
Análisis multivariante para el estudio socio-económico, parroquia Carmen de Pijili



Multivariate analysis for the socio-economic study - Carmen de Pijili

Geovanny Zamora Zamora.¹, Tito Max Banegas Peña.² María Eliza Flores Flores.³ & Jorge Edison Becerra Molina⁴

Abstract.

DOI: <https://doi.org/10.33262/cienciadigital.v3i2.3.489>

The complexity of the phenomena of the sciences in general means that researchers are forced to face complex, multivalent problems, in which intervening and coexisting variables called predictors and the calls of results, response or events, which require typification between large volumes of data. For this reason, a socioeconomic study of the community is carried out, using the metric techniques to determine the social and economic situation using advanced concepts and tools for the comprehensive examination and interpretation of the results. The multivariate analysis (Martínez Arias, 2008) "is a tool that, in recent years has achieved great development in social science research, although this tool is not easily applicable to scholars". The objective of this work is to perform a multivariate analysis for the socio-economic study of the parish of Carmen de Pijili and its surrounding areas, information to determine the greatest needs, make the best decisions, plan public policies and project assistance to the most disadvantaged people and to optimize the quality of life of the inhabitants. From a universe of 580 families, a sample of 69 families was taken, which were consulted on socio-economic variables and local equipment. The database was processed and refined in the Dyana and SPSS software, to obtain the statistical indicators required in the social analysis

Keywords: Multivariate, socio-economic, social project.

Resumen.

La complejidad de los fenómenos de las ciencias en general hace que los investigadores se vean obligados a enfrentarse a problemas complejos, multivalentes, en los cuales intervienen y coexisten variables llamadas predictoras y las llamadas de resultados,

¹ Universidad Católica de Cuenca, Cuenca, Ecuador, ezamoraz@ucacue.edu.ec

² Universidad Católica de Cuenca, Cuenca, Ecuador, tbanegasp@ucacue.edu.ec

³ Universidad Católica de Cuenca, Cuenca, Ecuador, mfloresf@ucacue.edu.ec

⁴ Universidad Católica de Cuenca, Cuenca, Ecuador, jbecerram@ucacue.edu.ec

respuesta o eventos, que requieren ser tipificadas entre grandes volúmenes de datos. Por esta razón se realiza un estudio socioeconómico de la comunidad, empleando las técnicas métricas para determinar la situación social y económica utilizando conceptos avanzados y herramientas para el examen e interpretación integral de los resultados. El análisis multivariante (Martínez Arias, 2008) “es una herramienta que, en los últimos años ha alcanzado gran desarrollo en las investigaciones de las ciencias sociales, a pesar de no ser esta herramienta de fácil aplicación para los estudiosos”. El objetivo del presente trabajo tiene como fin realizar un análisis multivariante para el estudio socioeconómico de la parroquia Carmen de Pijili y sus zonas aledañas, información que permita determinar las grandes necesidades, tomar las mejores decisiones, planificar las políticas públicas y proyectar la asistencia social a las personas más desfavorecidas y poder optimizar la calidad de vida de los habitantes. De un universo de 580 familias, se tomó una muestra de 69 familias, mismas que fueron consultadas sobre variables de índole socio-económico y equipamiento local. La base de datos, fue procesada y depurada en el software Dyana y SPSS, para obtener los indicadores estadísticos requeridos en el análisis social.

Palabras claves: Multivariante, socio-económico, proyecto social.

Introducción.

En el Ecuador, para una planificación económica eficiente, en toda su magnitud, ya sea nacional, regional o local, que incida en el mejoramiento de la calidad de vida de sus habitantes, se requiere de información actualizada de su población y sus localidades, básicamente información socio-económica que dé cuenta de las condiciones de vida de la población y de la infraestructura básica con la que dispone (Sánchez, Lardé, Chauvet, & Jaimurzina, 2017). De esta manera, los organismos que toman decisiones planifican su quehacer diario de una manera eficaz, para redistribuir el ingreso a las diferentes localidades del país en forma de obras o equipamiento básico que permita el desarrollo local sostenido (*Informe sobre el crecimiento sostenido y el desarrollo incluyente*, s. f.).

En este sentido, a fin de redistribuir recursos del estado en obras de equipamiento que mejoren las condiciones de vida de las poblaciones, es preciso realizar estudios que permitan conocer las condiciones socio-económicas de dichas localidades rurales. Uno de los procedimientos que se considera como una herramienta más factible para realizar estos estudios en poblaciones dispersas que necesitan atención gubernamental, es el análisis multivariante (Martínez Arias, 2008).

El objetivo del presente trabajo es estudiar las condiciones socio-económicas de la parroquia Carmen de Pijilí utilizando análisis multivariante en combinación con el análisis descriptivo. Ello con la finalidad de hacer un levantamiento más real de las variables determinantes de

las condiciones socio-económicas en que vive dicha población y demostrar que el análisis multivariantes es un modelo más real y veraz de las variables que explican determinados fenómenos como los sociales.

Para establecer una línea base inicial de las condiciones socio-económicas de la parroquia del Carmen de Pijili, y, en lo posterior, medir el impacto que producirá la intervención de una entidad de gobierno al implementar un proyecto de desarrollo social, se utilizó la cartografía del INEC, considerada a la unidad de registro a la manzana y no la edificación.

Cabe indicar que existe el aporte teórico y metodológico importante que ofrece el análisis multivariante, por cuanto de acuerdo a los resultados obtenidos, este análisis muestra relaciones que no se pueden observar con un simple análisis descriptivo. Bajo este contexto se afirma que el análisis multivariante es un aporte y complemento importante del análisis descriptivo de estudios de carácter social. En primera instancia, se realiza un análisis estadístico descriptivo de las variables de condiciones socio-económicas de la localidad, y en un segundo momento; se realiza un análisis multivariante.

El presente artículo se estructura de la siguiente manera: la sección uno presenta el marco lógico que conste en la revisión teórica sobre el análisis multivariante y revisión de trabajos empíricos sobre el tema y sus principales hallazgos; la sección dos detalla la metodología y el uso del software Dyanna y SPSS, para poder obtener los indicadores estadísticos, así como los instrumentos utilizados para medir el impacto que producirá la intervención de una entidad de gobierno al implementar un proyecto de desarrollo social; la sección tres muestra los resultados encontrados basados en un análisis de significancia individual, a través de la t estadística. Finalmente, la sección cuatro presenta las conclusiones.

El estudio socio-económico no solo se toma en cuenta el salario o los ingresos de la parroquia del Carmen de Pijili, sino también las condiciones de vivienda, salubridad, número de habitantes por vivienda, la convivencia. El análisis de componentes principales (ACP) (León González, Llinás Solano, & Tilano, 2011) es una técnica estadística propuesta a principios del siglo XX por Hotelling (1933) quien se basó en los trabajos de Karl Pearson (1901) y en las investigaciones sobre ajustes ortogonales por mínimos cuadrados. Sin embargo, el análisis multivariantes según (León González et al., 2011) existen diversas definiciones acerca de ésta técnica de análisis de datos multivariados, pero los dos concuerdan en conceptualizarla como "una herramienta que tiene como objetivo principal resumir grandes cantidades de datos por medio de pocos parámetros (simplificación), además busca encontrar relaciones entre: las variables de respuesta y unidades experimentales

El análisis multivariante se puede clasificar, según (Loares Colina & López Roldán, 1991) en tres grupos: 1). - *Métodos de dependencia*: el objetivo de este método consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes, 2). - *Métodos de interdependencia*: su objetivo consiste en identificar qué

variables están relacionadas, cómo lo están y por qué, y; 3). - *Métodos estructurales*: el objetivo de estos métodos es analizar, no sólo como las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí.

Metodología.

Para establecer una línea base inicial, de las condiciones socio-económicas de la parroquia del Carmen de Pijilí, se ha realizado una investigación socio-económica de campo a los hogares de la parroquia indicada; debido a que la implementación de un proyecto social incide de manera directa a su población. Al fijar el tamaño de la muestra o número de encuestas deberán ser lo suficientemente representativo del universo de tal manera que se pueda generalizar los resultados a todo el universo investigado. Por lo tanto, para el 2018 se tendría 532 familias

Para el análisis social se ha realizado una encuesta con miras a levantar información de la población, el levantamiento de la información se realizará los días sábado 6 y domingo 7 de enero del 2018, las variables analizadas dieron cuenta de los rasgos socio-económicos del hogar, características de la vivienda, dotación de servicios básicos, salubridad, ingresos y gastos de la familia, que fueron las principales variables investigadas.

En el Tamaño de la muestra se utiliza la fórmula del error estándar de la población, para una población finita de menos de 100.000 habitantes, en el caso de la parroquia El Carmen de Pijilí, y; de acuerdo a la población de 2.418 habitantes del año de 2018 se tendría un valor que está dentro del rango y valida el método utilizado. Para el año de 2018 las familias son de 532 (con un promedio de 4,3 habitantes por familia), se ha calculado una muestra mínima de 69 encuestas, la misma que representa un 12.27% del total de la población.

ENCUESTA ECONOMICA SOCIAL: Fórmula de la muestra

Detalle	Fórmula
Fórmula 1	$n = \frac{n'}{1 + \frac{n'}{N}}$
Fórmula 2	$n' = \frac{s^2}{\sigma^2}$
Fórmula 3	$s^2 = p(1-p)$ $\sigma^2 = (se)^2$

ENCUESTA ECONOMICA SOCIAL: Parámetros de construcción de la muestra mínima

Tamaño de la muestra	n = 69
Tamaño de la población	N = 532
Varianza de la muestra	S ² = 0.09
Varianza poblacional	σ^2 = 0.001156
Error estándar	Se = 3.4%
Nivel de confianza	p = 95%

En donde:

El S² = 0.09 varianza muestra; Nivel de confianza = Corresponde al nivel de confianza para que la muestra sea representativa del universo, la cual representa un 95% de confianza; Se = Error estándar de la muestra, se ha considerado un valor del 3.4% como aceptable; El universo = Está formado por los hogares estimados, considerando las proyecciones de la población para toda la vida útil del proyecto, para el año de 2018 se ha estimado un universo integrado por 532 hogares.

Para la realización de la encuesta socio-económica se partió de un supuesto básico, que conduce a una implicación inevitable, se carece de una base muestral completa y actualizada en donde se especifique la ubicación y el número de viviendas, por lo cual, cualquier procedimiento de muestreo aleatorio que se utilice será siempre imperfecto. Como se mencionó se utilizó la cartografía del INEC, pero, en esta cartografía la unidad de registro es la manzana y no la edificación, por ello se infirió la existencia de un número determinado de edificaciones por cada manzana. De esta manera se procedió de acuerdo con el plano a seleccionar las manzanas a partir de un cálculo de un promedio de 5 edificaciones por manzana, las mismas que existe alrededor de 17 manzanas. Seguidamente, respetando las manzanas de la cartografía, se dividieron para un total de 8.6 manzanas, dando un resultado de 2 (salto), es decir; de cada manzana se eligió una vivienda que fue encuestada.

De esta manera, se seleccionó a la población a encuestar organizada por manzanas, a partir de un cálculo de un promedio, a través de un procedimiento de selección muestral por azar sistemático para garantizar la distribución homogéneamente en toda la zona de la parroquia. Para la aplicación de las encuestas se respetó la numeración de las manzanas del plano del INEC, se fueron escogiendo una manzana, de cada 2, adicionando sistemáticamente el coeficiente de elevación, tantas veces fuese necesario, hasta completar el tamaño de manzanas requeridas. Mientras que en las zonas dispersas las encuestas se realizaron, a través de la búsqueda directa de las viviendas, puesto que en estas localidades no existe un registro cartográfico definido. En este caso se procedió a realizar un muestreo aleatorio simple.

Variables e indicadores de la investigación

Las variables estudiadas fueron:

- Identificación y ubicación del hogar: número de ficha, status de hogar, dirección de la

vivienda, uso de la vivienda, miembros del hogar, residen. migración.

- Característica de la edificación: tipo de materiales de la vivienda.
- Infraestructura básica: servicios de agua potable, servicios de la vivienda, pago del agua, eliminación de aguas servidas, servicios de recolección de basura, estado del sistema de alcantarillado u otros.
- Agua potable: conexión domiciliaria, continuidad en el servicio, equipamiento de medidores, abastecimiento de agua potable.
- Salubridad: miembros del hogar enfermos, tipo de enfermedad.
- Ingresos y Egresos del hogar: ingresos principales del hogar, gastos del hogar.

La digitación de la información se realizó en un software diseñado exclusivo para el caso (en Dyana), generando una base de datos, para el posterior estudio. Con la ayuda del programa SPSS se diseñaron las tablas que sirvieron para organizar los datos y proceder al análisis estadístico indicado.

Resultados.

A. Resultados obtenidos aplicando análisis clásico mono o bivalente

- Estructura del hogar

Considerando que el tamaño promedio de los hogares es de 4,3 personas, en el cuadro siguiente, se puede observar que la población del área de estudio es relativamente joven. Aproximadamente el 49,4% de la población es menor de 20 años, el 38,3% tiene entre 20 a 50 años, y el 12,3% tienen 50 años y más. La mayoría de la población es masculina. La mayor parte de los hogares saben leer y escribir 79,5%, no existe discriminación de sexo, en cuanto a la educación, puesto que el porcentaje anotado incluye a hombres y mujeres.

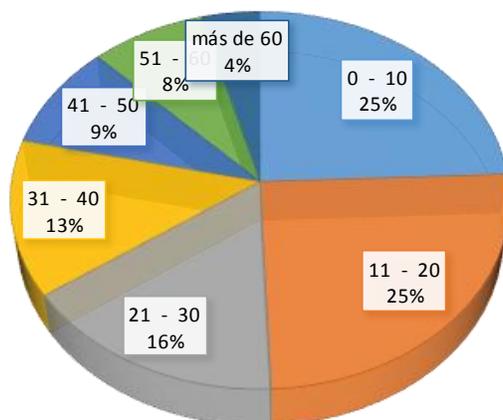
Tabla 1. Edad de los miembros del hogar según observaciones y porcentajes

Años cumplidos	Frecuencias	Porcentaje	% acumulado
0 - 10	75	24,35%	24,35%
11 - 20	77	25,00%	49,35%
21 - 30	49	15,91%	65,26%
31 - 40	41	13,31%	78,57%
41 - 50	28	9,09%	87,66%
51 - 60	25	8,12%	95,78%
más de 60	13	4,22%	100,00%
TOTAL	308	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Gráfico 1. Edad de los miembros del hogar según observaciones y porcentajes



Fuente: Encuesta aplicada socio-económica

- Renovación de la estructura por edades

El índice de renovación de la estructura por edades indica el número de menores de 15 años en relación con los mayores de 65. Se puede apreciar que, para toda el área, el índice señala que hay 8,7 niños por cada persona de edad avanzada, en el 2018. Evidentemente, se trata de una población con un alto índice de renovación de la población, probablemente se deba a procesos migratorios en la zona.

- Jefatura de hogar

La mayoría de hogares del área del proyecto tienen al padre como el jefe del hogar, seguido por la jefatura de la madre, en menor proporción los abuelos y los hijos/as.

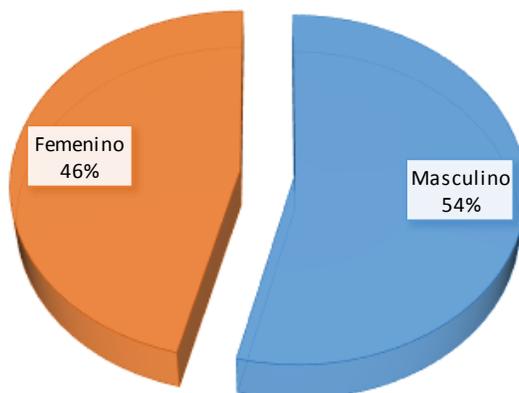
Tabla 2. Sexo de los miembros del hogar según observaciones y porcentajes

Sexo de los miembros del hogar	Frecuencias	Porcentaje	% acumulado
Masculino	166	53,90%	53,90%
Femenino	142	46,10%	100,00%
TOTAL	308	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Gráfico 2. Sexo de los miembros del hogar según observaciones y porcentajes



Fuente: Encuesta aplicada socio-económica

○ Migración

La migración se define como "el desplazamiento (duradero o permanente) de personas entre el lugar de origen o de partida y un lugar de destino. Hablar de migración hace referencia a dos procesos que se deben diferenciar: la inmigración que se refiere a la población que llega a un determinado territorio; y la emigración, se describe a la población que sale de su lugar de origen hacia otro territorio.

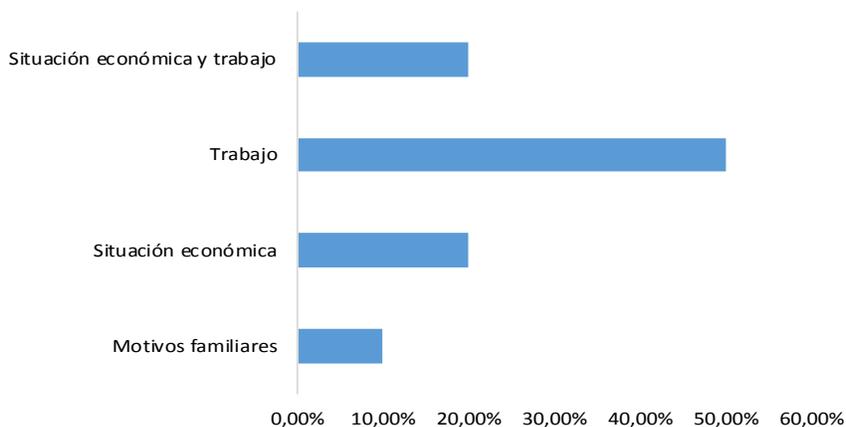
Tabla 3. Motivos de la Migración de los miembros del hogar según observaciones y porcentajes

Motivo de viaje	Frecuencias	Porcentaje	% acumulado
Motivos familiares	1	10,00%	10,00%
Situación económica	2	20,00%	30,00%
Trabajo	5	50,00%	80,00%
Situación económica y trabajo	2	20,00%	100,00%
TOTAL	10	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Gráfico 3. Motivos de la Migración de los miembros del hogar según observaciones y porcentajes



Fuente: Encuesta aplicada socio-económica

Del total de hogares con residentes en el exterior, el 15,9% tiene a uno de sus miembros, hombre o mujer fuera del país. Existen hogares hasta con 4 miembros varones que viven en el exterior.

○ Educación

El nivel de instrucción primaria se corresponde con la educación básica en la mayor parte de la población del área del proyecto. En la mayoría de los hogares se encuentran muchos de sus miembros que no han llegado a terminar este nivel de instrucción. El 23,4% dice no terminar la primaria y el 21,8% no tiene ningún tipo de formación.

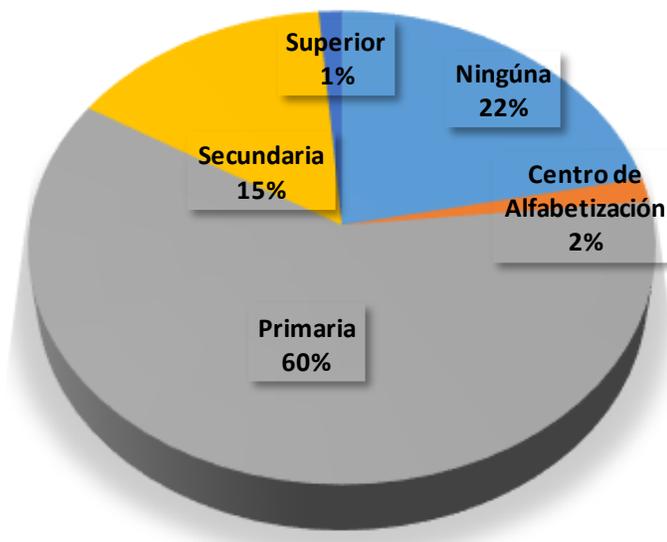
Tabla 4. Nivel de instrucción de los miembros del hogar según observaciones y porcentajes

Nivel de Instrucción	Frecuencias	Porcentaje	% acumulado
Ninguna	67	21,75%	21,75%
Centro de Alfabetización	5	1,62%	23,38%
Primaria	186	60,39%	83,77%
Secundaria	46	14,94%	98,70%
Superior	4	1,30%	100,00%
TOTAL	308	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Gráfico 4. Nivel de instrucción de los miembros del hogar según observaciones y porcentajes



Fuente: Encuesta aplicada socio-económica

Con una educación secundaria incompleta llegan el 14,9% de la población, este nivel es considerado como el de mayor aptitud para encontrar trabajo en las diferentes ocupaciones existentes en la economía del cantón o cantones. Ciertamente, solamente es necesario buscar la demanda de trabajadores en los diferentes periódicos de la ciudad de Cuenca para establecer que siempre se solicita nivel de instrucción secundaria, aún en casos de empleos no profesionales.

Al nivel de instrucción superior llega un 1,3% con instrucción superior; personas que en su mayoría no obtuvieron el título profesional. El nivel de postgrado no es común en nuestro medio por lo que ninguno ha realizado un postgrado.

La principal ocupación de la mayoría de jefes de hogar es como trabajadores agrícolas (80%), es decir, se ocupan en actividades como la agricultura, la ganadería, la horticultura, etc., las mismas que no requieren mayores niveles de instrucción y, además, por la mala calidad de los recursos, tampoco son rentables. Así mismo, el 1,3% son microempresarios.

- Salud

Como se observa en el cuadro adjunto, alrededor del 30,43% de los hogares han declarado que por lo menos uno de sus miembros ha enfermado en el último mes. La alta persistencia de enfermedades en la población del área del proyecto, se convierte en un problema no solo

de salud, sino además de carácter social y de demanda de servicios de salud de acuerdo al cuadro patológico que como se puede observar es muy diversificado.

El tipo de enfermedades de la población del área del proyecto, está muy asociada a problemas ambientales, pues en la mayoría de los casos se tratan de enfermedades infectas contagiosas, es decir, provenientes del medio ambiental en que reside cada uno de los hogares.

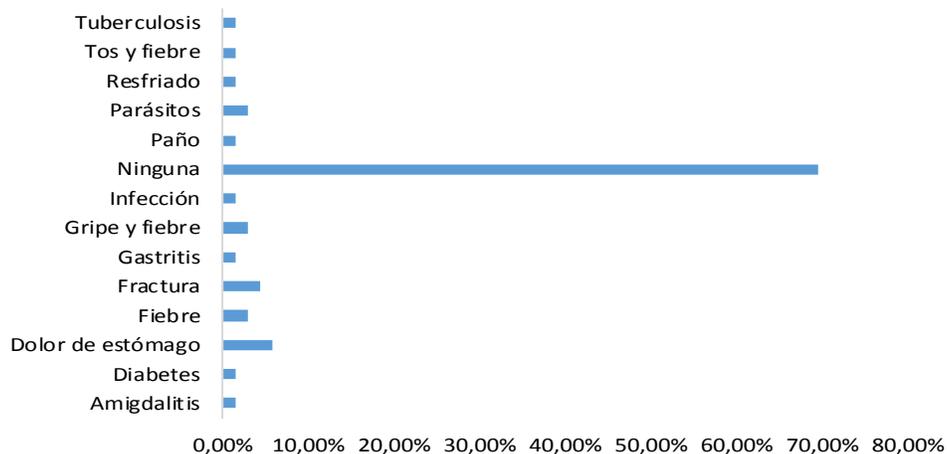
Tabla 5. Tipo de enfermedad que tuvo el enfermo según observaciones y porcentajes

Enfermedad	Frecuencias	Porcentaje	% acumulado
Amigdalitis	1	1,45%	1,45%
Diabetes	1	1,45%	2,90%
Dolor de estómago	4	5,80%	8,70%
Fiebre	2	2,90%	11,59%
Fractura	3	4,35%	15,94%
Gastritis	1	1,45%	17,39%
Gripe y fiebre	2	2,90%	20,29%
Infección	1	1,45%	21,74%
Ninguna	48	69,57%	91,30%
Paño	1	1,45%	92,75%
Parásitos	2	2,90%	95,65%
Resfriado	1	1,45%	97,10%
Tos y fiebre	1	1,45%	98,55%
Tuberculosis	1	1,45%	100,00%
TOTAL	69	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Gráfico 5. Tipo de enfermedad que tuvo el enfermo según observaciones y porcentajes



Fuente: Encuesta aplicada socio-económica

- Tenencia de la vivienda

La mayoría de hogares del área del proyecto tienen vivienda propia (84 de cada 100), seguido de arrendada con un 8,7% y las viviendas prestadas por servicios llega al 7,2%.

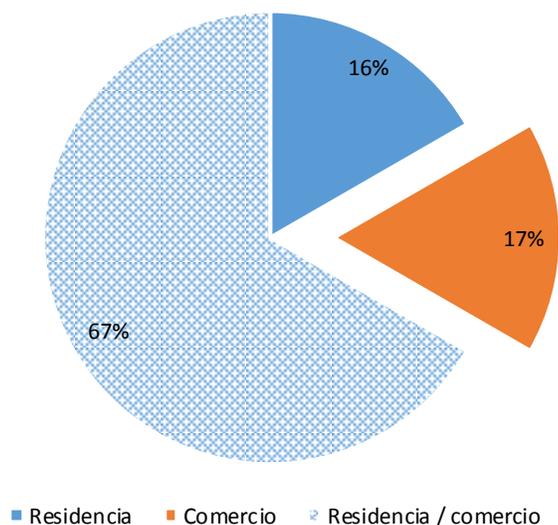
Tabla 6. Usos de la vivienda según observaciones y porcentajes

Uso de la Vivienda	Frecuencias	Porcentaje	% acumulado
Residencia	64	92,75%	92,75%
Comercio	4	5,80%	98,55%
Residencia / comercio	1	1,45%	100,00%
TOTAL	69	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Gráfico 6. Usos de la vivienda según observaciones y porcentajes



Fuente: Encuesta aplicada socio-económica

En el área del proyecto, aproximadamente 92,8 de cada 100 las viviendas son de uso residencial, 5,8 de cada 100 son utilizadas para el comercio. Un problema genérico, en el área rural, en la vivienda no se deriva tanto del tamaño de la vivienda y su tenencia, pero sí, de las múltiples actividades que se realizan en ella y el reducido espacio destinado para dormitorios.

- Servicios intra-domiciliarios

Las condiciones de vivienda y de saneamiento ambiental definen, en gran medida, la forma de vida de la población. La vivienda influye sobre la satisfacción de otras necesidades básicas como salud y educación. De la calidad de la vivienda depende, a su vez, la capacidad para proteger a los habitantes de agentes externos, brindarles seguridad y privacidad y controlar sus riesgos sanitarios.

En este sentido, para este análisis se ha diferenciado los servicios ubicados fuera de la vivienda y aquellos en su interior, considerando que éstos últimos son exclusivos y dan un mayor grado de satisfacción a los hogares.

Tabla 7. Servicios que tiene la vivienda según observaciones y porcentajes

Uso de la Vivienda	Frecuencias	Porcentaje	% acumulado
No sabe	3	4,35%	4,35%
Agua potable	4	5,80%	10,14%
Alcantarillado	1	1,45%	11,59%
Energía eléctrica	39	56,52%	68,12%
Agua potable / energía eléctrica	7	10,14%	78,26%
Agua potable / otros	1	1,45%	79,71%
Alcantarillado / energía eléctrica	1	1,45%	81,16%
Alcantarillado / otros	1	1,45%	82,61%
Agua potable / alcantarillado / energía eléctrica	4	5,80%	88,41%
Agua potable / alcantarillado / otros	1	1,45%	89,86%
Agua potable / energía eléctrica / otros	1	1,45%	91,30%
Agua potable / alcantarillado / energía eléctrica / otros	3	4,35%	95,65%
Agua potable / teléfono / energía eléctrica / otros	2	2,90%	98,55%
Agua potable / teléfono / energía eléctrica / barrido	1	1,45%	100,00%
TOTAL	69	100,00%	

Fuente: Encuesta socio-económicas

Elaborado por: Autores

Como se puede observar el servicio de energía eléctrica es el de mayor cobertura, en efecto más de 56,5% de los hogares cuenta con este servicio. De otra parte, el servicio con menor cobertura es el teléfono. Otro servicio con baja cobertura es el alcantarillado (no sobre pasa de 1,4%), y; consecuentemente contar con servicio higiénico al interior. Por otra parte, 5.8% hogares cuenta con agua al interior de la vivienda (si bien es cierto la mayoría de viviendas cuenta con agua, pero es entubada, en el mejor de los casos).

El servicio de alcantarillado sanitario es el medio masivo más efectivo para la eliminación de excretas y aguas servidas. El porcentaje de hogares con este servicio es un indicador que

refleja, por un lado, la calidad la vivienda y, por otro, el acceso de la población a un servicio urbano básico: la red pública de alcantarillado (1,4%), el mismo, que además debe ser condición para emprender un proyecto de mejoramiento vial. En este contexto, se tiene que los hogares ubicados en las localidades aledañas no tienen este servicio, en tanto que en Carmen de Pijilí la cobertura es muy baja, es decir si están mal servidos.

- Actividades económicas en la vivienda

Las actividades económicas al interior de la vivienda o microempresas son propias del área urbana, sin embargo, en el área rural también se desarrollan algunos negocios relacionados con la actividad agropecuaria, artesanal y tiendas al por menor. En efecto, en el área del proyecto aproximadamente el 20% de los hogares tienen negocios al interior de la vivienda, su actividad está relacionada con la agricultura, ganadería pues, procesan la leche para obtener quesillos y la crianza de cuyes, pollos, gallinas o chanchos.

- Análisis multivariantes de datos

“El análisis multivariante (Cuadras, 2007) es la rama de la estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico sobre un conjunto de $n > 1$ de variables, que pueden ser cuantitativas, cualitativas o una mezcla”. El análisis de datos multivariante se puede dividir en dos partes, la primera extrae información que contiene datos disponibles y segundo obtiene conclusiones sobre la población para construir un modelo que permite prever valores futuros.

Dentro de la localidad del Carmen de Pijilí existen determinadas variables socio - económicas que definen a la población, esta variable puede ser: El nivel de ingreso que determina el gasto del hogar, el nivel de Educación, la Ocupación, la Salubridad del hogar, etc. Dentro del análisis multivariante se ha realizado un estudio de dos de estas variables (por supuesto que es aplicable a todas las variables estudiadas) y son: El Gasto mensual del hogar a través de un análisis de componentes principales complementando con un análisis de Clasificación a través del método K.N.N., y la Salubridad, para analizar esta variable se ha definido un modelo lineal de comportamiento en donde las variables independientes son los diferentes servicios con los que dispone la vivienda, y la variable dependiente se ha definido a la variable del hogar que tiene miembros enfermos.

- Análisis del gasto en la vivienda

Una de las variables que son importantes analizar es el Ingreso familiar porque esto determina en gran medida el Gasto del hogar. En Ecuador el mayor porcentaje del Ingreso se los destina

al Gasto del Hogar, quedando únicamente alrededor de un 5% del Ingreso destinado a otras actividades, como son al pago de deudas o en inversión.

Por otra parte, existe siempre un comportamiento estratégico de los hogares cuando se les consulta sobre sus ingresos a no dar exactamente estos valores, debido al temor de que se les puede cobrar más impuestos, pero cuando se les consulta sobre los Gastos del Hogar, los encuestados si proporcionan de manera fidedigna estos valores.

Para el análisis de componentes principales sobre los Gastos del Hogar, partimos de una hipótesis que se explica de la siguiente manera: Existe una priorización del Gasto que será distribuidos en los siguientes rubros: Alimentación, Vestuario, Vivienda, Educación, Electricidad, Salud, Agua, Transporte, Teléfono y Otros-Gastos, la priorización del Gasto dependerá de las necesidades que el Hogar considera indispensables.

Tabla 8. Componentes principales (gasto familiar)

Tipo de gasto	Orden	Valor
Alimentación	1	2,454
Vestuario	2	1,507
Vivienda	3	1,248
Educación	4	1,130
Electricidad	5	0,983
Salud	6	0,819
Agua	7	0,671
Transporte	8	0,612
Teléfono	9	0,369
Otros	10	0,207

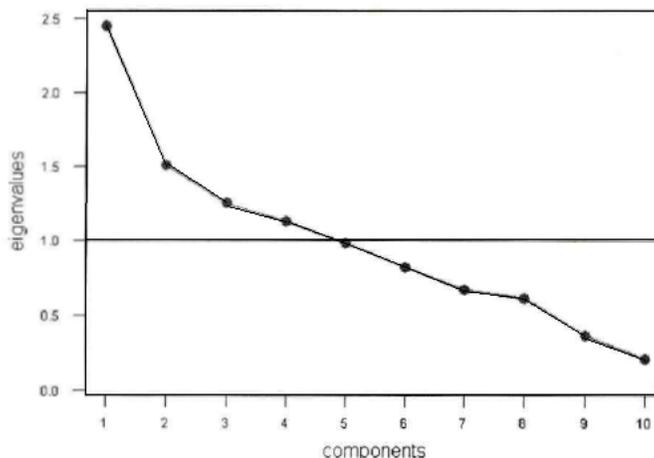
Fuente: Encuesta socio-económicas

Elaborado por: Autores

Utilizando en programa SCAN para el análisis del gasto del hogar, a través del método K.N.N. se ha obtenido los siguientes resultados: Alimentación (2,4), Vestuario (1,5), Vivienda (1,2), Educación (1,1), Electricidad (0,9) son las variables que explican en su mayoría al Gasto del Hogar, es decir, si analizamos únicamente estas 5 variables como componentes del Gasto del Hogar, se estaría estudiando a esta variable en un gran porcentaje.

Las otras variables del Gasto del Hogar, como: Salud (0.819), Agua (0.671), Transporte (0.612), Teléfono (0.369) y Otros-Gastos (0.207) no son representativos o no son determinantes en el comportamiento del Gasto del Hogar.

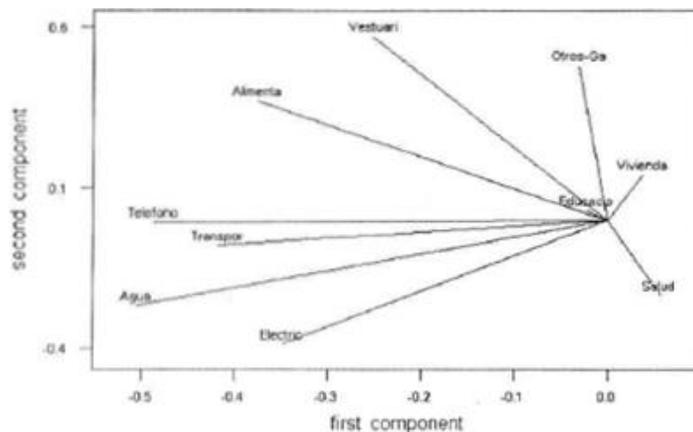
Gráfico 7. Principal componente eigenvalue



Fuente: Encuesta aplicada socio-económica

Para determinar que variables son determinantes, en el Gasto Familiar, se ha utilizado el siguiente criterio: De acuerdo al gráfico (scree plot), en el eje X están las variables o componentes definidos con los números siguientes: Alimentación (1), Vestuario (2), Vivienda (3), Educación (4), Electricidad (5), Salud (6), Agua (7), Transporte (8), Teléfono (9) y Otros-Gastos (10); y de acuerdo a los valores obtenidos en el eje y (eigenvalues), aquellas variables que superan de 1 o más se definen como los componentes principales.

Gráfico 8. Principal componente loading plot



Fuente: Encuesta aplicada socio-económica

Las tres primeras variables definidas como componentes principales son, en su orden: Alimentación (2,4), Vestuario (1,5) y Vivienda (1,2). Se puede definir, estadísticamente, que las variables Educación (1,1) y Electricidad (0,9) están en una cuarta y quinta ubicación de elección en los consumos familiares.

B. Resultados obtenidos al aplicar análisis multivariante.

El análisis estadístico multivariante, aplicado al conjunto de datos de la Encuesta socio – económica permite detectar la presencia de estructuras (grupos, categorías, clases) propias de la información en su conjunto. Estas estructuras no serían detectadas utilizando métodos clásicos mono o bivariante.

La detección de grupos o clases permite comprender mejor el fenómeno representado por la matriz de datos. Para los fines del presente informe se atribuye a los diferentes ingresos, el rol de "variables" y a las familias que corresponden los datos, el papel de "objetos".

La metodología utilizada se desarrolla por fases, las mismas que son las siguientes: Primero se hizo una Normalización de los datos, en este caso se desarrolló mediante el uso de un procedimiento de "Estandarización" o "Normalización" con lo cual es posible expresar las diferentes variables en las mismas unidades de medida. Luego se realiza una Exploración, para detectar indicios de estructuras intrínsecas de los datos, esta fase se ha llevado a cabo mediante técnicas de análisis de agrupamiento (Cluster Analysis) y finalmente se practicó una Validación de la consistencia de los grupos o categorías detectados mediante agrupamiento por medio de técnicas de clasificación, utilizando el algoritmo del "Vecino más cercano" o K.N.N. (K-Nearest Neighbor Method).

Se busca un modelo determinístico entre las enfermedades y Dominio de Vivienda, Enfermedades, Utilización de cultivos con Químicos, Eliminación de Basura, Tenencia de la Vivienda, Toma el Agua, Distancia de los Cultivos, Eliminación de Aguas Servidas, Estado de las Vías, Ingreso del Hogar, Evacuación de Aguas Negras, Servicios de la Vivienda y Estatus del Hogar.

En el caso de la Normalización de los datos, se procedió así: los valores de cada variable fueron normalizados a través de la siguiente transformación:

$$X'_i = \frac{x_i - \bar{x}}{S}$$

En donde:

x_i : Valor observado.

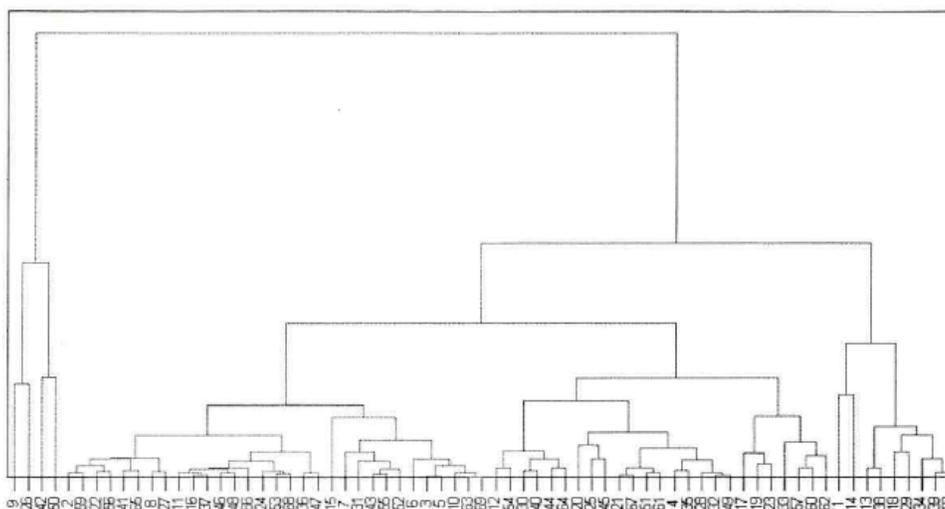
x : Valor promedio de las variables observado.

S : Desviación estándar de la variable.

X'_i : Valor estandarizado.

Por su parte, el agrupamiento (*clustering*) es del siguiente tipo. Se ha utilizado en paquete estadístico SCAN. El dendograma final se presenta en la figura siguiente:

Gráfico 9. Cluster análisis



Fuente: Criterio de selección del modelo

El primer grupo son aquellas familias cuyos ingresos oscilan entre 9 a 50 USD por mes. El segundo grupo va desde los 2 hasta los 62 USD por mes, y el tercer grupo va desde 1 hasta 58 USD por mes.

Lo relacionado con la clasificación. El algoritmo utilizado, K.N.N. se le que atribuye un objeto a una u otra clase según que la mayoría de sus K vecinos pertenezcan a una clase. En este caso se ha considerado un número K de vecinos igual a 3 y se ha trabajado con simple mayoría. Este algoritmo es muy robusto, es decir, no depende del tipo de función de distribución de cada clase.

Cuando se levantó la información a través de la encuesta socio - económica, dentro de las diferentes variables estadísticas a ser levantadas existe una variable definida como Estatus del Hogar esta variable define o categoriza a los hogares en tres tipos de clases: Alto = (1), Medio = (2), y Bajo = (3), dependiendo del estado de la vivienda, según los siguientes criterios:

- Estatus alto: Casas de Paredes de Ladrillo, columnas de fundición, espacios verdes dentro del predio.
- Estatus medio: Casas de Paredes de Ladrillo o Media Agua, sin espacios verdes en el predio.
- Estatus bajo: Casas de Adobe, Caña, etc., de un piso, techos de paja u otro material.

Utilizando el algoritmo K.N.N., de Clasificación se quiere validar esta clasificación inicial realizada por el encuestador, por lo cual se ha utilizado la variable Ingreso y la variable Estatus, para realizar una nueva clasificación a través del algoritmo K.N.N., el resultado de la clasificación se resume en la siguiente matriz:

Tabla 9. Variable Ingreso y la variable Estatus

Clases asignadas	Total	1	2	3
1	3	0	2	1
	4,35%	0,00%	66,67%	33,33%
2	47	2	43	2
	68,12%	4,26%	91,49%	4,26%
3	19	0	14	5
	27,54%	0,00%	73,68%	26,32%
Cross-validated Error Rate:	30,43%			
Cross-validated Risk	30,43%			
Observaciones:	69			

Fuente: Encuesta socio-económicas

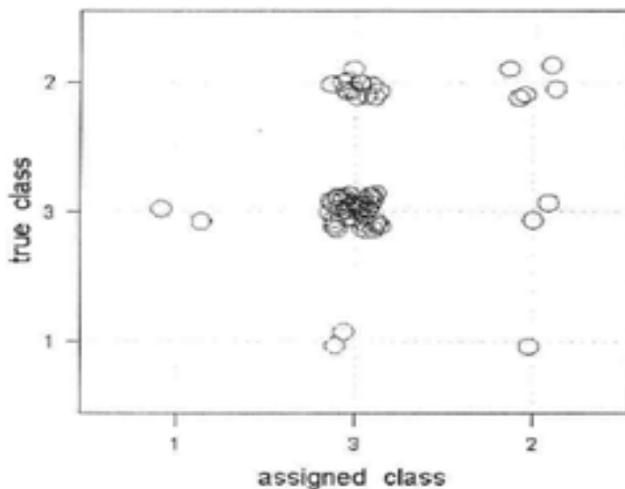
Elaborado por: Autores

La clasificación a través de la variable Estatus definió el siguiente resultado: Pertenecen al Estatus Alto (1) un 4,35% de la población, pertenecen al *Estatus Bajo* (3) un 68,12% de la población estudiada y pertenecen a un *Estatus Medio* un 27,54% de la población.

Utilizando el algoritmo K.N.N., de Clasificación y utilizando la variable Ingreso del Hogar se ha tenido el siguiente resultado: El *Cross-validated Error Rate* es de 30,43%, es decir que el modelo ha clasificado y ha coincidido en un 69,57% con la variable Estatus del Hogar, en un análisis individual se ha obtenido los siguientes resultados:

- El 66,7% y 33,3% se ha clasificado en la categoría 2 y 3, respectivamente, siendo de la categoría 1.
- El 2,3% y 91,5% se ha clasificado en la categoría 1 y 2, respectivamente, siendo de la categoría 3.
- El 26,3% se ha clasificado en la categoría 3, siendo de la categoría 2.

Gráfico 10. KNN validate class assignments

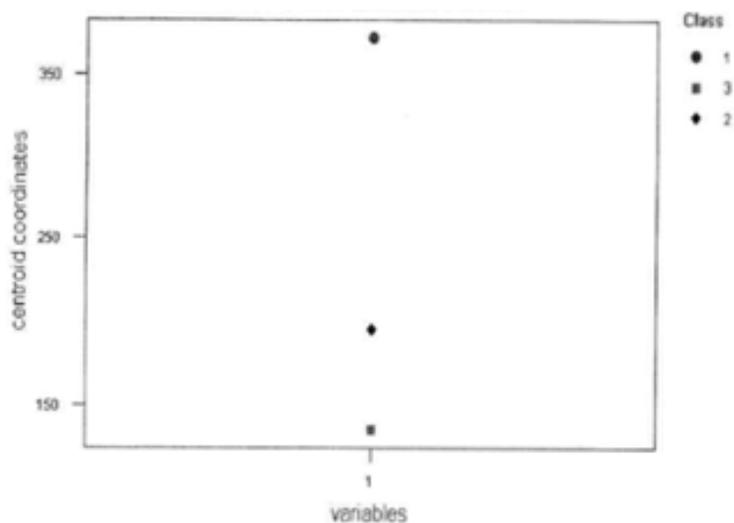


Fuente: Criterio de selección del modelo

Existe alrededor de un 70% de coincidencia en la clasificación a través de la variable *Estatus* con la variable *Ingresos*, por lo que se puede definir que existe una correspondencia entre el *Tipo de Vivienda* y los *Ingresos del Hogar*.

Las diferencias entre los valores de los *centroides* de las tres categorías se presentan en la siguiente figura:

Gráfico 11. KNN class centroid profile



Fuente: Criterio de selección del modelo

Centroide	1	372,67
	2	195,21
	3	134,53

El primer centroide es de 372,67 USD, el segundo *centroide* es de 195,21 USD. El tercer *centroide* es de 134,53 USD, de acuerdo a ésta clasificación las familias que pertenecen al Estatus Alto tendrían un ingreso que en promedio es de 372,67 USD al mes, los de Estatus medio tendrían un ingreso promedio mes de 195,21 USD, y los de Estatus Bajo tendrían un ingreso promedio de 134,53 USD.

La clasificación a través de la variable Ingreso definió el siguiente resultado: Pertenecen al *Estatus Alto* (1) un 4,35% de la población, pertenecen al *Estatus Bajo* (3) un 38,12% de la población estudiada y pertenecen a un *Estatus Medio* un 31,54% de la población.

Gráfico 12. KNN model selection



Fuente: Criterio de selección del modelo

En cuanto a la modelación. Conjuntamente, se elaboró un modelo de regresión lineal entre las Enfermedades del Hogar y las siguientes variables: *Dominio de Vivienda, Enfermedades, Utilización de cultivos con Químicos, Eliminación de Basura, Tenencia de la Vivienda, Toma el Agua, Distancia de los Cultivos, Eliminación de Aguas Servidas, Estado de las Vías, Ingreso del Hogar, Evacuación de Aguas Negras, Servicios de la Vivienda y Estatus del Hogar.*

Para realizar esta simulación se utilizó el Programa SPSS, los resultados se presentan en los siguientes cuadros:

Tabla 10. Resumen del modelo (b)

Modelo	R	R cuadrado	R cuadrado corregido	Error tip. De la estimación	Durbin-Watson
1	0,515(a)	0,265	0,091	0,538	1,799

a Variables predictoras: (Constante), DOMVIV, ENFERM, EDADES, QUIMCULT, ELIMBASURA, TENENVIV, TOMAGUA, DISTCULT, ELLMAS, ESTVIAS, INGRESO, EVACAGUA, SERWIV, ESTATUS.

b Variable dependiente: ENFERMOS

El ajuste del modelo no es consistente pues el R^2 ajustado es del 26,5%, es decir que existe una inconsistencia del 73,5% de error o que el modelo no explica.

Tabla 11. Coeficientes (a)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error tip.	Beta		
1	(Constante)	0,122	0,847		0,144	0,886
	Enfermedades	0,161	0,152	0,137	1,059	0,293
	QUIMCULT	-0,579	0,426	-0,211	-1,359	0,18
	DISTCULT	0,000	0,000	0,062	0,489	0,627
	INGRESO	0,000	0,000	0,115	0,853	0,397
	ESTATUS	-0,042	0,170	-0,040	-0,247	0,803
	TENENVIV	-0,263	0,114	-0,287	-2,307	0,025
	ELIMAS	0,171	0,054	0,472	3,167	0,003
	ElimBasura	-0,002	0,003	-0,108	-0,667	0,455
	EVACAGUA	-0,097	0,059	-0,229	-1,644	0,104
	ESTVIAS	-0,065	0,119	-0,720	-0,546	0,588
	TOMAGUA	0,018	0,042	0,060	0,429	0,662
	SERWIV	0,000	0,000	0,126	0,827	0,412
DOMVIV	0,149	0,330	0,054	0,452	0,653	

a Variable dependiente: enfermos.

El análisis de significancia individual, a través de la t estadística, indica que solamente las variables *Tenencia de la Vivienda* (-2.3) y *Eliminación de Agua Servidas* (3.17) son las variables de consistencia estadística a un nivel de significancia del 95%, La relación es inversa entre *Enfermos en el Hogar* y la variable *Tenencia de la Vivienda* y la relación es directa entre la variable *Enfermos* y *Eliminación de Aguas Servidas*. El resto de variables: *Dominio de Vivienda*, *Enfermedades*, *Utilización de cultivos con Químicos*, *Eliminación de Basura*, *Toma el Agua*, *Distancia de los Cultivos*, *Estado de las Vías*, *Ingreso del Hogar*, *Evacuación de Aguas Negras*, *Servicios de la Vivienda* y *Estatus del Hogar* no son significativos individualmente a un 95% de significancia de acuerdo al estadístico t.

El Modelo final. Del modelo inicial, únicamente 2 variables fueron significativas (*t estat*) al 95%; y son las variables *Tenencia de la Vivienda* y *Eliminación de Agua Servidas*, entonces; se realizó un nuevo modelo que explica la variable *Enfermos en el Hogar* (variable dependiente) con las variables *Tenencia de la Vivienda* y *Eliminación de Aguas Servidas* (variables independientes), obteniendo los siguientes resultados:

Tabla 12. Resumen del modelo (b)

Modelo	R	R cuadrado	R cuadrado corregido	Error tip. De la estimación	Durbin-Watson
1	0,905(a)	0,864	0,801	0,524	1,886

a Variables predictoras: (Constante), Ellmas, Tenenviv

b Variable dependiente: Enfermos.

El ajuste del modelo es consistente pues el R^2 ajustado es del 86,4%, es decir que existe una inconsistencia del 13,6% de error o que el modelo no explica.

Tabla 13. Coeficientes (b)

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error tip.	Beta		
1	(Constante)	0,223	0,172		1,998	0,199
	TENENVIV	-0,232	0,103	-0,253	-2,243	0,28
	ELIMAS	0,122	0,410	0,336	2,978	0,004

a Variable dependiente: Enfermos.

El análisis de significancia individual, a través de la *t* estadístico, indica que las variables *Tenencia de la Vivienda* (-2.2) y *Eliminación de Agua Servidas* (2,9) son las variables con una consistencia estadística, a un nivel de significancia del 95%. La relación es inversa entre *Enfermos en el Hogar* y la variable *Tenencia de la Vivienda*. La relación es directa entre la variable *Enfermos en el Hogar* y *Eliminación de Aguas Servidas*.

Tabla 14. Correlaciones de los coeficientes(a)

Modelo		ELIMAS	TENENVIV
1	Correlaciones	ELIMAS	1,000
		TENENVIV	-0,740
	Covarianza	ELIMAS	0,002
		TENENVIV	0,011

a Variable dependiente: Enfermos

La ecuación final que explica la variable *Enfermos del Hogar* queda definida de la siguiente manera:

$$\text{EnfermosHogar} = 0.223 - 0.232 * \text{TenenciaVivienda} + 0.122 * \text{EliminaciónSanitaria}$$

Con este modelo, *calibrado*; se puede definir la probabilidad de un *Hogar* de tener enfermos, a través; de la *Tenencia de la Vivienda*, además; si la Vivienda tiene un sistema fiable de evacuación de *Aguas Servidas*, puesto que incide en la salubridad de los miembros del hogar, según los resultados obtenidos.

Conclusiones.

- El análisis multivariante es, en esencia, un conjunto de técnicas orientadas al estudio de proyecciones de varias variables de modo simultáneo, el método puede generar una explicación más rica y cercana a la realidad, a lo que se deriva, a su vez, en un mayor conocimiento del fenómeno, con la utilización de matrices y, por lo tanto, en un lugar apropiado para la Selección optima de la decisión o respuesta de interés.
- De acuerdo con los resultados obtenidos en la investigación de campo de la parroquia Carmen de Pijilí y sus zonas aledañas. Utilizando el análisis multivariante, se pudo determinar que el hacimiento, el aislamiento y la falta de saneamiento ambiental influyen negativamente en el desarrollo de las poblaciones y en la calidad de vida de sus habitantes. Bajo esta consideración un número significativo de familias que habita en la parroquia y las zonas aledañas en viviendas presentan factores de riesgos para la salud y un detonante de grandes enfermedades por la situación precaria e insalubridad de evacuación de las aguas servidas. Este problema puede también ser ocasionados por otros factores que no son tan influyentes individualmente como el Dominio de Vivienda, Enfermedades, Utilización de cultivos con Químicos, Eliminación de Basura, Toma el Agua, Distancia de los Cultivos, Estado de las Vías, Ingreso del Hogar, Evacuación de Aguas Negras, Servicios de la Vivienda y Estatus del Hogar

Referencias bibliográficas.

- Cuadras, C. M. (2007). Nuevos Metodos de Analisis Multivariante. *Revista española de quimioterapia : publicación oficial de la Sociedad Española de Quimioterapia*, 20(3), 249. Recuperado de <http://www.ncbi.nlm.nih.gov/pubmed/19406528>
- Informe sobre el crecimiento sostenido y el desarrollo incluyente.* (s. f.).
- León González, Á., Llinás Solano, H., & Tilano, J. (2011). Ingeniería y Desarrollo. En *Revista Científica Ingeniería y Desarrollo* (Vol. 23). Recuperado de <http://rcientificas.uninorte.edu.co/index.php/ingenieria/article/viewArticle/2098/4467>
- Lozares Colina, C., & López Roldán, P. (1991). El análisis multivariado: definición,

criterios y clasificación. *Papers. Revista de Sociologia*, 37(37), 9.
<https://doi.org/10.5565/rev/papers/v37n0.1594>

Martínez Arias, R. (2008). *El análisis multivariante en la investigación científica*.
Recuperado de <https://www.casadellibro.com/libro-el-analisis-multivariante-en-la-investigacion-cientifica/9788471337788/2262741>

Sánchez, R. J., Lardé, J., Chauvet, P., & Jaimurzina, A. (2017). Inversiones en infraestructura en América Latina Tendencias , brechas y oportunidades. *CEPAL. Serie Recursos Naturales e Infraestructura*, (187), 1-94

PARA CITAR EL ARTÍCULO INDEXADO.

Zamora Zamora, G., Banegas Peña, T., Flores Flores, M., & Becerra Molina, J. (2019). Análisis multivariante para el estudio socio-económico - parroquia Carmen de Pijili. *Ciencia Digital*, 3(2.3), 66-91. <https://doi.org/10.33262/cienciadigital.v3i2.3.489>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Ciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Ciencia Digital**.

