalúan y valoran el impacto que tienen las variables socioeconómicas y demográficas proporcionadas por las bases de datos de la ENIGH y el Censo de Población y Vivienda con corte al año 2010, en el sector agua potable y alcantarillado.

Una vez revisadas diferentes publicaciones sobre el tema, se concluyó que la dinámica y relevancia del abasto de los servicios de agua potable y alcantarillado, justifica la aplicación de técnicas multivariadas para el análisis y planteamiento de estrategias para la problemática existente.

La amalgama de las técnicas propuestas, proporciona una visión integral de la situación que guardan los hogares de la República con respecto a los servicios de agua potable y alcantarillado, además de permitir la evaluación del impacto de los resultados de la aplicación de políticas y estrategias a través de un modelo informático.

Los altos resultados estadísticos alcanzados permiten la construcción de una visión de futuro del sector eligiendo las estrategias adecuadas para alcanzarlo.

4.2 Aportaciones originales

Conforme a la literatura consultada a nivel Nacional e Internacional, se afirma que este es el primer estudio de investigación en el que se construye un *índice de marginación del agua y alcantarillado* para análisis del sector agua potable y alcantarillado.

Administración del recurso.- La aplicación de modelos multilogísticos para la obtención del diagnóstico de la situación que guarda el sector agua potable y alcantarillado en los Municipios de la República.

El uso de la simulación para evaluar las estrategias y políticas propuestas para abatir la falta de servicios de agua potable y alcantarillado.

La inclusión de modelos de imputación para la obtener una base de datos para el Estado de Oaxaca con características de representatividad y precisión.

Esta investigación ayuda a la exploración de variables socioeconómicas y demográficas que inciden en un sector tan importante como el agua potable, lo cual sin duda, será aporte para la realización de estudios futuros.

4.4 Límites del modelo planteado

El modelo se limita a los datos obtenidos del Censo de Población y Vivienda y de la ENIGH, ambas fuentes para el año 2010, considerando como válido, el instrumento de medición propuesto por la Institución.

CAPÍTULO 5. RECOMENDACIONES

5.1 Recomendaciones a los Organismos gestores del agua potable y alcantarillado.

Dimensionar los niveles específicos de servicio que existen y se espera alcanzar. Los niveles en el abastecimiento de agua van desde el acceso al agua por medio de río o pozo (sin tratamiento), hasta el suministro de agua potable domiciliar, con medición, continuidad y presión adecuada, lo que deriva en costos, disposición de pago y administración. En alcantarillado, los niveles varían desde disposición de excretas a campo abierto, hasta el acceso a redes colectoras de alcantarillado sanitario.

Implementar sistemas de información que permitan la cuantificación de los recursos invertidos en obras orientadas a la ampliación de la cobertura o mejoramiento de los servicios, separando las inversiones hechas en las zonas rural y urbana. De esta forma, se estará en posibilidad de analizar políticas públicas y estrategias vinculadas con el impacto social alcanzado y la sostenibilidad de los Organismos debido a la inversión realizada.

Establecer subsidios para poblaciones vulnerables cuya capacidad económica le impiden cofinanciar las inversiones para la ampliación de infraestructura.

Implementar dentro de las Instituciones una estructura organizacional cuyas funciones sean las de promover políticas públicas, planeación y soporte técnico.

Establecer esquemas tarifarios sencillos que garanticen la suficiencia financiera y económica de la prestación del servicio.

Crear el Centro de Investigaciones del Agua, con funciones de carácter consultivo, de investigación, así como de certificación.

Establecer como estrategia para la recuperación de caudal, la ampliación de la cobertura de la macro y micro medición del suministro de agua potable logrando así un mayor control y disminución del consumo.

5.2 Recomendaciones relativas a los resultados obtenidos.

La georeferenciación del índice de marginación del agua y alcantarillado, permitió identificar que los Municipios que se encuentran al sur del País, presentan carencias importantes en el abastecimiento de estos servicios. Motivo por el cual se recomienda a los Gobiernos Federal, Estatal y Municipal, la instauración de planes y proyectos relacionados con las vocaciones productivas laborales que impulse a una mejor calidad de vida a los habitantes de esta zona.

La simulación aplicada a la información extraída de la ENIGH 2010, identificó que la variable mas significativa en el equipamiento de los hogares de los servicios de agua potable y alcantarillado es el ingreso, motivo por el cual, los Gobiernos Federal y Estatal deberán implementar planes y proyectos encaminados a la sustentabilidad

Otro factor de impacto es el nivel de estudios del jefe del hogar, por lo cual, se propone como estrategia, el desarrollar programas de alfabetización, educación primaria y secundaria y de capacitación para el trabajo. Así mismo, establecer vínculos con los diferentes centros educativos a nivel medio superior y superior con el fin de establecer esquemas que faciliten el acceso a la población para la realización de estudios en esos niveles.

5.3 Recomendaciones generales.

Este estudio es un esfuerzo inicial dirigido a entender el efecto que tienen las disparidades socioeconómicas y demográficas en las zonas urbana y rural sobre la distribución del agua potable y el servicio de alcantarillado en México e intenta fortalecer la discusión sobre estas desigualdades principalmente al sur de la República Mexicana.

Es innegable la relación que guardan los ingresos en el equipamiento del servicio de agua potable y alcantarillado sanitario motivo por el cual el Gobierno Federal, los gobiernos Estatales y los Municipios deberán implementar estrategias y alianzas colaborativas para que el vital líquido llegue a los sectores más marginados del País impulsando su desarrollo sustentable.

Utilizar las técnicas propuestas para analizar y probar estrategias de mejora haciendo muestreos propios en Municipios de interés sobre aspectos de importancia como pueden ser el sistema tarifario, proponer el uso de sistemas de información geográfica para la planeación y control del sistema de red hidráulica y simulación para construir escenarios que permitan realizar pruebas anticipadas de propuestas políticas orientadas hacia la demanda de los servicios de agua potable y alcantarillado sanitario.

REFERENCIAS

- Alderman H., M. Babita, G. Demombynes, N. Makhatha and B. Özler, (2002). "How long Can You Go?: Combining Census and Survey Data for Mapping Poverty in South Africa". *Journal of African Economics*, forthcoming.
- Aluja, T., Morineau, A. (1999) *Aprender de los datos: el análisis de componentes principales, una aproximación desde el data mining*. EUB, Barcelona.
- Anderson, D. Sweeney D. & Williams, T. (2004) *Estadística para administración y economía* (8ª. ed.). México: Thomson.
- Anderson, M. J. and T.J. Willis (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, 84, 511-525.
- Anderson, T. W. (1958) *An Introduction to Multivariate Analysis*. Wiley, N. York.
- Angrist, J. and A. Krueger, (1992). "The Effect of Age of School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples". *Journal of the American Statistical Association*. Vol. 87, Pp. 328-336.
- Anselin, Luc (1988), *Spatial Econometrics: Methods And Models*, Netherlands, Kluwer Academic Publishers.
- Arellano, M. and C. Meghir, (1992), "Female Labour Supply and on the Job Search: an Empirical Model Estimated Using Complementary Data Sets", *Review of Economic Studies* 59, pp. 537-559.
- Arrojo, P. (2003). *El Plan Hidrológico Nacional: una cita frustrada con la historia*. Barcelona: RBA Libros.

- Axelrod, R. M. (1997). Advancing the Art of Simulation in the Social Sciences. En Conte, R., Hegselmann, R. & Terna, P. (Eds), *Simulating Social Phenomena, Lecture Notes in Economics and Mathematical Systems* Berlin: Springer-Verlag.
- Banco Mundial (2007), datos y Estadísticas, PIB per cápita, [en línea], <http://www.bancomundial.org/datos/>
- Barreteau, O., Bousquet, F. & Attonaty, J. M. (2001). Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation*. Disponible en <http://jasss.soc.surrey.ac.uk/4/2/5.html>.
- Beamonte Garcia, Pimentel Conde, Gómez Cristóbal, Herrero de Andrés, Fernández Antón, (2010). Técnicas de teledetección y sistemas de información geográfica para la evaluación de la demanda de agua para usos de exterior en la Comunidad de Madrid. Canal de Isabel II.
- Beard, D. (1995). Intervención ante la Comisión Internacional de Grandes Presas Durban Sudáfrica. *Tecnología del Agua* 137: 77-78.
- Billings, B. R. (1982), "Specification of Block Rate Price Variables in Demand Models", *Land Economics*, Vol. 58, núm. 3.
- Bousquet, F., Barreteau, O., Le Page, C., Mullon, C. & Weber, J. (1999). An environmental modelling approach. The use of multi-agent simulations. En Blasco, F. & Weill, A. (Eds), *Advances in environmental and ecological modeling*. New York, NY: Elsevier.

- Brugué, Quim y Ricard Gomà (1998): Gobiernos locales y políticas públicas, Barcelona: Ariel.
- Cebrián, J.A. (1999), Información geográfica y sistemas de información geográfica, Santander, Universidad de Cantabria.
- Clarke, K. C. (1999). Getting started with Geographic Information Systems. Upper Saddle River, NJ: Prentice Hall.
- CNA (2004), Estadísticas del Agua en México 2004, Comisión nacional del Agua, México.
- CNA (2005), Estadísticas del Agua en México 2005, Comisión nacional del Agua, México.
- CNA (2006), Estadísticas del Agua en México 2006, Comisión nacional del Agua, México.
- CNA (2007), Estadísticas del Agua en México 2007, Comisión nacional del Agua, México.
- CNA (2008), Estadísticas del Agua en México 2008, Comisión nacional del Agua, México.
- Cooley, W.W. and P. R. Lohnes (1971) Multivariate Data Analysis. Wiley, N. York.
- Cuadras, C. M. and J. Fortiana (2000) The Importance of Geometry in Multivariate Analysis and some Applications. In: C.R. Rao and G. Szekely, (Eds.), Statistics for the 21st Century, pp. 93-108. Marcel Dekker, N. York.
- Dube E y Van Der Zaag, P. Analysing water use patterns for demand management: The case of the city of Masvingo, Zimbabwe. Physics and Chemistry of the Earth 28, 805 – 815. 2003.

- E. Manoli, P. Katsiardi, G. Arampatzis & D. Assimacopoulos. (2005).
“Comprehensive Water Management Scenarios For Strategic Planning”.
Global NEST Journal. Disponible en
http://www.gnest.org/journal/Vol7_No3/paper_18_Manoli_392.pdf
- Elbers, C., P. Lanjouw, and J. Lanjouw, (2003), “Micro-level Estimation of Poverty and Inequality”, *Econometrica* 71(1), Pp 355-364.
- Font, Joan (2001): *Ciudadanos y decisiones públicas*, Barcelona: Ariel.
- Freund, John E., Simon, Gary A. (2007), *Estadística elemental*, octava edición, México, Pearson & Prentice Hall.
- Hanemann, W. M. (1997) “Price Rate Structures”, D. D. Baumann, J. J. Boland y W. M. Hanemann (comps.), *Urban Water Demand Management and Planing*, Nueva York, McGraw-Hill.
- Hansen, M., W. Hurwitz and W. G. Madow, (1953). “Sample survey methods and theory”. John Wiley and Sons, New York. Reprinted in 1993.
- Harman, H. H. (1976) *Modern Factor Analysis*. The Univ. Chicago Press, Chicago, 3a ed.
- Hartley, H.O. (1958). “Maximum Likelihood Estimation from Incomplete Data”. *Biometrics*, 27, Pp.783-823.
- Hérmendez Sampieri, R., Fernández Collado, C. y Baptista, P. (2010). *Metodología de la investigación* (4ª ed.). México: Mc Graw-Hill.
- Hernández, D. y M. Székely (2005), “Medición del Bienestar en México en los Inicios del Siglo XXI”, Capítulo 3 en Székely, M., “Números que Mueven al

Mundo: la medición de la pobreza en México”, Editorial Porrúa, en Prensa, México DF.

Howe, C. W., y F. P. Linaweaver (1967), “The impact of Price on Residential Water Demand and its Relation to System Demand and Price Structure”, Water Resurces Research, Vol. 1, núm.

Huang, F., X. Wang, L. Liping, Z. Zhiqing, W. Jiaping. (2010). Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques, Water Research, 44:1562-1572.

Hubbell, L. K. (1977). “The Residential Demand for Water and Sewerage Service in Developing Countries: A Case Study of Nairobi,” Urban Reg. Rep. 77-14, Dev. Econ. Dep., World Bank, Washington D.C.

INEGI (2004), Encuesta Nacional Ingreso Gasto en los Hogares 2004, [en línea], <http://www.inegi.org.mx>

INEGI (2006), Encuesta Nacional Ingreso Gasto en los Hogares 2006, [en línea], <http://www.inegi.org.mx>

INEGI (2008), Encuesta Nacional Ingreso Gasto en los Hogares 2008, [en línea], <http://www.inegi.org.mx>

Katzman, M. (1977). “Income and Price Elasticities of Demand for Water in Developing Countries.” Water Resources Bulletin 13: 47–55.

Kazi, T. G., M. B. Arain, M. K. Jamali, N. Jalbani, H. I. Afridi, R. A. Sarfraz, J. A. Baig, A. Q. Shah. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotoxicology and Environmental Safety, 72(2): 301-309.

- Kremer, M., Leino, J., Miguel, E., and Zwane, A. P. (2007). Spring cleaning: A randomized evaluation of source water quality improvement. UC-Berkeley, Working Paper. Disponible en <http://elsa.berkeley.edu>.
- Mason, Robert D., Marchal, William C. Lind, A. Douglas (2004), Estadística para administración y economía, onceava edición, México, Alfaomega.
- Moore, P.G. (1952). "The estimation of the Poisson parameter from a truncated distribution". Biometrika 39, Pp247-251.
- Mu, X., D. Whittington, and J. Briscoe. (1990). "Modeling Village Water Demand Behavior: A Discrete Choice Approach," Water Resources Research.
- Novalés, Alfonso (2006), Econometría, Madrid, España, Editorial Mc Graw Hill.
- Pacheco, A. y Cruz, M. A. (2006). Metodología crítica de la investigación. México: CECSA.
- Phumpiu, Paul (2005) Participación privada en agua y saneamiento. Revista de Economía y Derecho, Vol. 2, No 7. pp. 17-25.
- R. Quentin Grafton, Clay Landry, Gary D. Libecap, and Robert J. O'Brien. (2010). "Water Markets: Australia's Murray-Darling Basin And The U.S. Southwest". National Bureau of Economic Research. Disponible en <http://www.nber.org/papers/w15797>.
- Reitsma, R., Zigurs, I., Lewis, C., Wilson, V. & Sloane, A. (1996). Experiment with simulation models in water-resources negotiations. Journal of water resources planning and management.
- Rojas Soriano, R. (2005). Guía para realizar investigaciones sociales (40ª. ed.). México: Plaza y Valdés.

- Romero Pérez (2002), “Evaluación de la transferencia del distrito de riego 011, Alto Lerma”, Brigitte Boehm.
- Shrestha, S., F. Kazama (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. *Environmental Modelling and Software*, 22: 464-475.
- Sierra, R. (2001). *Técnicas de investigación social*. España, Paraninfo.
- Star, J. & Estes, J. E. (1990). *Geographic Information Systems: An Introduction*. Upper Saddle River, NJ: Prentice Hall.
- Thobani, M. (2007): “Formal water markets: why, when, and how to introduce tradable water rights”. *The World Bank Research Observer*, pp. 161-179.
- UN (2003), *World Water Assessment programme, Water for people Water for Life*, The united nations World Water development report, UNESCO y Berghahn Books, Barcelona.
- UN (2006), *Water a shared responsibility*, The united nations World Water development report 2 (WWDR2), World Water Assessment programme, UNESCO, Berghahn Books, un Water.
- Vialle, C., C. Sablayrolles, M. Lovera, S. Jacob, M. C. Huau. (2011). Monitoring of water quality from roof runoff: Interpretation using multivariate analysis. *Water Research*, 45: 3765-3775.
- White, G., D. Bradley, and A. White. (1972). *Drawers of Water: Domestic Water Use in East Africa*. University of Chicago Press.
- Whittington, D., J. Briscoe, X. Mu, and W. Barron. (1990). “Estimating the Willingness to Pay for Water Services in Developing Countries: A Case

Study of the Use of Contingent Valuation Surveys in Southern Haiti.”

Economic Development and Cultural Change. pp. 293–311.

WHO-UNICEF (2006), Joint Monitoring programme for Water Supply and Sanitation, [en línea], <http://www.wssinfo.org/en/welcome.html>.

WHO-UNICEF (2004), Meeting the Mdg drinking water and sanitation target: a mid-term assessment of progress, 2004, [en línea], http://www.who.int/water_sanitation_health/monitoring/jmp04.pdf

Wunderlin, D., A., Díaz, M., P., Amé, M., V., Pesce, S., F., Hued, A., C., Bistoni M. A. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía river basin (Córdoba-Argentina). Water Res. 35 (12), 2881-2894.

Zeng, X., Rasmussen, T.,C.(2005): Multivariate statistical characterization of water quality in lake Lanier, Georgia, USA. J. Environ. Qual. 34 (6), 1980-1991.

ÍNDICE DE ANEXOS

Anexo	Título
A	Código fuente para la aplicación del Modelo multilogístico
B	Codigo fuente para la aplicación del modelo de componentes principales
C	Código fuente para la homologación de las variables de la ENIGH y el Censo de Población y Vivienda para el año 2010
D	Imputación de datos utilizando PovMap 2.0
E	Diagrama de relación del modelo multilogístico Nacional utilizando datos de la ENIGH 2010
F	Diagrama de relación para la explotación de las bases de datos del Censo de Población y Vivienda 2010
G	Producción científica.

ANEXO A

CÓDIGO FUENTE PARA LA APLICACIÓN DEL MODELO MULTILOGÍSTICO

* Análisis del agua 2010;

```
gl data="C:\Documents and Settings\Matg\Enigh 2010\Enigh 2010\Bdd Stata";  
log using "$data\análisis_tesis__2010.log", replace;
```

*Genera la tabla de disponibilidad de pago por hogar;

```
use folioviv foliohog gas_tri clave if clave == "A215" using "$data\gastos2010.dta", clear;  
gen folio = folioviv + foliohog;  
gen gas_agua = gas_tri/3;  
collapse (sum) gas_agua, by (folio);  
sort folio;  
save "$data\gas_mes_agua_2010.dta", replace;  
use folioviv foliohog gas_tri clave if clave == "R001" using "$data\gastos2010.dta", clear;  
gen folio = folioviv + foliohog;  
gen gas_luz = gas_tri/3;  
collapse (sum) gas_luz, by (folio);  
sort folio;  
save "$data\gas_mes_luz_2010.dta", replace;
```

*Guarda la tabla de pago por energía eléctrica;

*Genera la tabla de necesidades de agua por hogar;

```
use folioviv foliohog tot_resi using "$data\concen2010.dta", clear;  
gen folio = folioviv + foliohog;  
sort folio;  
save "$data\tam_hog_2010.dta", replace;
```

*Guarda la tabla de necesidades de agua por hogar;

*Genera la tabla de ingresos por hogar;

```
use folioviv foliohog ing_tri using "$data\ingresos2010.dta", clear;
```

```

gen folio = folioviv + foliohog;
gen ing_mes = ing_tri/3;

collapse (sum) ing_mes, by (folio);
sort folio;
save "$data\ing_hog_a_2010.dta", replace;

*Guarda la tabla de ingresos por hogar;

*Genera la tabla de analfabetismo por hogar;

*cambios;
use folioviv foliohog numren alfabe if alfabe==2 using "$data\poblacion2010.dta", clear;
*termina cambios;

gen double n_analfabe = 1;
gen folio = folioviv + foliohog;
collapse (sum) n_analfabe, by (folio);
gen double analf_hog = 1;
sort folio;
save "$data\analfabetismo2010.dta", replace;
*Guarda la tabla de analfabetismo por hogar;

*Genera la tabla de gradomáximo de educación de cualquier miembro del hogar;
use folioviv foliohog nivelaprob gradoaprob using "$data\poblacion2010.dta", clear;
gen folio = folioviv + foliohog;
* cambios;
gen double esc_1 = (nivelaprob);
gen double cur_1 = (gradoaprob);
* termina cambios;
egen double escolar = max(esc_1), by (folio);
keep if (esc_1==escolar);
egen double cursado=max(cur_1), by (folio);
keep if (cur_1==cursado);
collapse escolar cursado, by (folio);
sort folio;
merge folio using "$data\hogares2010.dta", sort keep(factor);

```

```
sort folio;  
drop _merge;
```

```
gen double ninguno=0;  
gen double pres_in=0;  
gen double prim_in=0;  
gen double sec_in=0;  
gen double prep_in=0;  
gen double norm_in =0;  
gen double tec_in =0;  
gen double prof_in =0;  
gen double maest_in =0;  
gen double doc_in =0;  
gen double pres_ter=0;  
gen double prim_ter=0;  
gen double sec_ter=0;  
gen double prep_ter=0;  
gen double norm_ter =0;  
gen double tec_ter =0;  
gen double prof_ter =0;  
gen double maest_ter =0;  
gen double doc_ter =0;  
replace ninguno = factor if (escolar==0);  
replace pres_in=factor if escolar ==1 & cursado < 3;  
replace prim_in=factor if escolar ==2 & cursado < 6;  
replace sec_in=factor if escolar ==3 & cursado < 3;  
replace prep_in=factor if escolar ==4 & cursado < 3;  
replace norm_in =factor if escolar ==5 & cursado < 4;  
replace tec_in =factor if escolar ==6 & cursado < 4;  
replace prof_in =factor if escolar ==7 & cursado < 5;  
replace maest_in = factor if escolar ==8 & cursado < 3;  
replace doc_in = factor if escolar ==9 & cursado < 3;  
replace pres_ter=factor if escolar ==1 & cursado ==3;  
replace prim_ter=factor if escolar ==2 & cursado ==6;  
replace sec_ter=factor if escolar ==3 & cursado == 3;  
replace prep_ter=factor if escolar ==4 & cursado == 3;
```

```

replace norm_ter =factor if escolar ==5 & cursado == 4;
replace tec_ter =factor if escolar ==6 & cursado == 4;
replace prof_ter =factor if escolar ==7 & (cursado == 5 | cursado ==6);
replace maest_ter = factor if escolar ==8 & (cursado == 3 | cursado ==4);
replace doc_ter = factor if escolar ==9 & cursado == 3;
gen double grado_max = 0;
replace grado_max = 0 if (escolar==0);
replace grado_max = 1.5 if escolar ==1 & cursado < 3;
replace grado_max = 6 if escolar ==2 & cursado < 6;
replace grado_max = 10.5 if escolar ==3 & cursado < 3;
replace grado_max = 13.5 if escolar ==4 & cursado < 3;
replace grado_max = 17 if escolar ==5 & cursado < 4;
replace grado_max = 17 if escolar ==6 & cursado < 4;
replace grado_max = 21.5 if escolar ==7 & cursado < 5;
replace grado_max = 25.5 if escolar ==8 & cursado < 3;
replace grado_max = 28.5 if escolar ==9 & cursado < 3;
replace grado_max = 3 if escolar ==1 & cursado ==3;
replace grado_max = 9 if escolar ==2 & cursado ==6;
replace grado_max = 12 if escolar ==3 & cursado == 3;
replace grado_max = 15 if escolar ==4 & cursado == 3;
replace grado_max = 19 if escolar ==5 & cursado == 4;
replace grado_max = 19 if escolar ==6 & cursado == 4;
replace grado_max = 24 if escolar ==7 & (cursado == 5 | cursado ==6);
replace grado_max = 27 if escolar ==8 & (cursado == 3 | cursado ==4);
replace grado_max = 30 if escolar ==9 & cursado == 3;
save "$data\gra_max_2010.dta", replace;

```

*Guarda la tabla de grado máximo de educación;

*Genera la tabla de gradomáximo de educación alcanzado por el jefe de familia;

```

use folioviv foliohog nivelaprob gradoaprob parentesco if (parentesco == 100) using $data\poblacion2010.dta",
clear;
gen folio = folioviv + foliohog;
sort folio;
merge folio using "$data\hogares2010.dta", sort keep(factor);
sort folio;
drop _merge;

```

*Cambios;

gen double escolarj = (nivelaprob);

gen double cursadoj = (gradoaprob);

* termina cambios;

gen double ningunoj=0;

gen double pres_inj=0;

gen double prim_inj=0;

gen double sec_inj=0;

gen double prep_inj=0;

gen double norm_inj =0;

gen double tec_inj =0;

gen double prof_inj =0;

gen double maest_inj =0;

gen double doc_inj =0;

gen double pres_terj=0;

gen double prim_terj=0;

gen double sec_terj=0;

gen double prep_terj=0;

gen double norm_terj =0;

gen double tec_terj =0;

gen double prof_terj =0;

gen double maest_terj =0;

gen double doc_terj =0;

replace ningunoj = factor if (escolarj==0);

replace pres_inj=factor if escolarj ==1 & cursadoj < 3;

replace prim_inj=factor if escolarj ==2 & cursadoj < 6;

replace sec_inj=factor if escolarj ==3 & cursadoj < 3;

replace prep_inj=factor if escolarj ==4 & cursadoj < 3;

replace norm_inj =factor if escolarj ==5 & cursadoj < 4;

replace tec_inj =factor if escolarj ==6 & cursadoj < 4;

replace prof_inj =factor if escolarj ==7 & cursadoj < 5;

replace maest_inj = factor if escolarj ==8 & cursadoj < 3;

replace doc_inj = factor if escolarj ==9 & cursadoj < 3;

```

replace pres_terj=factor if escolarj ==1 & cursadoj ==3;
replace prim_terj=factor if escolarj ==2 & cursadoj ==6;
replace sec_terj=factor if escolarj ==3 & cursadoj == 3;
replace prep_terj=factor if escolarj ==4 & cursadoj == 3;
replace norm_terj =factor if escolarj ==5 & cursadoj == 4;
replace tec_terj =factor if escolarj ==6 & cursadoj == 4;
replace prof_terj =factor if escolarj ==7 & (cursadoj == 5 | cursadoj ==6);
replace maest_terj = factor if escolarj ==8 & (cursadoj == 3 | cursadoj ==4);
replace doc_terj = factor if escolarj ==9 & cursadoj == 3;
sort folio;
save "$data\gra_max_jefe_2010.dta", replace;

```

*Guarda la tabla de grado máximo de educación alcanzado por el jefe de familia;

```

use folioviv foliohog drenaje dis_agua elect factor using "$data\hogares2010.dta", clear;
gen folio = folioviv + foliohog;
sort folio;
merge folio using "$data\gas_mes_agua_2010.dta", keep(gas_agua);
replace gas_agua=0 if gas_agua ==.;
drop _merge;
sort folio;
merge folio using "$data\gas_mes_luz_2010.dta", keep(gas_luz);
replace gas_luz=0 if gas_luz ==.;
drop _merge;
sort folio;
merge folio using "$data\tam_hog_2010.dta", keep(tot_resi);
replace tot_resi=0 if tot_resi ==.;
drop _merge;
sort folio;
merge folio using "$data\ing_hog_a_2010.dta", keep(ing_mes);
replace ing_mes=0 if ing_mes==.;
drop _merge;
sort folio;
merge folio using "$data\concen2010.dta", keep(n_ocup perocu tam_hog edad hombres tam_loc ingcor
    estrato);
ren edad edad_jefe;
sort folio;

```

```
drop _merge;
```

```
merge folio using "$data\analfabetismo2010.dta", keep(analf_hog);
```

```
replace analf_hog=0 if analf_hog==.;
```

```
sort folio;
```

```
drop _merge;
```

```
merge folio using "$data\gra_max_2010.dta", keep(grado_max);
```

```
replace grado_max=0 if grado_max==.;
```

```
sort folio;
```

```
drop _merge;
```

```
* Servicios públicos;
```

```
gen double red = (dis_agua ==1 | dis_agua ==2 | dis_agua ==4);
```

```
gen double hidrante = (dis_agua ==3);
```

```
gen double pipa = (dis_agua==5);
```

```
gen double río_pozo = (agua13==6);
```

```
gen condrenaje = (drenaje==1);
```

```
gen conluz = (elect==1);
```

```
*Termina servicios públicos;
```

```
* Generación de deciles;
```

```
gen double yct=ingcor;
```

```
replace yct=0 if yct==.;
```

```
table estrato [iw= factor], c(sum ingcor sum yct) f(%15.2f) row;
```

```
gen double yct_pc=yct/tam_hog;
```

```
sort yct_pc;
```

```
gen double pi = sum(factor);
```

```
gen double ppi = pi/pi[_N];
```

```
gen double qi = sum(yct_pc*factor);
```

```
gen double qqi = qi/qi[_N];
```

```
gen decil = 0;
```

```
for num 1/5:replace decil=X if ppi>((X-1)/5) & ppi<=(X/5);
```

```
rename decil dhn_yct_pc;  
table dhn_yct_pc [iw=factor], c(freq sum ingcor min yct_pc max yct_pc) f(%15.2f) row;
```

```
drop pi ppi qi qqi;  
sort yct_pc;  
gen double pi = sum(factor*tam_hog);  
gen double ppi=pi/pi[_N];  
gen double qi = sum(yct_pc*tam_hog*factor);  
gen double qqi = qi/qi[_N];  
gen decil = 0;  
for num 1/5:replace decil=X if ppi>((X-1)/5) & ppi<=(X/5);  
rename decil dpn_yct_pc;  
table dpn_yct_pc [iw=factor], c(freq sum ingcor min yct_pc max yct_pc) f(%15.2f) row;
```

* Termina generación de deciles;

```
gen double zona =(tam_loc == 1 | tam_loc == 2);
```

* Combinaciones de los servicio;

```
gen double combina = 0;  
replace combina = 1 if (red ==0 & hidrante==0 & pipa==0 & río_pozo==0);  
replace combina = 2 if (red ==0 & hidrante==0 & pipa==0 & río_pozo==1);  
replace combina = 3 if (red ==0 & hidrante==0 & pipa==1 & río_pozo==0);  
replace combina = 5 if (red ==0 & hidrante==1 & pipa==0 & río_pozo==0);  
replace combina = 9 if (red ==1 & hidrante==0 & pipa==0 & río_pozo==0);  
gen double hog_combina = 0;  
replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==0);  
replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==1);  
replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==1 & río_pozo==0);  
replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==1 & río_pozo==1);  
replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==0 & río_pozo==0);  
replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==0 & río_pozo==1);  
replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==1 & río_pozo==0);  
replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==1 & río_pozo==1);  
replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==0 & río_pozo==0);  
replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==0 & río_pozo==1);
```

```
replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==1 & río_pozo==0);
replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==1 & río_pozo==1);
replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==0 & río_pozo==0);
replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==0 & río_pozo==1);
replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==1 & río_pozo==0);
replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==1 & río_pozo==1);
```

```
gen sin_serv=0;
replace sin_serv = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==0);
gen con_río_pozo=0;
replace con_río_pozo = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==1);
gen con_pipa=0;
replace con_pipa = factor if (red ==0 & hidrante==0 & pipa==1 & río_pozo==0);
gen con_hidrante=0;
replace con_hidrante = factor if (red ==0 & hidrante==1 & pipa==0 & río_pozo==0);
gen con_red=0;
replace con_red = factor if (red ==1 & hidrante==0 & pipa==0 & río_pozo==0);
```

* Guarda la base de datos nacional;

```
save "$data\Nacional.dta", replace;
```

*Clasifica por estado;

```
gen estado = substr(folio,1,2);
```

* Modelo multilogístico;

```
tab combina;
```

```
bysort combina: sum ing_mes gas_agua gas_luz tot_resi analf_hog zona condrenaje conluz estrato grado_max;
estsimp mlogit combina ing_mes gas_agua gas_luz tot_resi analf_hog zona condrenaje conluz estrato
grado_max [fweight=factor], base(9);
```

```
log close;
```

ANEXO B

CÓDIGO FUENTE PARA LA APLICACIÓN DEL MODELO DE COMPONENTES PRINCIPALES

```
# delimit;
set mem 1100m;
set more off;
set matsize 800;
* Análisis del agua 2010;

gl ruta="C:\Users\Documents\Tesis Doctoral\Bdd stata";
gl viviendas="C:\Users\Documents\Tesis Doctoral\Censo 2010\Dtas";

*log using "$data\factor análisis.log", replace;
* Construye la base de datos viviendas 2010 para el cálculo de componentes principales;

do "$ruta\Bdd.do";

* Termina la construcción de la base de datos;

use "$viviendas\viviendas_2010.dta", clear;

* Componentes principales;
* Cálculo de índices;
gen p_sin_red = (100-(( tred/ factor)*100));
gen p_río_pozo=((trío_pozo/factor)*100);
gen p_dot_devez=((tdot_devez/factor)*100);
gen p_hidrante= ((thidrante/factor)*100);
gen psin_drenaje=((tsin_drenaje/factor)*100);

*Termina Cálculo de índices;

*Estandarización de índices;
egen med_sinred = mean(p_sin_red);
egen desv_sinred= sd(p_sin_red);
gen z_sinred = ((p_sin_red-med_sinred)/desv_sinred);
egen med_rpozo = mean(p_río_pozo);
egen desv_rpozo= sd(p_río_pozo);
```

```

gen z_rpozo = (p_río_pozo-med_rpozo)/desv_rpozo;
egen med_dvez = mean(p_dot_devez);
egen desv_dvez= sd(p_dot_devez);
gen z_dvez = (p_dot_devez-med_dvez)/desv_dvez;
egen med_hidra = mean(p_hidrante);
egen desv_hidra= sd(p_hidrante);
gen z_hidra = (p_hidrante-med_hidra)/desv_hidra;
egen med_sdrena = mean(psin_drenaje);
egen desv_sdrena= sd(psin_drenaje);
gen z_sdrena = (psin_drenaje-med_sdrena)/desv_sdrena;

```

*Termina estandarización de índices;

```

save "$viviendas\viviendas_2010.dta", replace;

```

```

pca z_sinred z_rpozo z_dvez z_hidra z_sdrena, correlation;
sdtest z_sinred==z_hidra;
estat summarize;
estat anti;
estat loading;
estat residual;
estat kmo;
predict marg_agua c_2 c_3 c_4 c_5;
summarize marg_agua;
gen estratos = 0;
replace estratos = 6 if (marg_agua <= -2.986274000568);
replace estratos = 5 if ((marg_agua < -1.493137000568) & (marg_agua > -2.986274000568));
replace estratos = 4 if ((marg_agua < -0.000000000568) & (marg_agua >= -1.493137000568));
replace estratos = 3 if ((marg_agua > -0.000000000568) & (marg_agua <= 1.493136999432));
replace estratos = 2 if ((marg_agua > 1.493136999432) & (marg_agua < 2.986273999432));
replace estratos = 1 if (marg_agua >= 2.986273999432);

```

```

tabulate estratos;
gen cveest2=0;

```

```

label def estratos2 1 "Muy alto" 2 "Alto" 3 "Medio alto" 4 "Medio bajo" 5 "Bajo" 6 "Muy bajo";
label values estratos estratos2;

```

```

replace cveest = real( ent);
label def aux 1 "Aguascalientes" 2 "Baja California" 3 "Baja California Sur" 4 "Campeche" 5
"Coahuila de Zaragoza" 6 "Colima" 7 "Chiapas" 8 "Chihuahua" 9 "Distrito Federal" 10 "Durango" 11
"Guanajuato" 12 "Guerrero" 13 "Hidalgo" 14 "Jalisco" 15 "Estado de México" 16 "Michoacán de
Ocampo" 17 "Morelos" 18 "Nayarit" 19 "Nuevo León" 20 "Oaxaca" 21 "Puebla" 22 "Querétaro de
Arteaga" 23 "Quintana Roo" 24 "San Luis Potosí" 25 "Sinaloa" 26 "Sonora" 27 "Tabasco" 28
"Tamaulipas" 29 "Tlaxcala" 30 "Veracruz de Ignacio de la Llave" 31 "Yucatán" 32 "Zacatecas";

```

```

label values cveest2 aux;
tab cveest2 estratos, chi2;
screepplot, yline(1) ci(het);
correlate marg_agua z_sinred z_rpozo z_dvez z_hidra z_sdrena;
save "$viviendas\viviendas_2010.dta", replace;

```

***Módulo Bdd.do;**

*Base de datos general de características de las viviendas;

```

gl rutavar="C:\Users\Documents\Tesis Doctoral\Bdd stata";
gl datos="C:\Users\Documents\Tesis Doctoral\Censo 2010\Dtas";

```

```

use "$datos\viviendas_01.dta", clear;

```

*Construcción de variables;

```

include "$rutavar\variables.do";

```

*Termina construcción de variables;

*Construcción de base de datos del primer estado;

```

collapse (sum) tred tredf thidrante tot_red tpipa trío_pozo tdot_devez tdot_sem tsin_san tsin_drenaje
tsin_elec factor, by (mun);
gen ent = "01";
sort ent mun;

```

```

save "$datos\viviendas_2010.dta", replace;

```

* Termina proceso de primer estado;

```

gl datos2= ".dta";

```

```

local i=2;
while `i' <=32 {;

if `i' <= 9 {;
    gl cero ="0";
    gl datos="C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo 2010\Dtas\viviendas_0";
        gl datos4 = string(`i');
    gl datos3 = "$datos$datos4$datos2";
    use "$datos3", clear;
        *Construcción de variables;
    include "$rutavar\variables.do";
    *Termina construcción de variables;
    collapse (sum) tred tref thidrate tot_red tpipa trío_pozo tdot_devez tdot_sem tsin_san
tsin_drenaje tsin_elec factor, by (mun);
        gen ent = "$cero$datos4";
        sort ent mun;

    save "C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo 2010\Dtas\temporal.dta",
replace;
        use "C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo
2010\Dtas\viviendas_2010.dta", clear;
        merge ent mun using "C:\Users\ratonmiguelito\Documents\Tesis
Doctoral\Censo 2010\Dtas\temporal.dta", sort;
        drop _merge;
        save "C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo
2010\Dtas\viviendas_2010.dta", replace;
        clear all;
    };
else {;
        gl      datos="C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo
2010\Dtas\viviendas_";
        gl datos4 = string(`i');
    gl datos3 = "$datos$datos4$datos2";
    use "$datos3", clear;
        *Construcción de variables;
    include "$rutavar\variables.do";

```

```

*Termina construcción de variables;

collapse (sum) tred tredf thidrante tot_red tpipa río_pozo tdot_devez tdot_sem tsin_san
tsin_drenaje tsin_elec factor, by (mun);

gen ent = "$datos4";
sort ent mun;

save "C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo 2010\Dtas\temporal.dta",
replace;

use "C:\Documents\Tesis Doctoral\Censo 2010\Dtas\viviendas_2010.dta", clear;
merge ent mun using "C:\Users\ratonmiguelito\Documents\Tesis
Doctoral\Censo 2010\Dtas\temporal.dta", sort;

drop _merge;
save "C:\Users\ratonmiguelito\Documents\Tesis Doctoral\Censo
2010\Dtas\viviendas_2010.dta", replace;
clear all;

};

local i=`i'+1;
};

```

```
# delimit;
```

```
*Módulo Variables.do;
```

```
*Genera variables para cálculo de componentes;
```

```

*gen double red = (disagu ==1 | disagu ==2);
gen double red = (disagu ==1);
gen double red_f = (disagu ==2);
gen double hidrante = (disagu ==3);
gen double ot_red = (disagu ==4);
gen double pipa = (disagu==5);
gen double río_pozo = (disagu==6);
gen double dot_devez= (dotaguad==5);
gen double dot_sem = (dotaguad==4);
gen double sin_san = (sersan==3);
gen double sin_drenaje = (drenaje==5);
gen double sin_elec = (electri==3);
gen double tred = red * factor;

```

```
gen double thidrante = hidrante * factor;  
gen double tot_red = ot_red * factor;  
gen double tpipa = pipa * factor;  
gen double trío_pozo = río_pozo * factor;  
gen double tdot_devez= dot_devez* factor;  
gen double tdot_sem = dot_sem * factor;  
gen double tsin_san = sin_san * factor;  
gen double tsin_drenaje = sin_drenaje * factor;  
gen double tsin_elec = sin_elec * factor;  
gen double tredf = red_f * factor;
```

*Termina generación de variables;

ANEXO C

CÓDIGO FUENTE PARA LA HOMOLOGACIÓN DE LAS VARIABLES DE LA ENIGH Y EL CENSO DE POBLACIÓN Y VIVIENDA PARA EL AÑO 2010

* Dase de datos del censo 2010;

```
use ent mun id_viv edad sexo nivacad escolar parent if(parent ==1) using
```

```
"$datacenso\personas_20.dta", clear;
```

```
gen folio = ent + mun + id_viv;
```

```
ren edad edad_jefe;
```

```
ren sexo sexo_jefe;
```

```
ren nivacad nivelaprob;
```

```
ren escolar gradoaprob;
```

```
sort folio;
```

```
save "$data\jefe_censo.dta", replace;
```

```
use ent mun id_viv alfabet if (alfabet==1) using "$datacenso\personas_20.dta", clear;
```

```
gen double n_analfabe = 1;
```

```
gen folio = ent + mun + id_viv;
```

```
collapse (sum) n_analfabe, by (folio);
```

```
gen double analf_hog = 1;
```

```
replace n_analfabe = 0 if (n_analfabe==.);
```

```
replace analf_hog = 0 if (analf_hog==.);
```

```
sort folio;
```

```
save "$data\analfa_censo.dta", replace;
```

```
clear all;
```

```
use ent mun id_viv nivacad escolar factor using "$datacenso\personas_20.dta", clear;
```

```
*drop folio;
```

```
gen folio = ent + mun + id_viv;
```

```
* cambios;
```

```
ren nivacad nivel;
```

```
ren escolar grado;
gen double esc_1 = nivel;
gen double cur_1 = grado;
* termina cambios;

egen double escolar = max(esc_1), by (folio);
keep if (esc_1==escolar);
egen double cursado=max(cur_1), by (folio);
keep if (cur_1==cursado);
collapse  escolar cursado, by (folio);
sort folio;
merge folio using "$datacenso\factorviv.dta", sort keep(factor);
sort folio;
drop _merge;
```

```
gen double ninguno=0;
gen double pres_in=0;
gen double prim_in=0;
gen double sec_in=0;
gen double prep_in=0;
gen double norm_in =0;
gen double tec_in =0;
gen double prof_in =0;
gen double maest_in =0;
gen double doc_in =0;
```

```
gen double pres_ter=0;
gen double prim_ter=0;
gen double sec_ter=0;
gen double prep_ter=0;
gen double norm_ter =0;
gen double tec_ter =0;
gen double prof_ter =0;
gen double maest_ter =0;
gen double doc_ter =0;
```

```
replace ninguno = factor if (escolar==0);
replace pres_in=factor if escolar ==1 & cursado < 3;
replace prim_in=factor if escolar ==2 & cursado < 6;
replace sec_in=factor if escolar ==3 & cursado < 3;
replace prep_in=factor if escolar ==4 & cursado < 3;
replace norm_in =factor if escolar ==5 & cursado < 4;
replace tec_in =factor if escolar ==6 & cursado < 4;
replace prof_in =factor if escolar ==7 & cursado < 5;
replace maest_in = factor if escolar ==8 & cursado < 3;
replace doc_in = factor if escolar ==9 & cursado < 3;
```

```
replace pres_ter=factor if escolar ==1 & cursado ==3;
replace prim_ter=factor if escolar ==2 & cursado ==6;
replace sec_ter=factor if escolar ==3 & cursado == 3;
replace prep_ter=factor if escolar ==4 & cursado == 3;
replace norm_ter =factor if escolar ==5 & cursado == 4;
replace tec_ter =factor if escolar ==6 & cursado == 4;
replace prof_ter =factor if escolar ==7 & (cursado == 5 | cursado ==6);
replace maest_ter = factor if escolar ==8 & (cursado == 3 | cursado ==4);
replace doc_ter = factor if escolar ==9 & cursado == 3;
```

```
gen double grado_max = 0;
```

```
replace grado_max = 0 if (escolar==0);
replace grado_max = 1.5 if escolar ==1 & cursado < 3;
replace grado_max = 6 if escolar ==2 & cursado < 6;
replace grado_max = 10.5 if escolar ==3 & cursado < 3;
replace grado_max = 13.5 if escolar ==4 & cursado < 3;
replace grado_max = 17 if escolar ==5 & cursado < 4;
replace grado_max = 17 if escolar ==6 & cursado < 4;
replace grado_max = 21.5 if escolar ==7 & cursado < 5;
replace grado_max = 25.5 if escolar ==8 & cursado < 3;
replace grado_max = 28.5 if escolar ==9 & cursado < 3;
```

```
replace grado_max = 3 if escolar ==1 & cursado ==3;
replace grado_max = 9 if escolar ==2 & cursado ==6;
```

```

replace grado_max = 12 if escolar ==3 & cursado == 3;
replace grado_max = 15 if escolar ==4 & cursado == 3;
replace grado_max = 19 if escolar ==5 & cursado == 4;
replace grado_max = 19 if escolar ==6 & cursado == 4;
replace grado_max = 24 if escolar ==7 & (cursado == 5 | cursado ==6);
replace grado_max = 27 if escolar ==8 & (cursado == 3 | cursado ==4);
replace grado_max = 30 if escolar ==9 & cursado == 3;

```

```

save "$data\gra_max_censo.dta", replace;

```

*Genera la tabla de gradomáximo de educación alcanzado por el jefe de familia del censo 2010;

```

clear all;
use ent mun id_viv nivacad escolar parent factor if (parent == 1) using
"$datacenso\personas_20.dta", clear;
gen folio = ent + mun + id_viv;
* cambios;
ren nivacad escolarj;
ren escolar cursadoj;
* termina cambios;

```

```

gen double ningunoj=0;
gen double pres_inj=0;
gen double prim_inj=0;
gen double sec_inj=0;
gen double prep_inj=0;
gen double norm_inj =0;
gen double tec_inj =0;
gen double prof_inj =0;
gen double maest_inj =0;
gen double doc_inj =0;

```

```

gen double pres_terj=0;
gen double prim_terj=0;
gen double sec_terj=0;
gen double prep_terj=0;
gen double norm_terj =0;

```

```
gen double tec_terj =0;
gen double prof_terj =0;
gen double maest_terj =0;
gen double doc_terj =0;
```

```
replace ningunoj = factor if (escolarj==0);
replace pres_inj=factor if escolarj ==1 & cursadoj < 3;
replace prim_inj=factor if escolarj ==2 & cursadoj < 6;
replace sec_inj=factor if escolarj ==3 & cursadoj < 3;
replace prep_inj=factor if escolarj ==4 & cursadoj < 3;
replace norm_inj =factor if escolarj ==5 & cursadoj < 4;
replace tec_inj =factor if escolarj ==6 & cursadoj < 4;
replace prof_inj =factor if escolarj ==7 & cursadoj < 5;
replace maest_inj = factor if escolarj ==8 & cursadoj < 3;
replace doc_inj = factor if escolarj ==9 & cursadoj < 3;
```

```
replace pres_terj=factor if escolarj ==1 & cursadoj ==3;
replace prim_terj=factor if escolarj ==2 & cursadoj ==6;
replace sec_terj=factor if escolarj ==3 & cursadoj == 3;
replace prep_terj=factor if escolarj ==4 & cursadoj == 3;
replace norm_terj =factor if escolarj ==5 & cursadoj == 4;
replace tec_terj =factor if escolarj ==6 & cursadoj == 4;
replace prof_terj =factor if escolarj ==7 & (cursadoj == 5 | cursadoj ==6);
replace maest_terj = factor if escolarj ==8 & (cursadoj == 3 | cursadoj ==4);
replace doc_terj = factor if escolarj ==9 & cursadoj == 3;
```

```
gen double grado_jefe = 0;
```

```
replace grado_jefe = 0 if (escolarj==0);
replace grado_jefe = 1.5 if escolarj ==1 & cursadoj < 3;
replace grado_jefe = 6 if escolarj ==2 & cursadoj < 6;
replace grado_jefe = 10.5 if escolarj ==3 & cursadoj < 3;
replace grado_jefe = 13.5 if escolarj ==4 & cursadoj < 3;
replace grado_jefe = 17 if escolarj ==5 & cursadoj < 4;
replace grado_jefe = 17 if escolarj ==6 & cursadoj < 4;
replace grado_jefe = 21.5 if escolarj ==7 & cursadoj < 5;
replace grado_jefe = 25.5 if escolarj ==8 & cursadoj < 3;
```

```
replace grado_jefe = 28.5 if escolarj ==9 & cursadoj < 3;
```

```
replace grado_jefe = 3 if escolarj ==1 & cursadoj ==3;
```

```
replace grado_jefe = 9 if escolarj ==2 & cursadoj ==6;
```

```
replace grado_jefe = 12 if escolarj ==3 & cursadoj == 3;
```

```
replace grado_jefe = 15 if escolarj ==4 & cursadoj == 3;
```

```
replace grado_jefe = 19 if escolarj ==5 & cursadoj == 4;
```

```
replace grado_jefe = 19 if escolarj ==6 & cursadoj == 4;
```

```
replace grado_jefe = 24 if escolarj ==7 & (cursadoj == 5 | cursadoj ==6);
```

```
replace grado_jefe = 27 if escolarj ==8 & (cursadoj == 3 | cursadoj ==4);
```

```
replace grado_jefe = 30 if escolarj ==9 & cursadoj == 3;
```

```
sort folio;
```

```
save "$data\gra_max_jefe_censo.dta", replace;
```

*Guarda la tabla de grado máximo de educación alcanzado por el jefe de familia del censo 2010;

```
use ent mun id_viv factor disagu drenaje electri numpers tam_loc ingtrhog using  
"$datacenso\viviendas_20.dta", clear;
```

```
gen folio = ent + mun + id_viv;
```

```
*egen media =mean ( ingtrhog) if (ingtrhog != 999999 & ingtrhog != 999998);
```

```
ren disagu dis_agua;
```

```
ren drenaje condrenaje;
```

```
ren electri conluz;
```

```
ren numpers tot_resi;
```

```
ren ingtrhog ing_mes;
```

```
merge folio using "$data\jefe_censo.dta", keep(edad_jefe sexo_jefe nivelaprob gradoaprob parent);
```

```
drop _merge;
```

```
merge folio using "$data\analfa_censo.dta", keep(n_analfabe analf_hog);
```

```
drop _merge;
```

```
sort folio;
```

```
sort ing_mes;
```

```
merge folio using "$data\gra_max_censo.dta", sort keep(grado_max);
drop _merge;
sort folio;
```

```
merge folio using "$data\gra_max_jefe_censo.dta", sort keep(grado_jefe);
drop _merge;
sort folio;
```

```
replace condrenaje = 0 if (condrenaje != 1);
replace conluz = 0 if (conluz != 1);
ren tot_resi tot_resi_tmp;
gen tot_resi =real(tot_resi_tmp);
drop tot_resi_tmp;
replace sexo_jefe = 2 if ( sexo_jefe==3);
replace analf_hog=0 if( analf_hog==.);
gen ing_pp = (ing_mes/tot_resi);
```

* Une los ingresos imputado a la base de datos;

```
merge folio using "$data\ImputadosPP.dta", sort keep(ing_pp_imp);
drop _merge;
sort folio;
```

* Servicios públicos;

```
gen double red = (dis_agua ==1 | dis_agua ==2| dis_agua ==4);
gen double hidrante = (dis_agua ==3);
gen double pipa = (dis_agua==5);
gen double río_pozo = (dis_agua==6);
```

*Termina servicios públicos;

```
gen double zona =(tam_loc == 3 | tam_loc == 4);
```

* Combinaciones de los servicio;

gen double combina = 0;

replace combina = 1 if (red ==0 & hidrante==0 & pipa==0 & río_pozo==0);

replace combina = 2 if (red ==0 & hidrante==0 & pipa==0 & río_pozo==1);

replace combina = 3 if (red ==0 & hidrante==0 & pipa==1 & río_pozo==0);

replace combina = 5 if (red ==0 & hidrante==1 & pipa==0 & río_pozo==0);

replace combina = 9 if (red ==1 & hidrante==0 & pipa==0 & río_pozo==0);

gen double hog_combina = 0;

replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==0);

replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==1);

replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==1 & río_pozo==0);

replace hog_combina = factor if (red ==0 & hidrante==0 & pipa==1 & río_pozo==1);

replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==0 & río_pozo==0);

replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==0 & río_pozo==1);

replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==1 & río_pozo==0);

replace hog_combina = factor if (red ==0 & hidrante==1 & pipa==1 & río_pozo==1);

replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==0 & río_pozo==0);

replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==0 & río_pozo==1);

replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==1 & río_pozo==0);

replace hog_combina = factor if (red ==1 & hidrante==0 & pipa==1 & río_pozo==1);

replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==0 & río_pozo==0);

replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==0 & río_pozo==1);

replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==1 & río_pozo==0);

replace hog_combina = factor if (red ==1 & hidrante==1 & pipa==1 & río_pozo==1);

gen sin_serv=0;

replace sin_serv = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==0);

gen con_río_pozo=0;

replace con_río_pozo = factor if (red ==0 & hidrante==0 & pipa==0 & río_pozo==1);

gen con_pipa=0;

replace con_pipa = factor if (red ==0 & hidrante==0 & pipa==1 & río_pozo==0);

gen con_hidrante=0;

replace con_hidrante = factor if (red ==0 & hidrante==1 & pipa==0 & río_pozo==0);

gen con_red=0;

replace con_red = factor if (red ==1 & hidrante==0 & pipa==0 & río_pozo==0);

```
saveold "$data\Oaxaca_censo9.dta", replace;
```

```
* Modelo multilogístico para Oaxaca con datos imputados;
```

```
tab combina;
```

```
bysort combina: sum ing_pp_imp tot_resi analf_hog zona condrenaje conluz grado_max grado_jefe;
```

```
estsimp mlogit combina ing_pp_imp tot_resi analf_hog zona condrenaje conluz grado_max  
grado_jefe sexo_jefe edad_jefe [fweight=factor], base(9);
```

```
* Termina el modelo multilogístico con datos de censo 2010;
```

```
saveold "$data\Oax_imputados.dta", replace;
```

ANEXO D

Utilización del PovMap 2.0.

La herramienta PovMap 2.0 es una solución informática que permite la imputación de datos de niveles inferiores como municipales y locales recopilados mediante una encuesta a un conjunto de datos de un censo con la finalidad de realizar estudios con mayor precisión.

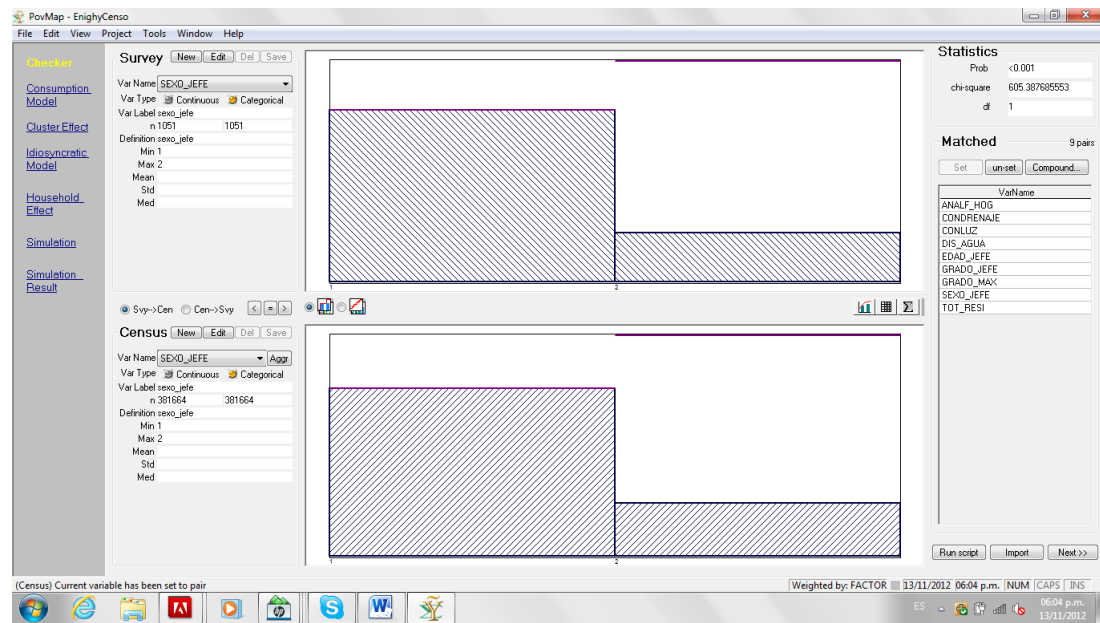
En su primera fase, el software busca parejas de variables candidatas, compara las estadísticas descriptivas de la encuesta y los datos del censo, editar las propiedades de variables y establece las variables que serán utilizadas en la regresión.

Cuando se cotejan dos variables categóricas de la encuesta y del censo, el sistema ejecuta una prueba estadística chi-cuadrado, que compara las frecuencias de la encuesta con las frecuencias del censo. La importancia de la estadística de chi-cuadrado es indicar si ambas poblaciones tienen distribuciones de frecuencias similares (ver figura 18).

En el caso de la comparación de variables continuas, se aplica la prueba de kolmogorov-Smirnov a los datos de la muestra y el Censo, que proporciona una medida de la correlación entre las funciones de distribución de probabilidad acumulativa de la encuesta y el censo. Esta medida, es la distancia máxima entre la

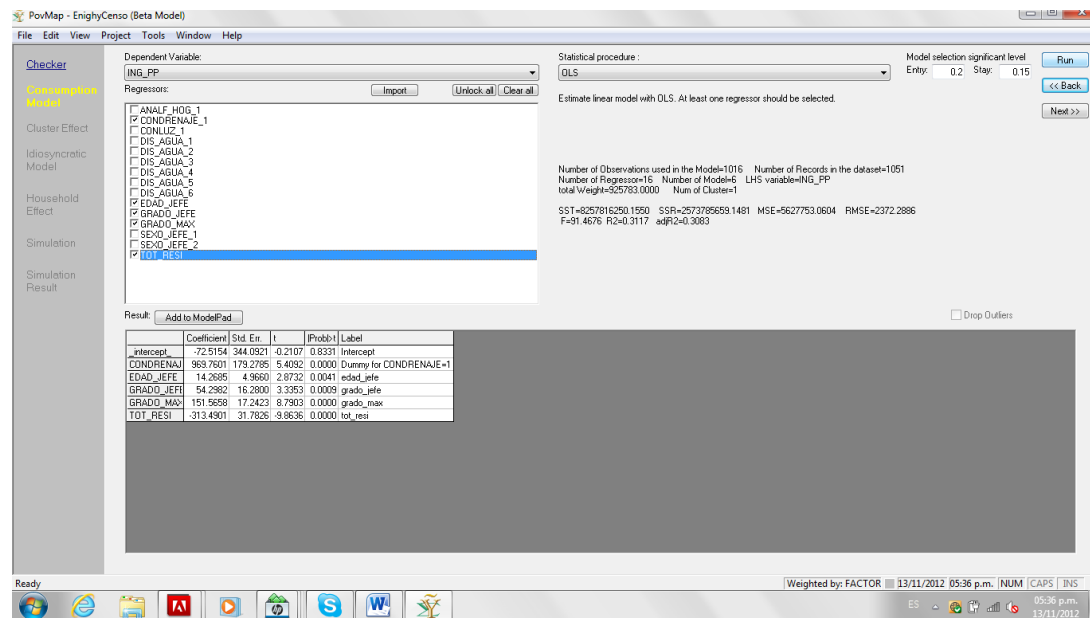
encuesta y distribución censo; valores pequeños indican que la variable de la encuesta es representativa de la variable del censo (Novales, 2006).

Figura 18.- Búsqueda de variables candidatas



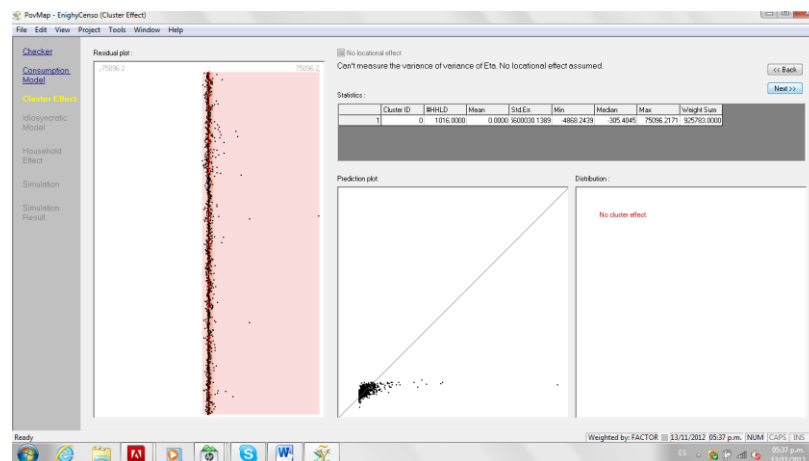
En la figura 19 se muestra el modelo de consumo. Las funciones y rutinas incorporadas en este módulo son para la búsqueda de un modelo de consumo que proporcione la mejor predicción con datos de la encuesta. Los elementos básicos manejados aquí son regresores. El regresor podría ser una variable continua, o un grupo de variables ficticias generadas a partir de una variable categórica.

Figura 19. Modelo de consumo



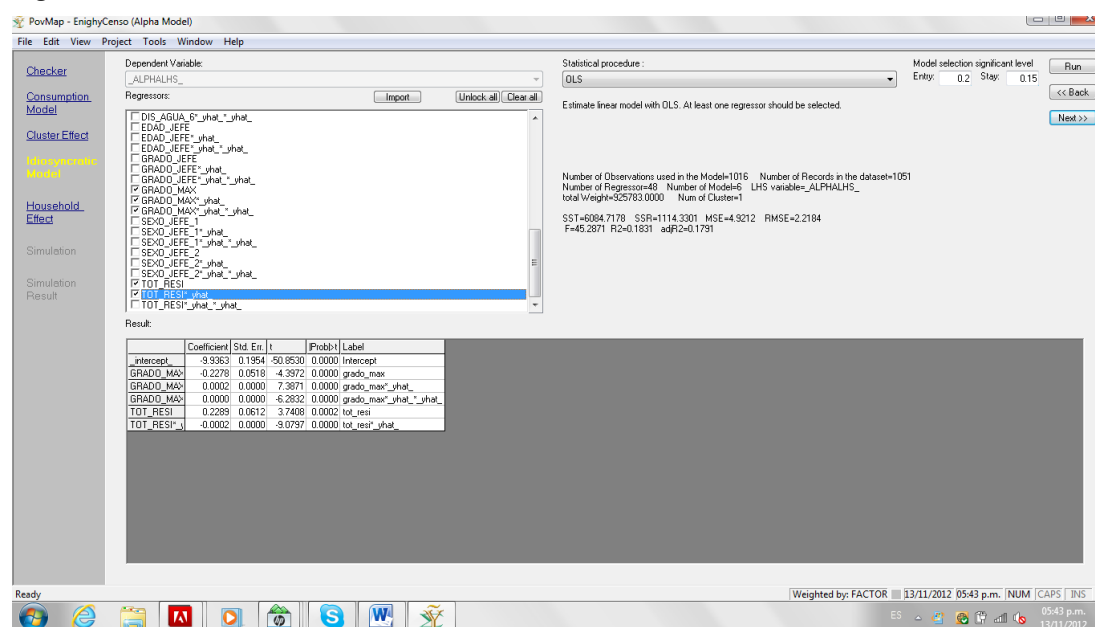
Esta opción, muestra el efecto a nivel de clúster y permite a los usuarios deshabilitar su efecto y ubicación además de determinar la distribución de la componente de localización que se aplicará a los datos del censo durante la simulación, ver figura 20.

Figura 20. Efecto cluster



El modelo idiosincrático.- También se conoce como el modelo alfa. El modelo idiosincrático realiza las estimaciones del hogar. Dado que la varianza del hogar no es constante, se permite que varíe con algunas variables explicativas considerando un modelo de heteroscedasticidad. La variable dependiente, identificada como `_ALPHALHS`, sólo se ve afectada por variables cuyo valor afecta a la varianza del término de error, por lo tanto, es posible estimar los parámetros utilizando una regresión paso a paso. Siguiendo los modelos de heteroscedasticidad, se generaran regresores a partir de las variables coincidentes y los términos adicionales que implican interacción $y^{\wedge}$ y $y^{\wedge 2}$.

Figura 21. Modelo idiosincrático.

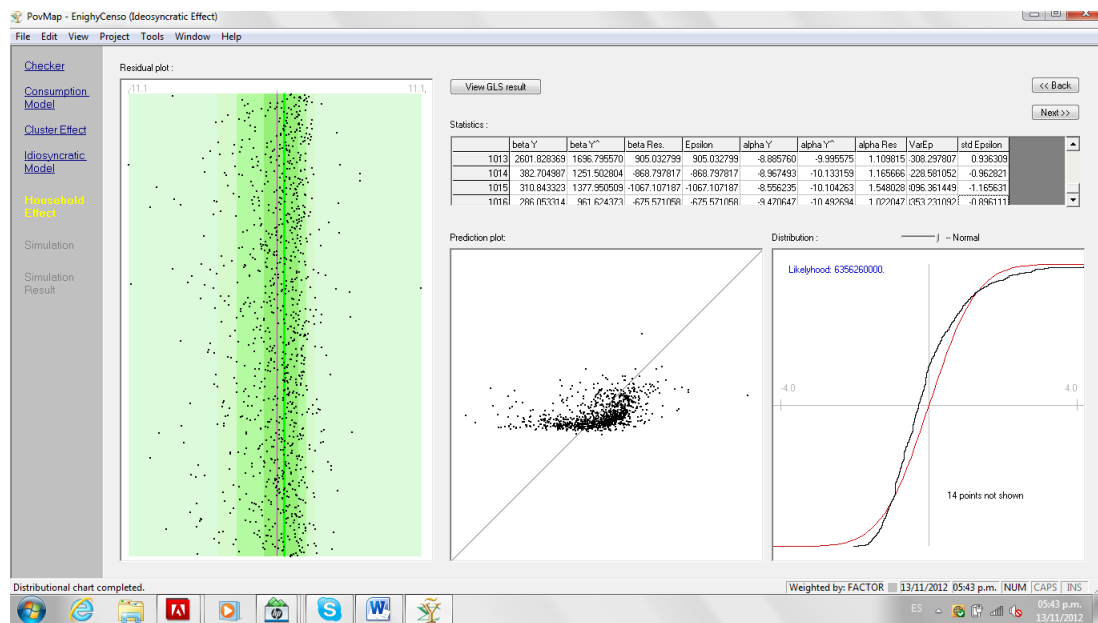


La opción *efecto hogar*, examina el grado de variación heteroscedasticos que puede explicarse por el modelo idiosincrático. Esto es considerando los siguientes 4 elementos:

- El gráfico de los residuales de alfa
- Tabluación de las estimaciones de Alpha Beta y el hogar
- Un gráfico de dispersión de la actual (αy) y la estimación ($\alpha \hat{y}$) para cada hogar, y
- La distribución acumulada la aproximación de los términos de error.

Sin embargo, a pesar de toda la información proporcionada, solo se logran pequeños ajustes en la predicción del ingreso del hogar, sin embargo, la predicción puede ser mejorada si se utiliza una muestra mayor de datos de la encuesta, ver figura 22.

Figura 22. Efecto hogar



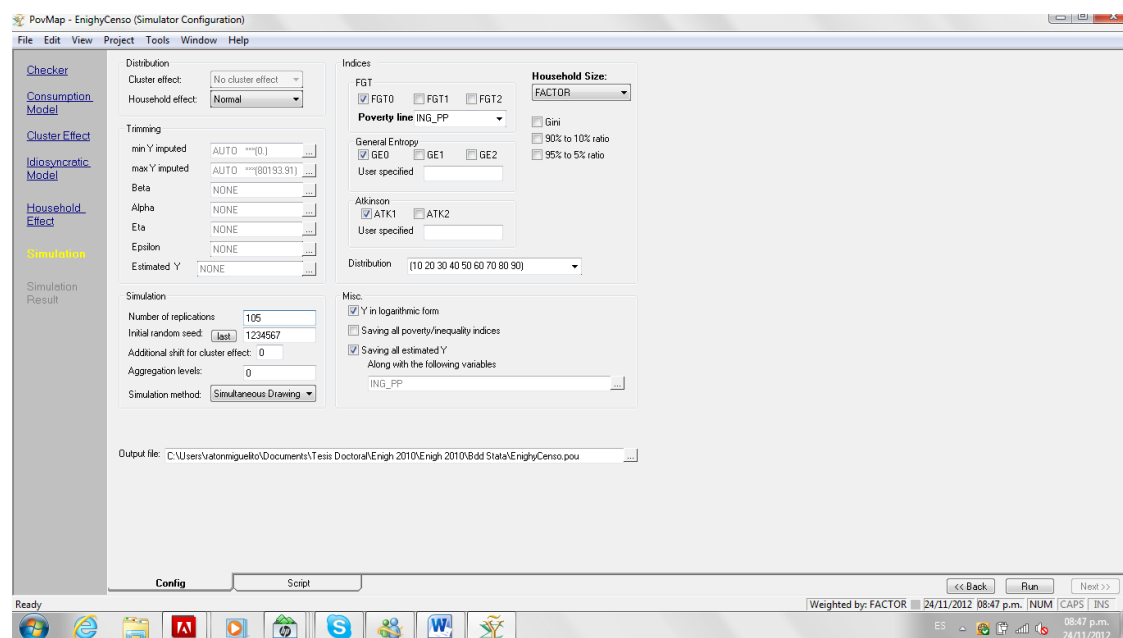
El proceso de imputación, termina con la simulación de los datos, esto se refiere al punto en el mapeo de la pobreza cuando las estimaciones de los parámetros y el error

de la encuesta se aplicó a los datos del censo. Esta opción permite a los usuarios modificar la configuración considerando:

- Cambiar la distribución del cluster y el efecto idiosincrásico,
- Establecer los parámetros para el tamaño de la simulación,
- Especificar los parámetros de simulación,
- Especificar información adicional para los datos de salida, y
- La identificación del tipo y nivel de simulación.

PovMap proporciona un archivo de salida con formato PCF, con la opción a ser exportado a formato texto.

Figura 23. Simulación



Finalmente, el sistema proporciona los resultados de la simulación, la página resumen proporciona una visión general de la configuración del modelo de simulación y del registro de la generación de números aleatorios. Los resultados son mostrados en un formato similar a una hoja de cálculo. Las columnas incluyen el identificador del hogar el valor mínimo y máximo imputado y la imputación de Y (YHAT), ver figura 24.

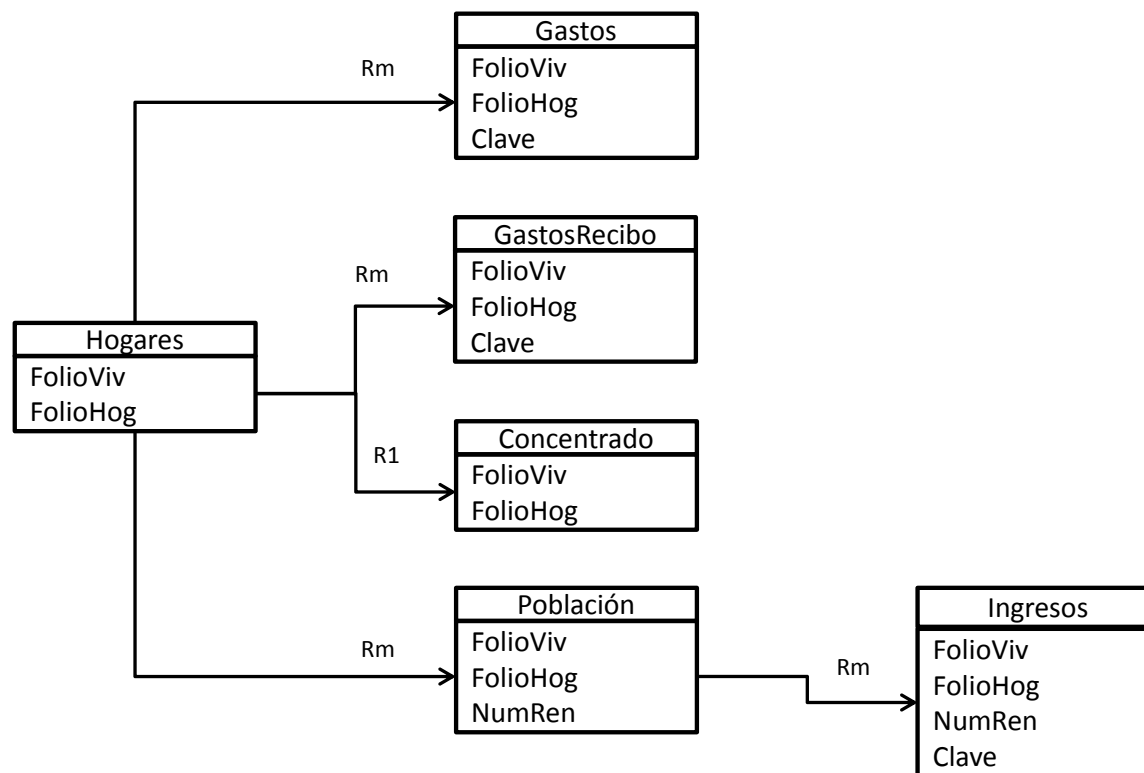
Los resultados pueden ser exportados para su tratamiento a un archivo con formato *.dat, *.dbf, *.dta

Figura 24. Resultados de la simulación

ID	HH\$	HH\$PP	YHAT
363357	2054417149057.0000	1.0000	2338.4355
363358	2054417149059.0000	1.0000	1590.7550
363359	2054417149061.0000	1.0000	1829.4153
363360	2054417149064.0000	1.0000	1275.4523
363361	2054417149067.0000	1.0000	1204.1631
363362	2054417149070.0000	1.0000	1132.8739
363363	2054417149071.0000	1.0000	1700.8065
363364	2054417149072.0000	1.0000	1345.4934
363365	2054417809636.0000	1.0000	1409.4154
363366	2054417809641.0000	1.0000	1210.1152
363367	2054417809643.0000	1.0000	705.5953
363368	2054417817257.0000	1.0000	1353.8804
363369	2054417817294.0000	1.0000	1975.0790
363370	2054417817295.0000	1.0000	1602.8262
363371	2054417817302.0000	1.0000	605.6906
363372	2054417817311.0000	1.0000	1072.4691
363373	2054417817314.0000	1.0000	1196.1456
363374	2054417817316.0000	1.0000	1592.6807
363375	2054417817318.0000	1.0000	1814.7069
363376	2054417817320.0000	1.0000	2053.5071
363377	2054417820127.0000	1.0000	1592.3680
363378	2054417820178.0000	1.0000	1624.3029
363379	2054417820179.0000	1.0000	1790.7941
363380	2054417820180.0000	1.0000	1546.2455
363381	2054417820181.0000	1.0000	1121.5415
363382	2054417820182.0000	1.0000	1290.9781
363383	2054417820183.0000	1.0000	1263.3824
363384	2054417820184.0000	1.0000	2004.4333
363385	2054417820185.0000	1.0000	1602.8262
363386	2054417820532.0000	1.0000	1725.4569
363387	2054417820536.0000	1.0000	1092.1871

Anexo E

Diagrama de relación del modelo multilogístico Nacional
utilizando datos de la ENIGH 2010.

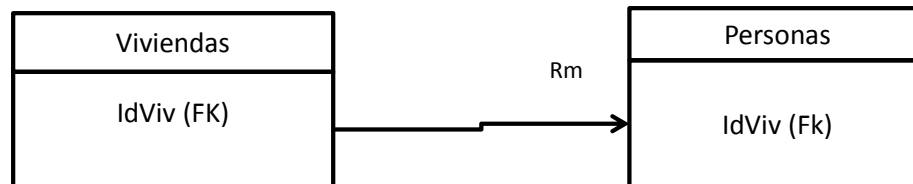


R1: Relación 1 a 1.

Rm: Relación 1 a muchos

Anexo F

Diagrama de relación para la explotación de las bases de datos del Censo de Población y Vivienda 2010



R1: Relación 1 a 1.
Rm: Relación 1 a muchos
Fk: Llave primaria

Anexo G

Producción científica.

Artículos producidos.	
Título	Universidad/Institución
Perspectivas del recurso agua en el estado de hidalgo	Libro: retos alternativas de desarrollo para México ante la crisis agroalimentaria mundial. Universidad Autónoma de Chapingo & Universidad Politécnica de Tulancingo. ISBN: 978-607-12-0150-8 Impreso en Octubre 2010
Análisis econométrico del sector agua potable a partir de la ENIGH 2008 que permita el desarrollo de planes y estrategias para un manejo eficiente de los recursos hídricos	Memorias: tercer coloquio interdisciplinario de doctorado “innovación e investigación en la solución de problemas sociales”, universidad popular autónoma de puebla ISBN: 978-607-95186-8-4
Diagnóstico del desarrollo y disponibilidad del agua en el estado de Hidalgo.	Miguel Ángel Torres González, en Tomás Serrano Aviles y Zeus Salvador Hernández Veleros (coordinadores). Estudios Demográficos del Estado de Hidalgo, Tomo II (Cambio Social y Económico). Universidad Autónoma del Estado de Hidalgo, 2010, 253 pp. ISBN: 978-607-482-188-8.

Artículos producidos.	
Título	Universidad/Institución
Situación de la cobertura del servicio de agua potable en los Municipios del Estado de Hidalgo a partir del Censo de población y Vivienda del 2000.	Miguel Ángel Torres González y Antonio Sánchez de la Vega en Tomás Serrano Aviles y Zeus Salvador Hernández Veleros (coordinadores). Estudios Demográficos del Estado de Hidalgo, Tomo II (Cambio Social y Económico). Universidad Autónoma del Estado de Hidalgo, 2010, 253 pp. ISBN: 978-607-482-188-8.
Velocidad, aceleración y ritmo económicos en las actividades empresariales: La toma de decisiones mediante simulaciones computacionales.“	Nuevas estrategias competitivas, Universidad Autónoma del Estado de Hidalgo. ISBN 978-607-482-292-2 (a publicarse en 2013).

Producción científica.

Soporte documental.

2.6 PERSPECTIVAS DEL RECURSO AGUA EN EL ESTADO DE HIDALGO

Miguel Ángel Torres González¹

*¹ Candidato a Doctor por la Universidad Popular
Autónoma de Puebla
mief2000@hotmail.com*

INTRODUCCIÓN

Uno de los retos de mayor trascendencia ante los que se encuentra la sociedad en la actualidad, es el manejo de los recursos naturales, donde el agua, emerge como el principal elemento que se relaciona ampliamente en todos los aspectos de la civilización

A través del tiempo, cuando el ser humano se estableció en comunidades en espacio y tiempo definidos, la necesidad de dotarse del agua se convirtió en primicia, llevándolo a adoptar al inicio mecanismos naturales, mismos que con el paso del tiempo se depuraron, fundamentados en técnicas y metodologías basadas en el conocimiento científico. Aunado a la tecnología, el agua se constituyó como factor de crecimiento religioso, social y político de los pueblos, ya que por su misma naturaleza desde el principio fue contemplada como símbolo de pureza así como de salud.

Con el paso del tiempo, y debido al crecimiento en ocasiones anárquico de la población, y la aparición de la revolución industrial que tuvo por consecuencia un alto impulso en la creación de procesos más automatizados, las expectativas de suministro de agua se vieron rebasadas, y al mismo tiempo se comenzó a gestar un problema hasta entonces poco

Alternativas de desarrollo para México ante la crisis agroalimentaria mundial



Alma Velia Garay • Amparo Nidia Castillo Santos • César Santiago Tepantlán • Gustavo Almaguer Vargas



Alternativas de desarrollo para
México ante la crisis
agroalimentaria mundial

ISBN: 978-607-12-0150-8



9 786071 201508



UNIVERSIDAD
POLITÉCNICA
TULANCINGO



"Enseñar la explotación de
la tierra no la del hombre"



UNIVERSIDAD POLITÉCNICA DE TULANCINGO

Otorga el presente

Reconocimiento

Foro Nacional... En busca de alternativas
a: M.I.E.F. MIGUEL ÁNGEL TORRES GONZÁLEZ

POR SU PARTICIPACIÓN COMO PONENTE EN EL

**FORO NACIONAL: ALTERNATIVAS DE DESARROLLO PARA MÉXICO
ANTE LAS CRISIS ECONÓMICA Y AGROALIMENTARIA MUNDIAL**

M.C. Luis Téllez Reyes
Rector





UNIVERSIDAD POLITÉCNICA DE TULANCINGO

Otorga el presente

Reconocimiento

**Mtro. Miguel Ángel Torres
González**

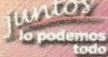
a:

*Por su participación como ponente en el 2do. Encuentro de
Investigación de la Universidad Politécnica de Tulancingo*

Tulancingo de Bravo, Hgo., 20 de octubre de 2009.



M.C. Luis Téllez Reyes
Rector



Zempoala Hidalgo, a 15 de Julio de 2011

M. A. TORRES GONZÁLEZ, S. GÓMEZ GÓMEZ, M. A. VERA JIMÉNEZ, E. OLMOS BLANQUEL, A. SÁNCHEZ DE LA VEGA

Por medio de la presente, se informa que su contribución titulada **“SIMULACIÓN APLICADA AL RIESGO EN PROYECTOS DE INVERSIÓN UN CASO DE ESTUDIO”** identificada con el código de recepción TIF25 ha sido ACEPTADA para ser presentada en la modalidad de PRESENTACIÓN ORAL en el Simposio Iberoamericano Multidisciplinario de Ciencias e Ingenierías 2011 (SIMCI-2011), que se llevará a cabo del 21 al 23 de Septiembre de 2011 en el Centro de Convenciones TuzoForum, Pachuca, Hidalgo.

Si desea que el trabajo en extenso sea publicado en la **Revista Iberoamericana de Ciencias e Ingenierías Emergentes con ISBN: 978-607-95344**, será necesario enviarlo para su evaluación en el formato que se encuentra disponible en <http://www.upp.edu.mx/simci/presentacion-de-articulos.php>. La fecha límite de recepción de trabajos es el 1 de Septiembre del presente año.

Para mayor información puede comunicarse con el Comité Organizador del evento al correo: simci@site.upp.edu.mx

Sin otro particular, reciba saludos cordiales.

ATENTAMENTE



Dr. Marco Antonio FLORES-GONZÁLEZ
COMITÉ ORGANIZADOR

Zempoala Hidalgo, a 08 de Agosto de 2011

M. A. VERA JIMÉNEZ, S. GÓMEZ GÓMEZ, M. A. TORRES GONZÁLEZ, Y E. OLMOS BLANQUEL

Por medio de la presente, se informa que su contribución titulada **“Comercio exterior y la educación, aplicación actual”** identificada con el código de recepción TIF29 ha sido ACEPTADA para ser presentada en la modalidad de POSTER en el Simposio Iberoamericano Multidisciplinario de Ciencias e Ingenierías 2011 (SIMCI-2011), que se llevará a cabo del 21 al 23 de Septiembre de 2011 en el Centro de Convenciones TuzoForum, Pachuca, Hidalgo.

Si desea que el trabajo en extenso sea publicado en la **Revista Iberoamericana de Ciencias e Ingenierías Emergentes con ISBN: 978-607-95344**, será necesario enviarlo para su evaluación en el formato que se encuentra disponible en <http://www.upp.edu.mx/simci/presentacion-de-articulos.php>. La fecha límite de recepción de trabajos es el 1 de Septiembre del presente año.

Para mayor información puede comunicarse con el Comité Organizador del evento al correo: simci@site.upp.edu.mx

Sin otro particular, reciba saludos cordiales.

ATENTAMENTE



Dr. Marco Antonio FLORES-GONZÁLEZ
COMITÉ ORGANIZADOR

Zempoala Hidalgo, a 08 de Agosto de 2011

S. GÓMEZ GÓMEZ, M. A. TORRES GONZÁLEZ, M. A. VERA JIMÉNEZ Y E. OLMOS BLANQUEL

Por medio de la presente, se informa que su contribución titulada **“Análisis de la Ventaja Competitiva en la Industria Textil”** identificada con el código de recepción TIF27 ha sido ACEPTADA para ser presentada en la modalidad de POSTER en el Simposio Iberoamericano Multidisciplinario de Ciencias e Ingenierías 2011 (SIMCI-2011), que se llevará a cabo del 21 al 23 de Septiembre de 2011 en el Centro de Convenciones TuzoForum, Pachuca, Hidalgo.

Si desea que el trabajo en extenso sea publicado en la **Revista Iberoamericana de Ciencias e Ingenierías Emergentes con ISBN: 978-607-95344**, será necesario enviarlo para su evaluación en el formato que se encuentra disponible en <http://www.upp.edu.mx/simci/presentacion-de-articulos.php>. La fecha límite de recepción de trabajos es el 1 de Septiembre del presente año.

Para mayor información puede comunicarse con el Comité Organizador del evento al correo: simci@site.upp.edu.mx

Sin otro particular, reciba saludos cordiales.

ATENTAMENTE



Dr. Marco Antonio FLORES-GONZÁLEZ
COMITÉ ORGANIZADOR



SIMPOSIO IBEROAMERICANO MULTIDISCIPLINARIO DE CIENCIAS E INGENIERÍAS

La Universidad Politécnica de Pachuca, otorga el presente

Reconocimiento a:

M. A. Torres González, S. Gómez Gómez, M. A. Vera Jiménez,
E. Olmos Blanquel, A. Sánchez de la Vega

Por haber participado con la exposición oral “SIMULACIÓN APLICADA AL RIESGO EN PROYECTOS DE INVERSIÓN UN CASO DE ESTUDIO”, en el marco del **Simposio Iberoamericano Multidisciplinario de Ciencias e Ingenierías**, efectuado del 21 al 23 de septiembre de 2011, en la ciudad de Pachuca, Hidalgo, México.


M.A.E. Sergio Alejandro Arteaga Carreño
Rector





SIMPOSIO IBEROAMERICANO MULTIDISCIPLINARIO DE CIENCIAS E INGENIERÍAS

La Universidad Politécnica de Pachuca, otorga el presente

Reconocimiento a:

**M. A. Vera Jiménez, S. Gómez Gómez,
M. A. Torres González, y E. Olmos Blanquel**

Por haber participado con el póster "COMERCIO EXTERIOR Y LA EDUCACIÓN, APLICACIÓN ACTUAL", en el marco del Simposio Iberoamericano Multidisciplinario de Ciencias e Ingenierías, efectuado del 21 al 23 de septiembre de 2011, en la ciudad de Pachuca, Hidalgo, México.



Dr. Marco Antonio Flores González
Presidente del Comité Organizador del SIMCI



La Universidad Politécnica de Pachuca, otorga el presente

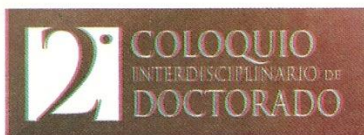
Reconocimiento a:

**S. Gómez Gómez, M. A. Torres González,
M. A. Vera Jiménez y E. Olmos Blanquel**

Por haber participado con el póster “ANÁLISIS DE LA VENTAJA COMPETITIVA EN LA INDUSTRIA TEXTIL”, en el marco del Simposio Iberoamericano Multidisciplinario de Ciencias e Ingenierías, efectuado del 21 al 23 de septiembre de 2011, en la ciudad de Pachuca, Hidalgo, México.



Dr. Marco Antonio Flores González
Presidente del Comité Organizador del SIMCI



Otorgan la presente

CONSTANCIA

a

Miguel Ángel Torres González

por su participación como: Asistente

“2° Coloquio Interdisciplinario de Doctorado”

Con una duración de 8 horas

Puebla, Pue., A 24 de Abril de 2009

“La Cultura al Servicio del Pueblo”

José Pablo Nuño de la Parra
Vicerrector de Posgrados e Investigación

Mayte Alcántara Gutiérrez
Directora de Posgrados



La Universidad Autónoma del Estado de Hidalgo, otorga el presente

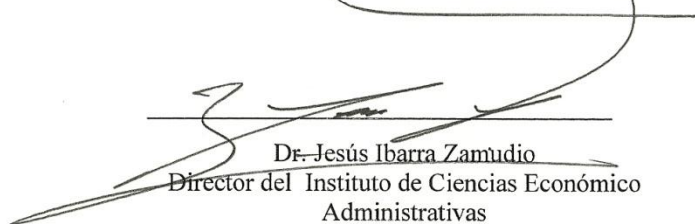
RECONOCIMIENTO al

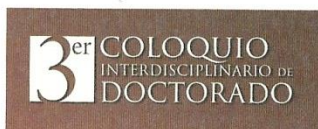
Mtro. Miguel Torres González

por su valiosa participación con la ponencia titulada:

***"Los pobres no tan pobres en México: ejemplo de errores de medición
a partir de la encuesta nacional del ingreso/gasto de los hogares
2008"***

Presentada en el Segundo Congreso de Investigación Económica y Políticas Públicas y Primer Foro Estudiantil de Investigación, efectuada en el Instituto de Ciencias Económico Administrativas de la UAEH, el 10 al 13 de noviembre de 2009 en la ciudad de Pachuca, Hidalgo.


Dr. Jesús Ibarra Zamudio
Director del Instituto de Ciencias Económico
Administrativas



Otorgan la presente

CONSTANCIA

a

Miguel Ángel Torres González

por su valiosa participación como: **Ponente** del:

Tercer Coloquio Interdisciplinario de Doctorado
"Innovación e Investigación en la Solución de Problemas Sociales"

Con una duración de 8 horas

Puebla, Pue., A 19 de Marzo de 2010

"La Cultura al Servicio del Pueblo"

José Pablo Nuño de la Parra
Vicerrector de Posgrados e Investigación

Mayte Alcántara Gutiérrez
Directora de Posgrados

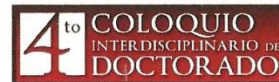


Universidad Popular Autónoma del Estado de Puebla
otorga la presente

CONSTANCIA

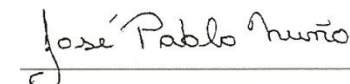
a MIGUEL ÁNGEL TORRES GONZÁLEZ

por su valiosa participación en el
4to Coloquio Interdisciplinario de Doctorado



como PONENTE

Puebla, Pue., México, 8 y 9 de abril de 2011.


Dr. José Pablo Nuño De la Parra
VICERRECTOR DE POSGRADO E INVESTIGACIÓN


Dra. María Esther Alcántara Gutiérrez
DIRECTORA DE POSGRADO



El Instituto Tecnológico Superior del Oriente del Estado de Hidalgo. ITESA

Otorga la presente

CONSTANCIA

Al: M. en E. Miguel Ángel Torres González

Por haber participado en la

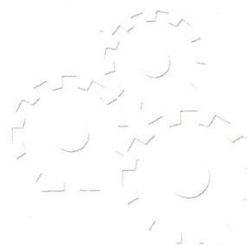


Semana de las Ingenierías y la Administración

Impartiendo la Conferencia:

"Aplicación de la Gestión Empresarial en el Sector Público y Privado"

Durante el Simposio *"Optimizando Recursos"*
de la carrera de Ingeniería en Gestión Empresarial



Mtro. Gerardo Tellez Reyes
Director General

Apan, Hidalgo, a 26 de Mayo de 2011

