

MODELO PREDICTIVO DE RESIDUOS ORGÁNICOS DOMICILIARIOS PER CÁPITA SEGÚN VARIABLES SOCIOECONÓMICAS Y DEMOGRÁFICAS EN POBLACIONES CON MENOS DE 70.000 HABITANTES

PREDICTIVE MODEL OF PER CAPITA HOUSEHOLD ORGANIC WASTE BASED ON SOCIOECONOMIC AND DEMOGRAPHIC VARIABLES IN POPULATIONS UNDER 70,000 INHABITANTS

 ^{1,2} María Gabriela Arias Garnica*

 ^{1,2} Pablo Javier Flores Muñoz

¹ Universidad Politécnica Estatal del Carchi (UPEC), Posgrado, Av. Universitaria y Antisana, Tulcán, 040101, Ecuador.

² Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Ciencias, Riobamba, Ecuador.

*maria.arias@upec.edu.ec

RESUMEN

En localidades menores a 70.000 habitantes, la gestión de residuos suele basarse en estimaciones teóricas poco exactas. Para corregir esto, el estudio desarrolló un modelo para predecir la cantidad de residuos orgánicos generados por persona, analizando factores socioeconómicos y demográficos de 15 pueblos en Ecuador mediante el monitoreo de 1.195 hogares durante siete días. El análisis empleó técnicas estadísticas avanzadas (reducción de dimensiones y un Modelo Lineal Generalizado tipo Gamma). Estas herramientas superaron a los métodos tradicionales al adaptarse mejor a la irregularidad real de los datos. Los resultados indican que el nivel de ingresos es el factor de mayor influencia, mostrando que a más ingresos, mayor generación general de residuos. Sin embargo, en familias numerosas, la cantidad de basura por persona disminuye. Un hallazgo clave demuestra que los sectores de menores recursos generan más residuos orgánicos per cápita que los sectores altos, debido al consumo frecuente de alimentos frescos. Se concluye que aplicar modelos estadísticos precisos es fundamental para diseñar sistemas de recolección eficientes y evitar errores en poblaciones con realidades socioeconómicas variadas.

Palabras claves: *Residuos orgánicos domiciliarios; Regresión robusta; Modelado predictivo; Factores socioeconómicos; Gestión municipal de residuos; Ecuador.*

ABSTRACT

Waste management planning in small settlements lacks locally calibrated data. This study develops and validates a predictive model for per capita household organic waste production in 15 Ecuadorian settlements under 70,000 inhabitants. A sample of 1,195 households was analysed. Scaled Principal Component Analysis (PCA) reduced dimensionality. A Generalised Linear Model (GLM) with Gamma distribution and log-link outperformed ordinary least squares methods. Household income was the strongest positive predictor ($z = 35.50$). Household size showed a negative relationship, confirming domestic economies of scale. Low socio-economic strata generated more organic waste per capita than higher strata. Gamma-family models provide reliable tools for waste collection system design in heterogeneous, data-scarce contexts.

Keywords: *Household solid waste; Robust regression; Predictive modelling; Socio-economic factors; Organic waste; Waste management.*

I. INTRODUCCIÓN

La gestión integral de residuos sólidos urbanos constituye uno de los desafíos ambientales y sanitarios más apremiantes del siglo XXI. El crecimiento demográfico acelerado, sumado a patrones de consumo lineales y una urbanización no planificada, ha exacerbado la generación de desechos a nivel global (1). Según estimaciones recientes del Banco Mundial, la generación mundial de residuos aumentará un 70% para el año 2050 si no se implementan medidas urgentes, alcanzando los 3.400 millones de toneladas anuales (2). Esta situación es especialmente grave en América Latina, ya que la cantidad de basura que genera cada persona aumenta mucho más rápido de lo que los municipios pueden recoger y procesar. Esto provoca la aparición de basureros al aire libre, lo que afecta seriamente la salud de la población (3, 4).

Por lo tanto, para organizar de forma eficiente el manejo de estos desechos, es indispensable contar con información exacta y de confianza. Como establece la máxima de gestión: no se puede gestionar lo que no se mide. Sin embargo, la cuantificación de la Producción Per Cápita (PPC) de residuos sólidos domiciliarios (RSD) presenta una variabilidad estocástica significativa, influenciada por una compleja red de factores socioeconómicos, demográficos, culturales y estacionales (5). En las grandes metrópolis, existen recursos para monitorear estas variables mediante pesajes continuos en básculas de vertedero; sin embargo, esta realidad difiere drásticamente en ciudades intermedias y pequeñas (6).

En el contexto específico de poblados menores a 70.000 habitantes, que constituyen una gran parte de la estructura demográfica de países como Ecuador, la carencia de registros históricos y la limitación presupuestaria impiden la caracterización continua de los residuos (7). En Ecuador, estudios sobre la gestión de residuos sólidos municipales han documentado que la mayoría de los gobiernos autónomos descentralizados de escala pequeña e intermedia carecen de datos primarios de generación per cápita, recurriendo a estimaciones nacionales que no reflejan la heterogeneidad territorial del país (8, 9). Por lo general, estos municipios organizan sus sistemas de recolección de basura utilizando promedios nacionales o cálculos teóricos que no se ajustan a su propia realidad (10). Esta

forma de planificar provoca errores comunes: diseñar rutas de recolección insuficientes, lo que causa problemas de salud, o construir basureros demasiado grandes, lo que termina desperdiciando el dinero público (11).

La literatura científica contemporánea ha evidenciado que la generación de RSD no es un fenómeno aleatorio puro, sino que está fuertemente correlacionada con indicadores socioeconómicos. Diversos estudios han validado la existencia de una relación directa entre el nivel de ingresos económicos y la tasa de generación de residuos, fenómeno a menudo asociado a la Curva Ambiental de Kuznets (12, 13). A mayor poder adquisitivo, se observa un incremento en el consumo de productos procesados y, consecuentemente, en la generación de materiales inorgánicos y embalajes (14). En cambio, en los sectores de menores recursos y zonas rurales, la mayor parte de la basura es orgánica y se produce menos cantidad en total. Allí, los desechos suelen manejarse de otras maneras dentro del hogar, como quemándolos o usándolos para alimentar animales (15). Para poder aprovechar estos recursos (Economía Circular), es indispensable conocer con exactitud qué tipo de basura se genera, especialmente la orgánica. Esto ayuda a planificar proyectos útiles, como crear abono o biogás, y depender menos de los basureros tradicionales (16).

Además de los ingresos, la cantidad de personas en una casa es un factor fundamental, ya que se genera una especie de ahorro al compartir recursos (17). Los estudios recientes demuestran que la basura no aumenta en la misma proporción por cada persona extra en el hogar. De hecho, las familias numerosas aprovechan mejor las cosas y producen menos desechos por individuo en comparación con las personas que viven solas (18, 19). Ignorar esta dinámica no lineal al proyectar la generación de residuos en poblaciones en crecimiento conduce a predicciones sesgadas y políticas públicas ineficaces (20).

A pesar de la abundante literatura sobre modelos predictivos de residuos sólidos, existe un sesgo marcado hacia el estudio de grandes urbes o naciones desarrolladas, utilizando técnicas como redes neuronales artificiales o análisis de series temporales que requieren grandes volúmenes de datos históricos (21-23). En contraste, la

aplicación de modelos estadísticos adecuados en ciudades pequeñas de países en vías de desarrollo es escasa (24). Los métodos convencionales de regresión lineal múltiple (OLS) a menudo fallan en este contexto debido a la presencia de valores atípicos severos (outliers) y la naturaleza asimétrica de los datos de residuos, lo que viola los supuestos de normalidad y estabilidad requeridos para una estimación fiable (25, 26). Una revisión sistemática de (27) confirma que los modelos de regresión con variables socioeconómicas son los más frecuentes en la literatura, pero rara vez han sido validados en poblados de escala pequeña con datos primarios locales.

Ante esta brecha de conocimiento, surge la necesidad de implementar técnicas de Regresión Robusta, específicamente aquellas basadas en M-estimadores, que han demostrado ser superiores para modelar variables donde la presencia de observaciones influyentes distorsiona los estimadores de mínimos cuadrados ordinarios (28). Estas herramientas permiten atenuar el efecto de los valores extremos y capturar la tendencia central real de la generación de residuos sin la pérdida masiva de información que implicaría la eliminación arbitraria de datos (29).

El presente estudio tiene como objetivo desarrollar y validar un modelo predictivo de la producción per cápita de residuos orgánicos domiciliarios en 15 poblados del Ecuador, distribuidos en seis provincias: Chimborazo (Penipe, San Andrés, Yaruquíes, Chambo, Calpi, Guamote, Licán y Guano), Tungurahua (Cevallos y Patate), Bolívar (Guaranda y Las Naves), Cotopaxi (Salcedo), Imbabura (Otavalo) y Galápagos (Puerto Ayora), todos con poblaciones menores a 70.000 habitantes. Para predecir estos valores, se analizaron datos como el tamaño de la familia, sus ingresos y su nivel social.

La información proviene de varios estudios independientes realizados entre 2020 y 2025, que luego se agruparon en un solo archivo. En total, se estudiaron 1.195 hogares durante siete días seguidos. El trabajo consistió en entregar recipientes marcados, recoger y pesar la basura diariamente, y separarla en distintos grupos (orgánicos, reciclables, desechos tecnológicos, peligrosos, etc.). Las casas se eligieron al azar, pero divididas por niveles socioeconómicos (A, B, C y D), considerando un margen de ajuste por si alguna familia decidía no participar.

Se estableció el límite de 70.000 habitantes por tres razones principales: (i) el método utilizado está comprobado y funciona bien en poblaciones de hasta 150.000 personas (30); (ii) el pueblo más grande analizado en el estudio está dentro de este rango; y (iii) en ciudades con más habitantes, el comercio, el turismo y la cantidad de viviendas mezclan mucho más la basura, lo que obligaría a usar fórmulas matemáticas mucho más complicadas. Cabe señalar que la incorporación de la población rural de cada cantón es metodológicamente coherente, dado que los servicios municipales de recolección en Ecuador tienen cobertura cantonal; bajo este criterio, localidades como Otavalo —con población cantonal total próxima a los 70.000 habitantes— quedan dentro del ámbito del modelo.

II. MATERIALES Y MÉTODOS

La presente investigación adoptó un enfoque cuantitativo de alcance correlacional y explicativo, orientado a modelar la producción de residuos sólidos domiciliarios en función de determinantes socioeconómicos y demográficos.

2.1. Análisis exploratorio y estadística descriptiva

El análisis inicial de los datos se orientó a caracterizar la distribución de las variables y evaluar la calidad de la información, se calcularon medidas de tendencia central (media, mediana) y dispersión (desviación estándar, rango, coeficiente de variación) para las variables de producción per cápita PPC y los determinantes socioeconómicos y demográficos (*Ingreso Familiar, Tamaño del Hogar y estrato social*).

Además, se analizó la forma en que se agrupan los datos utilizando medidas matemáticas de asimetría (*skewness*) y curtosis, junto con pruebas específicas (Shapiro-Wilk) para asegurar la elección correcta de las fórmulas a utilizar más adelante. Para facilitar la revisión visual, se crearon gráficos especiales (*boxplots*) y gráficos de dispersión, lo que permitió identificar fácilmente los datos que se salían de lo común (*outliers*) y observar cómo se relaciona el nivel de ingresos con la cantidad de basura generada en cada sector social.

2.2. Estrategia de reducción de dimensionalidad y multicolinealidad

Para examinar la estructura de covarianza entre las diferentes fracciones de residuos y reducir la dimensionalidad del conjunto de datos, se aplicó un Análisis de Componentes Principales (PCA). Con el objetivo de capturar la variabilidad de todas las fracciones de forma equitativa, el análisis se ejecutó sobre la matriz de correlación, aplicando centralización ($\text{center} = \text{TRUE}$) y escalado ($\text{scale} = \text{TRUE}$). Esta forma de trabajar asegura que los datos con cantidades más grandes (como los residuos orgánicos) no alteren de forma artificial la creación de los grupos estadísticos (componentes principales). De esta manera, se logra un análisis justo y equilibrado de cómo se genera la basura, evitando errores causados simplemente por el gran tamaño de los números.

Por otro lado, para decidir cuántos de estos grupos estadísticos se debían conservar en el estudio, se utilizó una herramienta visual (gráfico de sedimentación) y se calculó cuánta información útil lograban representar en su conjunto (varianza acumulada explicada). La interpretación de los patrones latentes se realizó mediante la inspección de los vectores propios y la construcción de Biplots, permitiendo visualizar simultáneamente la dispersión de los hogares (scores) y la dirección de las variables originales (cargas).

Es importante también dar a notar que, se evaluó la redundancia entre los predictores mediante el cálculo del Factor de Inflación de la Varianza (VIF). Dada la correlación teórica esperada entre el ingreso familiar y el estrato social, se realizó un análisis comparativo entre el modelo saturado y modelos simplificados (omitiendo variables redundantes) utilizando el Criterio de Información de Akaike (AIC). Esta aproximación permitió justificar la selección final de variables, asegurando un equilibrio entre la parsimonia del modelo y su robustez predictiva.

2.3. Estrategia de modelado predictivo

Basado en los hallazgos del Análisis de Componentes Principales (PCA) escalado, en el cual la primera componente explicó el 60.8% de la variabilidad total, representando una escala de generación global en la que todas las fracciones aportan positivamente a

la varianza, se seleccionó a la Producción Per Cápita de Residuos Sólidos Orgánicos como la variable dependiente principal para el modelado. Inicialmente, se ajustó un modelo de Regresión Lineal Múltiple por Mínimos Cuadrados Ordinarios (OLS), incorporando como predictores exógenos al Tamaño del Hogar, Ingreso Familiar, Estrato Social y la Localidad (estas dos últimas codificadas como variables dummy). Para confirmar que este modelo inicial funcionara correctamente, se revisaron a fondo sus posibles errores (residuos). Se verificó si los datos tenían un comportamiento normal mediante herramientas visuales (gráficos Q-Q) y pruebas matemáticas precisas (Shapiro-Wilk y Lilliefors). Además, se detectaron los valores fuera de lo común utilizando gráficos específicos (diagramas de caja) y cálculos especiales (distancias de Cook).

Al notar que los datos no seguían un patrón normal y que la fórmula tradicional (estimador OLS) se alteraba demasiado con los valores extremos, se decidió no borrar información de forma arbitraria. En su lugar, se aplicaron y compararon dos métodos estadísticos más resistentes. Primero, se usó un método que reduce el impacto de los datos exagerados (Regresión Robusta o RLM, empleando la técnica de estimación-M de Huber). Segundo, debido a que la cantidad de basura siempre es un número positivo y los datos se concentraban hacia un lado (asimetría hacia la derecha), se aplicó un modelo avanzado ideal para este tipo de información (Modelo Lineal Generalizado o GLM, asumiendo una distribución Gamma con función de enlace logarítmico). La selección del modelo definitivo se realizó priorizando las métricas de error de predicción (RMSE y MAE) sobre un conjunto de validación independiente, optándose por aquel que minimizara la discrepancia entre los valores observados y predichos para garantizar la robustez del diseño de los sistemas de recolección.

2.4. Esquema de validación del modelo

La comparación entre los modelos candidatos (OLS, RLM y GLM Gamma) se realizó mediante un esquema de partición aleatoria estratificada (*holdout*) en proporción 80/20: el 80% de los hogares ($n \approx 956$) se destinó al entrenamiento y el 20% restante ($n \approx 239$) constituyó el conjunto de validación independiente, sobre el cual se calcularon el RMSE y el MAE reportados. Para

dividir los datos, se aseguró que todos los niveles socioeconómicos estuvieran representados de forma equitativa tanto en la fase de aprendizaje del modelo como en la de prueba (utilizando la función *createDataPartition* del paquete *caret* en R y una semilla aleatoria = 42 para que el proceso pueda repetirse con exactitud). No se utilizó un método de validación por repetición (*k-fold cross-validation*) porque la cantidad de hogares analizados (1.195) es lo suficientemente grande como para hacer una sola separación de la información (partición holdout) y seguir obteniendo resultados totalmente confiables. Además, como el objetivo final es que los municipios apliquen esta herramienta en la vida real, demostrar que la fórmula funciona correctamente en un grupo de prueba independiente y fijo es la mejor evidencia para respaldar las decisiones de trabajo diarias.

2.5. Justificación de la retención de predictores con alta multicolinealidad

El diagnóstico de multicolinealidad reveló $VIF = 10.94$ para el Ingreso Familiar y $VIF = 29.94$ para el Estrato Social. Ambas variables son conceptualmente complementarias y no redundantes: el ingreso captura el flujo monetario corriente (variabilidad intra-estrato), mientras el estrato refleja una posición socioeconómica estructural que integra educación, tipo de vivienda y acceso a servicios básicos. Estadísticamente, (20) señala que en un modelo predictivo la multicolinealidad afecta la interpretabilidad individual de los coeficientes, pero no afecta su capacidad general de predicción. Esta afirmación se comprobó en la práctica mediante un análisis de ablación: al eliminar el Ingreso Familiar del modelo GLM Gamma, el RMSE de validación se incrementó de 0.045 a 0.108 (+138%) y el pseudo- R^2 (D^2) cayó de 0.840 a 0.448; al retirar el Estrato Social, el deterioro fue equivalente ($RMSE = 0.110$, +142%; $D^2 = 0.476$). La Prueba de Razón de Verosimilitud confirma que ambas contribuciones son altamente significativas (ING_FAM: $\chi^2 = 106.6$, $p < 2.2 \times 10^{-16}$; EST_SOC: $\chi^2 = 99.1$, $p < 2.2 \times 10^{-16}$). Adicionalmente, a pesar de los VIF elevados, ambas variables conservan estadísticos z elevados en el modelo (ING_FAM: $z = 35.50$; EST_SOC_D: $z = 30.11$), o que confirma que la inflación de varianza no compromete su significancia individual. Por lo tanto, los coeficientes deben interpretarse como parte de un sistema de predicción en conjunto.

2.6. Herramientas Computacionales y Reproducibilidad

Todo el flujo de procesamiento de datos, imputación y modelado estadístico se ejecutó en el entorno de programación R (versión 4.3.9), utilizando la interfaz RStudio. Para la manipulación de datos se empleó la colección de paquetes *tidyverse*, mientras que el ajuste de los Modelos Lineales Generalizados se realizó mediante las librerías *stats* (*familia Gamma*) y *tweedie*. La generación de gráficos y diagnósticos visuales se soportó en *ggplot2* y *gridExtra*, asegurando la calidad editorial de las figuras presentadas.

III. RESULTADOS

A continuación, se exponen los hallazgos empíricos derivados del procesamiento estadístico y el modelado matemático descritos en la metodología. La sección III se estructura iniciando con la caracterización descriptiva de la muestra, seguida por la evaluación comparativa del desempeño de los modelos predictivos candidatos y, finalmente, la interpretación de los coeficientes determinantes en la generación de residuos sólidos domiciliarios.

3.1. Caracterización de la Muestra

El análisis descriptivo de los 1195 hogares encuestados revela una marcada heterogeneidad en la generación de residuos. Como se detalla en la **Tabla 1**, la producción media de residuos orgánicos (PPC_RSo) se situó en $0.34 \pm 1.31 \frac{kg}{hab/día}$, constituyendo la fracción mayoritaria del flujo de residuos. En contraste, la fracción reciclable (PPC_RSr) presenta un promedio significativamente menor $0.083 \frac{kg}{hab/día}$.

Un hallazgo crítico es la alta asimetría positiva observada en las variables de producción (*Coef. Asimetría* > 3.0), lo que indica que la distribución de los datos no es normal y presenta una cola pesada hacia la derecha; esto confirma la presencia de hogares que generan grandes cantidades de desechos, lo que desvía el promedio y refuerza la necesidad de utilizar métodos de regresión robusta en lugar de mínimos cuadrados ordinarios.

Variable	Descripción y Unidad	Media	D.E.	Mínimo	Máximo	Asimetría
PPC_RS0	PPC Orgánicos (kg/hab/día)	0.334	0.187	0.025	2.650	3.02
PPC_RSr	PPC Reciclables (kg/hab/día)	0.083	0.050	0.005	0.651	3.19
ING_FAM	Ingreso Familiar (USD/mes)	593.94	152.79	250.00	1033.00	-0.03
EST_HOG	Tamaño del Hogar (nº miembros)	4.12	1.52	1.00	12.00	0.92

Tabla 1. Estadística descriptiva de variables críticas.

Para visualizar la relación bivariada entre la capacidad económica y la generación de residuos, controlando por la estratificación social, se generaron gráficos de dispersión facetados. Como se observa en la **Figura 1**, existe una tendencia general al alza: al aumentar el ingreso mensual de la familia, también crece la cantidad de residuos orgánicos que genera cada persona. Sin embargo, este comportamiento no es igual en todos los casos; las líneas del gráfico varían según el nivel social, y la dispersión de los datos indica que el nivel de ingresos, aunque influye mucho, no es el único factor que determina cuánta basura orgánica se produce.

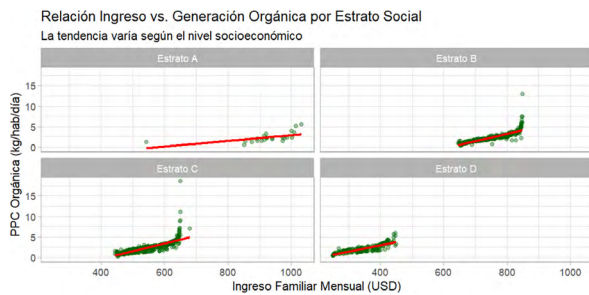


Figura 1. Relación ingreso vs generación orgánica por estrato social.

La lectura por estrato revela patrones diferenciados. En el Estrato A (nivel alto), la relación ingreso–generación es relativamente moderada y con mayor dispersión, lo que sugiere que el consumo de alimentos procesados y semipreparados atenúa la carga orgánica pese al mayor poder adquisitivo. En el Estrato B (nivel medio-alto), la dispersión es menor y la tendencia positiva más pronunciada, indicando mayor homogeneidad en los hábitos de compra de alimentos frescos en el mercado. El Estrato C (nivel medio) presenta la relación más consistente entre ingreso y generación, con escasos valores atípicos y una pendiente estable. El Estrato D (nivel bajo) muestra, paradójicamente, la mayor generación orgánica per cápita a pesar de los ingresos más reducidos: este patrón refleja los hábitos de preparación de alimentos desde insumos crudos —compras en mercados locales, cocina tradicional sin procesados— propios de las economías domésticas con menor acceso a bienes industrializados. Esta heterogeneidad

entre estratos valida la inclusión del estrato social como variable explicativa independiente del ingreso en el modelo.

3.2. Patrones multivariantes de generación y relaciones de multicolinealidad

El Análisis de Componentes Principales, ejecutado sobre datos escalados para evitar sesgos por magnitud, reveló que la dimensionalidad del problema puede reducirse significativamente. Como se observa en la **Figura 2**, las dos primeras dimensiones explican en conjunto el 86.3% de la varianza total de los datos.

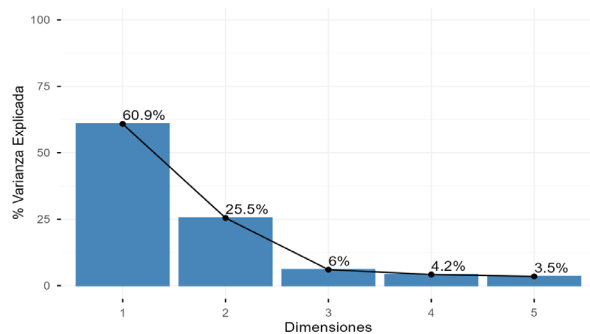


Figura 2. Resultados de análisis de componentes principales

La primera componente (Dimensión 1) explica el 60.8% de la varianza y representa la "escala de generación global", donde todas las fracciones de residuos aportan positivamente a la variabilidad. Por otro lado, la segunda componente (Dimensión 2) explica el 25.5% y está asociada principalmente a la diferenciación entre residuos orgánicos y reciclables.

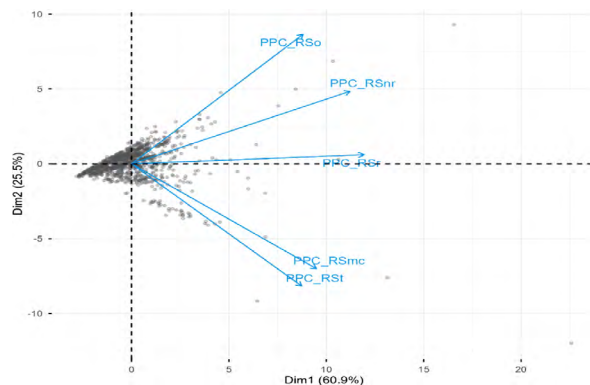


Figura 3. Ortogonalidad entre los vectores de residuos orgánicos.

Un hallazgo visual importante es la separación angular entre los vectores de residuos orgánicos y reciclables mostrados en la **Figura 3**. Esto sugiere que estas dos conductas mantienen una independencia parcial, la generación elevada de residuos orgánicos no predice obligatoriamente una mayor generación de reciclables en los poblados estudiados, lo que indica hábitos de consumo y desecho diferenciados para cada corriente.

El análisis matemático mediante el Factor de Inflación de la Varianza (VIF) reveló una fuerte conexión entre el Ingreso Familiar ($VIF = 10.94$) y el Estrato Social ($VIF_{máx} = 29.94$). Como estos resultados superan el límite tradicional de 10, se confirma que existe un vínculo muy estrecho entre el poder adquisitivo de las familias y el sector socioeconómico al que pertenecen.

A pesar de esta similitud de datos, al comparar los resultados se comprobó que el sistema de cálculo completo funciona mucho mejor que las versiones más simples, logrando una mejora importante en la precisión general ($AIC_{completo} = -3255$ frente a $AIC_{simplificado} = -1796$). Estos números demuestran que es necesario mantener ambos factores en la fórmula final, ya que cada uno aporta detalles únicos sobre cómo se genera la basura orgánica, detalles que el otro no puede explicar por sí solo.

En la Figura 4 se observa una separación muy clara, donde los ingresos promedios cambian de un grupo a otro sin cruzarse ni mezclarse. Esta distribución demuestra que conocer la clase social ayuda a predecir casi de forma directa el nivel de ingresos de una familia. Esto explica matemáticamente por qué se detecta tanta relación entre ambas variables ($VIF > 10$), pero al mismo tiempo confirma que es indispensable usar ambas para que el modelo predictivo sea totalmente preciso.

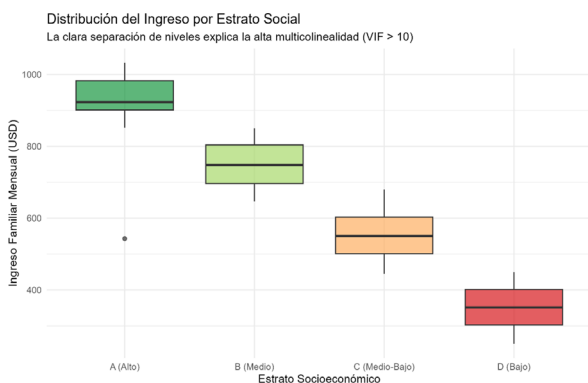


Figura 4. Distribución del ingreso familiar mensual agrupado por estrato socioeconómico.

3.2. Ajuste y validación del modelo predictivo

La evaluación inicial mediante regresión lineal clásica (OLS) reveló deficiencias estructurales en el ajuste. Las pruebas de diagnóstico sobre los residuos confirmaron la violación del supuesto de normalidad Test de Shapiro-Wilk:

$W = 0.3805, p < 2.2 \times 10^{-16}$, evidenciada visualmente en los gráficos Q-Q (*Quantile-Quantile*) por una desviación sistemática en las colas de la distribución. Esto confirmó que un modelo lineal tradicional no es adecuado para capturar la dinámica de generación de residuos en la zona de estudio debido a la alta variabilidad intra-domiciliaria y la presencia de observaciones influyentes. A continuación, se muestra la ecuación matemática que gobierna este modelo con los respectivos coeficientes.

$$PPC_{RSO} = -1.782 + 0.0023 * ingreso_{fam} - 0.0096 * tam_{hogar} + 0.392 * SOCb + 0.883 * SOCc + 1.328 * SOCd + \sum_{j \neq Cev} \beta_j * Pob_j \quad (1)$$

En la ecuación presentada, la variable PPC_RSO representa el cálculo estimado de la basura orgánica diaria generada por persona. Los valores de la fórmula miden el impacto de cada factor: el número asociado al Ingreso muestra cuánto aumenta la generación de residuos por cada dólar extra, mientras que la variable Hogar adapta el cálculo según la cantidad de personas que viven en la casa, reflejando cómo se ahorran recursos al compartir gastos. Las variables Estrato sirven para ajustar la predicción según el nivel socioeconómico (B, C y D), tomando siempre al Estrato A como punto de comparación. Finalmente, la parte de la sumatoria agrupa las características geográficas de las catorce localidades estudiadas, midiendo las diferencias de cada territorio frente a un pueblo de referencia (Cevallos), cuyo valor base es cero.

Por otro lado, la Figura 5 compara la precisión de los errores del modelo OLS frente a un patrón normal ideal (representado por una línea roja). En el gráfico se nota una fuerte desviación de los puntos en la esquina superior derecha, alejándose de la línea de referencia. Este comportamiento demuestra la existencia de datos muy extremos y una clara inclinación en la información, confirmando tanto de forma visual como matemática ($p < 0.05$) que los datos no tienen un comportamiento normal. Toda esta evidencia justifica y respalda la decisión de utilizar un sistema de cálculo más resistente.

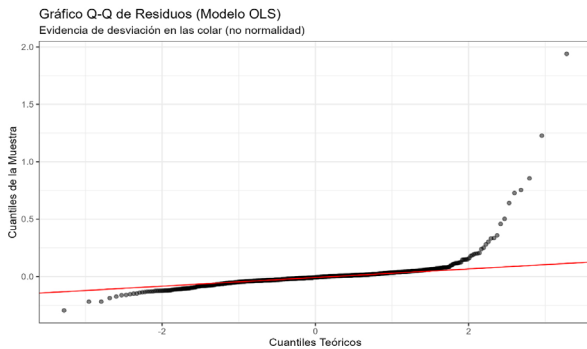


Figura 5. Diagrama Q-Q (Quantile-Quantile) de los residuos del modelo de regresión lineal estándar.

Ante estas limitaciones, la aplicación del Modelo de Regresión Robusta (RLM) corrigió las desviaciones, atenuando el impacto de los hogares “super-generadores” mediante el uso de estimadores-M. Los parámetros estimados del modelo final para la producción de residuos orgánicos se presentan en la Tabla 2.

Variable Predictora	Coefficiente (β)	Error Estándar	Valor-t
Intercepto	-1.671	0.025	-67.46
Ingreso Familiar	0.0021	0.0000	90.99
Tamaño del Hogar	-0.0026	0.0008	-3.24
Estrato Social B (Ref: A)	0.378	0.009	44.46
Estrato Social C (Ref: A)	0.833	0.011	72.95
Estrato Social D (Ref: A)	1.250	0.0015	82.67

Tabla 2. Coeficientes estimados mediante Regresión Robusta para la Producción de Residuos Orgánicos.

Para validar visualmente la superioridad del estimador robusto frente a la presencia de observaciones influyentes, se contrastaron las curvas de ajuste de ambos modelos. Como se aprecia en la Figura 6, la recta de regresión clásica tiende a sobreestimar la producción en los rangos medios y altos de ingreso al ser "arrastrada" por los hogares super-generadores. En contraste, el ajuste robusto (línea azul) resiste esta distorsión, proporcionando una estimación más conservadora y representativa del comportamiento poblacional típico.

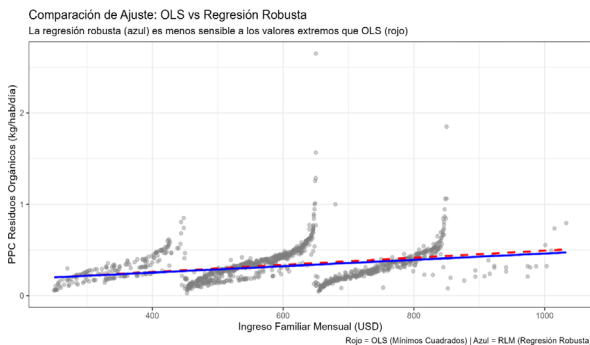


Figura 6. Comparación entre el modelo lineal tradicional (OLS) y el método estadístico más resistente (RLM) para estimar la producción de residuos orgánicos.

A continuación, se muestra la ecuación matemática para el modelo lineal robusto.

$$PPC_{RSO}^{rob} = -1.671 + 0.0021 * ingreso_{fam} - 0.0026 * tam_{hogar} + 0.378 * SOCb + 0.833 * SOCc + 1.250 * SOCd + \sum_{j \neq Cev} \beta_j^{rob} * Pob_j$$

En este modelo estadístico más resistente, los valores se calcularon utilizando un método que no se altera fácilmente por los errores extremos (estimador-M de Huber), a diferencia del sistema tradicional (suma de cuadrados del modelo OLS). En términos matemáticos, esto significa que el impacto de los ingresos (+0.0021) y de la cantidad de personas en la familia (-0.0026) refleja el comportamiento real de la mayoría de los hogares, ignorando las alteraciones provocadas por los datos inusuales o extremos. Nótese que el coeficiente del ingreso es ligeramente menor que en el modelo OLS, confirmando que la regresión clásica estaba siendo inflada por hogares con alto consumo y alta generación

A pesar de las mejoras introducidas por la regresión robusta, la naturaleza de los datos de residuos sólidos presenta dos desafíos estructurales que los modelos lineales (aditivos) no resuelven plenamente, la estricta positividad de la variable (nunca se generan residuos negativos) y la heterocedasticidad (la varianza del error aumenta conforme aumenta la media de producción). Para abordar esto, se implementó un Modelo Lineal Generalizado (GLM) con distribución Gamma y función de enlace logarítmica. Este método resulta mejor en teoría para datos que siempre son mayores a cero y que tienden a agruparse hacia un extremo (variables continuas asimétricas positivas). Esto se debe a que calcula cómo interactúan y se multiplican los factores con el promedio (relación multiplicativa), logrando entender de forma más exacta cómo el consumo crece de manera acelerada (crecimiento exponencial) a medida que aumentan los ingresos.

Por su parte, la **Tabla 3** resume los valores matemáticos (coeficientes) calculados a través del Modelo GLM Gamma. Esta herramienta estadística fue la que logró adaptarse con mayor precisión a la naturaleza irregular e impredecible de la información recolectada (estructura estocástica de los datos). Se observa que tanto el Ingreso Familiar como el Estrato

Social (B, C, D) presentan coeficientes positivos y estadísticamente significativos ($p < 0.001$), validando su fuerte influencia incremental sobre la generación per cápita. Por el contrario, el Tamaño del Hogar exhibe un efecto negativo significativo, confirmando la hipótesis de eficiencia por economías de escala.

Variable Predictora	Coefficiente β	Error Estándar	Valor-z
(Intercepto)	-7.105	0.199	-35.71
Ingreso Familiar	0.0067	0.0002	35.50
Tamaño del Hogar	-0.020	0.0065	-3.07
Estrato Social B (Ref: A)	0.914	0.068	13.38
Estrato Social C (Ref: A)	2.344	0.092	25.54
Estrato Social D (Ref: A)	3.659	0.121	30.11

Tabla 3. Estimaciones del Modelo GLM Gamma (Log-Link).

Dado que el GLM Gamma opera con función de enlace logarítmica, los coeficientes tienen interpretación multiplicativa. La Tabla 4 presenta los efectos multiplicativos $\exp(\beta)$ con sus respectivos intervalos de confianza al 95%, calculados como $\exp(\beta \pm 1.96 \times SE)$.

Variable Predictora	$\exp(\beta)$	IC 95% Inferior	IC 95% Superior
Ingreso Familiar	1.0067	1.0063	1.0071
Tamaño del Hogar	0.980	0.967	0.993
Estrato Social B (Ref: A)	2.49	2.18	2.85
Estrato Social C (Ref: A)	10.42	8.71	12.46
Estrato Social D (Ref: A)	38.83	30.65	49.23

Tabla 4. Efectos multiplicativos del modelo GLM Gamma — $\exp(\beta)$ con intervalos de confianza al 95%.

En términos prácticos, por cada dólar adicional de ingreso mensual, la producción media de residuos orgánicos se multiplica por 1.0067, equivalente a un incremento del 0.67%. Para ilustrarlo mejor, si los ingresos mensuales de una familia aumentan en \$100, la cantidad esperada de basura crece en una proporción de $\exp(0.0067 \times 100) \approx 1.96$, lo que significa que prácticamente se duplica. En cuanto al tamaño de la familia, por cada persona adicional que vive en la casa, la cantidad de residuos por individuo se reduce al 98% del nivel anterior ($\exp(-0.020) = 0.98$), lo que confirma el ahorro de recursos al compartir gastos. Por otro lado, el valor calculado para el sector de menores recursos (Estrato D) demuestra que, en promedio, estos hogares producen casi 39 veces más desechos orgánicos por persona que el sector de mayores ingresos tomado como referencia (Estrato A). Esto evidencia la gran diferencia que existe entre los distintos grupos sociales.

Para evaluar la exactitud matemática del modelo, se revisaron otros indicadores: el error restante

(devianza residual) fue de 43.41 con 939 grados de libertad, en comparación con un error inicial (devianza nula) de 271.93 (958 g.l.). Estos números dan como resultado un indicador de precisión (pseudo- R^2 de devianza D^2) de 0.840, lo que significa que la fórmula logra explicar el 84.0% del comportamiento de los datos. Finalmente, otra medida clave (Criterio de Información de Akaike o AIC = -2561.3) confirma que el sistema de cálculo es excelente, logrando un equilibrio perfecto entre su nivel de complejidad y su capacidad para ofrecer resultados precisos. El parámetro de dispersión estimado de la distribución Gamma ($\phi = 0.074$) es bajo, lo que refleja alta precisión del modelo en la caracterización de la variabilidad intra-grupal de la generación de residuos orgánicos.

A continuación, se muestra la fórmula matemática que gobierna el modelo GLM

$$PPC_{RSO} = \exp(-7.105 + 0.0066 * ingreso_{fam} - 0.0020 * tam_{hogar} + 0.914 * SOCb + 2.344 * SOCc + 3.659 * SOCd + \sum_{j \neq Cev} \beta_j^{rob} * Pobj) \quad (3)$$

La ecuación Gamma utiliza una estructura multiplicativa debido a la función de enlace logarítmica. Esto significa que los coeficientes no suman cantidades fijas de residuos, sino que actúan como tasas de cambio porcentual. Por ejemplo, el valor calculado para el Ingreso (0.0067) indica que, por cada unidad económica extra, la cantidad de basura producida aumenta en una proporción constante. Esto provoca que la generación de desechos crezca de forma mucho más rápida en los sectores de mayores recursos.

Como se muestra en la Figura 6, la precisión de las predicciones cambia dependiendo de la medida que se evalúe. El modelo estadístico avanzado (Modelo GLM Gamma) resulta ser el más efectivo porque logra reducir al mínimo los errores de cálculo (Error Cuadrático Medio o RMSE: 0.045). Esto demuestra que este sistema maneja mucho mejor la irregularidad de la información y los datos que resultan demasiado extremos. Por otro lado, el Modelo Robusto alcanza el menor Error Absoluto (MAE: 0.029), debido a su optimización específica para resistir observaciones influyentes en la media. En conjunto, estos resultados confirman que, aunque la regresión robusta es altamente precisa en promedio, el modelo Gamma ofrece la estructura estocástica y la capacidad predictiva global más consistente para los datos de residuos domiciliarios.

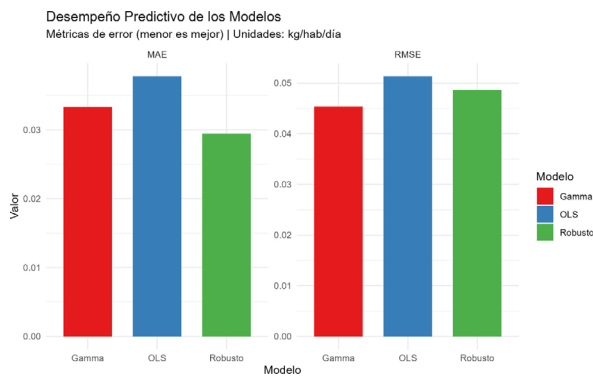


Figura 7. Comparativa de modelos.

Por otro lado, la **Figura 7** muestra las estimaciones del modelo estadístico (modelo Gamma) para diferentes situaciones familiares. En el gráfico se observa un crecimiento acelerado (tendencia exponencial): cuando los ingresos mensuales sobrepasan los \$600 (eje X), la línea sube de forma pronunciada. Esto indica que, en las familias con mayor poder adquisitivo, el nivel de consumo y la cantidad de basura que se desecha aumentan mucho más rápido.

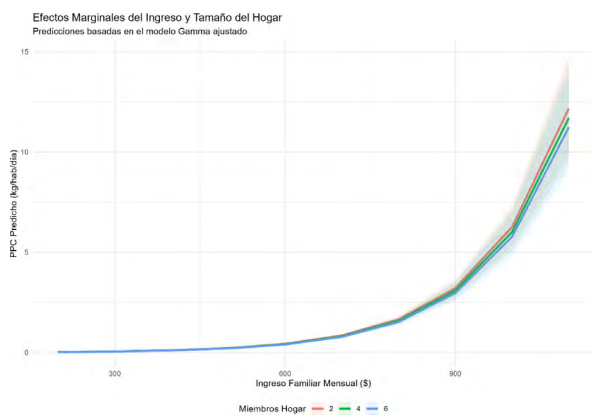


Figura 8. Efectos Marginales.

Además, las líneas de colores revelan el efecto de escala. La línea roja (hogares pequeños de 2 personas) está siempre por encima de la línea azul (hogares grandes de 6 personas). Esto prueba visualmente que vivir en familias numerosas es más eficiente, generando menor cantidad de residuos per cápita en comparación con vivir en pareja o solo, debido al uso compartido de alimentos y recursos."

Habiendo establecido las propiedades estadísticas del modelo seleccionado, a continuación, se discuten las implicaciones de estos hallazgos en el contexto de la gestión de residuos, contrastando la evidencia empírica con la teoría económica y estudios previos la evidencia empírica con la teoría económica y estudios previos

IV. DISCUSIÓN

Los resultados obtenidos en esta investigación validan la premisa de que la generación de residuos sólidos domiciliarios en poblaciones menores a 70.000 habitantes no sigue una distribución normal, sino que presenta una marcada asimetría positiva. La superioridad estadística del Modelo Lineal Generalizado (GLM) con distribución Gamma sobre los enfoques lineales (OLS y RLM) —demostrada por la significativa reducción del Error Cuadrático Medio (RMSE)— evidencia que, en contextos de alta heterogeneidad socioeconómica y valores atípicos severos, asumir una relación estocástica adecuada es indispensable para evitar sesgos en la planificación de la gestión integral. Este resultado es consistente con la revisión (19), quienes señalan que los modelos de regresión con variables socioeconómicas son los más apropiados para poblaciones de escala pequeña, aunque advierten que su desempeño depende críticamente del ajuste distribucional a los datos locales.

Un hallazgo central del estudio es la relación positiva y altamente significativa entre el ingreso familiar y la producción de residuos orgánicos ($z = 35.50$). Este comportamiento es consistente con la teoría macroeconómica y lo reportado por Hoornweg y Bhada-Tata (1), quienes establecen que el incremento en el poder adquisitivo impulsa el consumo per cápita de bienes. Sin embargo, nuestro Análisis de Componentes Principales (PCA) escalado matiza esta afirmación al revelar una separación angular significativa entre la fracción orgánica y la reciclable. Esto sugiere una independencia parcial entre ambas corrientes, indicando que el aumento del ingreso no incrementa todas las fracciones de residuos por igual, sino que responde a dinámicas de consumo diferenciadas.

Resulta particularmente interesante el patrón observado en la estratificación social. Contrario a la intuición lineal de que "a mayor estatus, mayor generación total", el modelo estadístico (modelo Gamma) demostró que las familias del sector de menores recursos (Estrato D) producen una cantidad significativamente mayor de residuos orgánicos por persona ($\beta = 3.659$) en comparación con los sectores más altos. Esta aparente contradicción se debe a las costumbres al cocinar: en estas zonas, las familias de menores ingresos suelen comprar alimentos frescos y preparar

la comida desde cero, lo que produce más desechos naturales. Por el contrario, los sectores de mayores ingresos consumen más alimentos procesados o prefabricados, cambiando su tipo de basura hacia envases y materiales inorgánicos, tal como lo indican estudios en otros países en desarrollo (26). Desde el punto de vista del aprovechamiento de recursos (Economía Circular), este descubrimiento tiene una utilidad práctica muy directa. Saber que los sectores de bajos recursos concentran la mayor cantidad de desechos orgánicos representa una excelente oportunidad para crear programas de abono casero (compostaje) y sistemas de recolección separada. De esta forma, se le da un nuevo valor a esta basura y se reduce drásticamente la cantidad de desperdicios que terminan en los basureros tradicionales, lo cual es una meta ambiental prioritaria para Ecuador y toda la región (5).

Finalmente, el valor negativo correspondiente al tamaño de la familia ($\beta = -0.020$) confirma que se produce un ahorro natural de recursos al convivir bajo un mismo techo. Este fenómeno, ya estudiado por investigadores como Dennison et al. (27), significa que los productos se aprovechan mejor a medida que aumenta la cantidad de habitantes. Dicho de otro modo, la basura generada por cada persona disminuye un poco en las familias numerosas porque comparten bienes y organizan de forma mucho más eficiente la preparación de sus comidas. Si los municipios no toman en cuenta este comportamiento al planificar sus presupuestos y actividades, podrían cometer el error de calcular una necesidad exagerada de servicios de recolección de basura en aquellos sectores donde vive mucha gente agrupada.

Desde el punto de vista de la planificación municipal, los resultados de este estudio tienen una utilidad práctica y directa para organizar el manejo de basura en localidades pequeñas y medianas. Al confirmar que los ingresos y el nivel social son los factores más importantes, los municipios pueden dividir su territorio según las características económicas de cada barrio y programar los recorridos de los camiones de forma distinta. Por ejemplo, los sectores de menores recursos necesitan una recolección de basura orgánica más frecuente, ya que producen una mayor cantidad por persona y estos desechos se descomponen rápidamente. Además, el ahorro de recursos que se genera al interior de las familias numerosas indica que, al calcular las

necesidades de limpieza en zonas con muchas viviendas (principalmente en los sectores B y C), se debe tomar en cuenta el tamaño promedio de los hogares. Esto evita el error de comprar o movilizar más vehículos de los que realmente se necesitan. En Ecuador, donde la mayoría de los municipios de este tamaño no cuenta con información propia o actualizada, este modelo matemático ofrece cálculos exactos y adaptados a la realidad local. Su uso resulta fundamental para elaborar proyectos de planificación eficientes (planes de gestión integral de residuos sólidos o PGIRS) que cumplan correctamente con las leyes ambientales vigentes.

Al analizar cómo reacciona el sistema de cálculo ante los cambios económicos de cada persona (sensibilidad del modelo), los resultados matemáticos (efectos marginales) obtenidos del modelo estadístico (modelo Gamma) indican que, dentro de los niveles de dinero estudiados (USD 250–1033 mensuales), la producción de basura crece de forma acelerada cuando los ingresos superan los USD 600 al mes. A partir de esa cantidad, la línea de estimación sube con mucha más fuerza. Esto significa que, si se aplican planes de ayuda económica o se aumentan los sueldos de manera constante en las familias de recursos medios y bajos, la cantidad de desechos orgánicos aumentará de forma considerable. Por lo tanto, es fundamental que los municipios tomen en cuenta estos posibles aumentos al momento de planificar y organizar sus sistemas de recolección para los próximos años (proyecciones de mediano plazo). La robustez del modelo ante estos escenarios es consistente dentro del dominio empírico observado; sin embargo, su capacidad predictiva decrece en escenarios de ingreso superiores al máximo observado (USD 1033/mes), para los cuales se requeriría recalibración con datos adicionales.

V. CONCLUSIONES

Se demostró que los modelos lineales clásicos (OLS) son insuficientes para predecir la generación de residuos en poblados menores a 70.000 habitantes, debido a la violación sistemática del supuesto de normalidad (Shapiro-Wilk: $W=0.3805$, $p < 2.2 \times 10^{-16}$). Si bien el uso del método estadístico más resistente (Regresión Robusta o RLM) logró disminuir el impacto de los datos inusuales al reducir el error promedio (MAE: 0.029), el modelo avanzado (Modelo Lineal

Generalizado o GLM) con distribución Gamma resultó ser la mejor opción para la planificación general. Este último sistema no solo controló los efectos de los datos exagerados al reducir al mínimo el error cuadrático (RMSE: 0.045), sino que también logró adaptarse correctamente al hecho de que los resultados siempre son números positivos y que la información tiende a agruparse de forma desigual (naturaleza asimétrica de los datos). Esto confirma que, en contextos de alta heterogeneidad y variabilidad multivariante (validada mediante PCA escalado), los estimadores basados en la familia Gamma proporcionan la base más confiable y teórica para el diseño de sistemas de recolección.

Por otra parte, el ingreso familiar se consolidó como el predictor más influyente en la producción de residuos orgánicos, presentando el estadístico de prueba más elevado del estudio ($z = 35.50$) y una relación positiva ($\beta = 0.0067$). Sin embargo, se detectó un comportamiento distinto según el nivel social: las familias con menos recursos (Estrato D) tienen una producción inicial de basura orgánica mucho más alta ($\beta = 3.659$; entendido como un multiplicador de la cantidad base) que las familias de nivel alto (Referencia A). Este hallazgo indica que, aunque tener más dinero hace que se consuma más en total, la forma de alimentarse en los sectores bajos (donde se preparan más alimentos frescos) aumenta específicamente la cantidad de desechos orgánicos.

También se comprobó matemáticamente que existe un ahorro de recursos dentro de la familia. El valor negativo relacionado con el tamaño

del hogar ($\beta = -0.020$) indica que la basura no aumenta de forma proporcional a la cantidad de personas; al contrario, se nota que la producción por individuo disminuye un poco a medida que más miembros viven en la misma casa. Tomar en cuenta este factor es fundamental para no calcular mal y evitar el uso de demasiados camiones de basura en las zonas donde vive mucha gente.

Finalmente, el estudio avanzado de los datos (Análisis de Componentes Principales o PCA escalado) reveló que la producción de basura es un sistema complejo donde influyen muchos factores. La primera parte de este análisis (primera componente) explica el 60.8% de toda la basura generada. La diferencia detectada en el gráfico entre los desechos orgánicos y los materiales reciclables demuestra que son situaciones que ocurren de forma independiente. Por lo tanto, las leyes municipales no deben asumir que quienes más basura producen son, de forma automática, quienes más reciclan; esto exige crear estrategias e incentivos distintos y específicos para cada tipo de residuo.

IV. AGRADECIMIENTOS

Los autores agradecen a la Universidad Nacional de Chimborazo por la provisión de la base de datos, a los Gobiernos Autónomos Descentralizados de los 15 poblados por las facilidades de acceso territorial, y a la ESPOCH y la UPEC como instituciones de adscripción del equipo investigador.

VII. REFERENCIAS

1. Kaza S, Yao L, Bhada-Tata P, Van Woerden F. What a waste 2.0: a global snapshot of solid waste management to 2050. Washington (DC): World Bank; 2018.
2. Chen DMC, Bodirsky BL, Krueger T, Mishra A, Popp A. The world's growing municipal solid waste: trends and impacts. *Environ Res Lett.* 2020;15(7):074021. doi: 10.1088/1748-9326/ab8659.
3. Ferronato N, Torretta V. Waste mismanagement in developing countries: a review of global issues. *Int J Environ Res Public Health.* 2019;16(6):1060. doi: 10.3390/ijerph16061060.
4. Abdallah M, Abu Talib M, Feroz S, Nasir Q, Abdalla H, Mahfood B. Artificial intelligence applications in solid waste management: a systematic research review. *Waste Manag.* 2020;109:231–46. doi: 10.1016/j.wasman.2020.04.057.
5. Gómez-Sanabria A, Kiesewetter G, Klimont Z, Schöpp W, Haberl H. Potential for future reductions of global GHG and air pollutant emissions from circular waste management systems. *Nat Commun.* 2022;13(1):106. doi: 10.1038/s41467-021-27624-7.

6. Sáez A, Urdanego J. Manejo de residuos sólidos en zonas rurales y urbanas pequeñas de América Latina: desafíos y oportunidades. *Rev Int Contam Ambient.* 2021;37:67–79. doi: 10.20937/RICA.53671.
7. Miezah K, Obiri-Danso K, Kádár Z, Fei-Baffoe B, Mensah MY. Municipal solid waste characterization and quantification as a measure towards effective waste management in Ghana. *Waste Manag.* 2015;46:15–27. doi: 10.1016/j.wasman.2015.09.009.
8. Ayeleru OO, Dlamini S, Mbohwa C. The reality of solid waste management in South Africa: a review. *J Clean Prod.* 2020;272:122695. doi: 10.1016/j.jclepro.2020.122695.
9. Ercolano S, Gaeta GL, Ghinoi S, Silvestri F. The role of institutions in the Environmental Kuznets Curve for municipal solid waste generation. *Ecol Econ.* 2018;150:66–73. doi: 10.1016/j.ecolecon.2018.04.005.
10. Ali S, Pervaiz A, Afzal S, Hamid N, Yasmin A. Socio-economic factors affecting household solid waste generation in selected vicinities of Lahore City. *J Mater Cycles Waste Manag.* 2019;21(3):619–31. doi: 10.1007/s10163-018-0826-y.
11. Nguyen TT, Zhu D, Le NP. Factors influencing waste separation intention of residential households in a developing country: evidence from Vietnam. *Habitat Int.* 2019;90:102008. doi: 10.1016/j.habitatint.2019.102008.
12. Odonkor ST, Sallar AM. Determinants of household solid waste management practices in the urban areas of Africa: a systematic review. *Environ Challenges.* 2021;4:100137. doi: 10.1016/j.envc.2021.100137.
13. Khan D, Kumar A, Samadder SR. Impact of socioeconomic status on municipal solid waste generation rate. *Waste Manag.* 2016;49:15–25. doi: 10.1016/j.wasman.2016.01.019.
14. Ojeda-Benitez S, Cruz-Sotelo SE, Taboada-Gonzalez P. Impact of household size on solid waste generation in Northern Mexico. *Resour Conserv Recycl.* 2020;155:104652. doi: 10.1016/j.resconrec.2019.104652.
15. Trang PTT, Dong HQ, Toan TQ, Hanh NTX, Thu NT. The effects of socio-economic factors on household solid waste generation and composition: a case study in Thu Dau Mot, Vietnam. *Energy Procedia.* 2017;107:253–8. doi: 10.1016/j.egypro.2016.12.144.
16. Kannangara M, Dua R, Ahmadi L, Bensebaa F. Modeling and prediction of municipal solid waste generation: a review. *Waste Manag.* 2018;74:105–18. doi: 10.1016/j.wasman.2018.01.030.
17. Hettiarachchi H, Meegoda JN, Ryu S. Organic waste management: new avenues for capacity building and smart waste management systems. *Bioresour Technol.* 2018;270:732–40. doi: 10.1016/j.biortech.2018.09.071.
18. Dunn PK, Smyth GK. *Generalized linear models with examples in R.* New York: Springer; 2018.
19. Beigl P, Lebersorger S, Salhofer S. Modelling municipal solid waste generation: a review. *Waste Manag.* 2008;28(1):200–14. doi: 10.1016/j.wasman.2006.12.011.
20. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate data analysis.* 8th ed. Andover: Cengage Learning; 2019.
21. Torkashvand J, Farzadkia M, Namin MS, Rahmatian A. Predicting the generation of municipal solid waste based on independent variables of the impact: a systematic review. *Iran J Public Health.* 2021;50(3):441–52. doi: 10.18502/ijph.v50i3.5583.
22. Noufal M, Liu Y, Maalla Z, Adipah S. Determinants of household solid waste generation and composition in Homs City, Syria. *J Environ Public Health.* 2020;2020:7460356. doi: 10.1155/2020/7460356.
23. Cubillos M. Análisis comparativo de modelos de predicción de residuos sólidos urbanos mediante técnicas de machine learning. *Inf Tecnol.* 2021;32(4):11–22. doi: 10.4067/S0718-

07642021000400011.

24. Malebana P, Ayeleru OO, Mbohwa C. Statistical modeling of municipal solid waste generation rates in Johannesburg, South Africa. *Procedia Manuf.* 2019;35:269–74. doi: 10.1016/j.promfg.2019.05.037.
25. Faris H, Aljarah I, Al-Madi N, Mirjalili S. Optimizing the learning process of feedforward neural networks using multi-verse optimizer. In: Mirjalili S, Faris H, Aljarah I, editors. *Recent advances in intelligent paradigms and applications*. Cham (CH): Springer; 2019. p. 121–42. doi: 10.1007/978-3-030-01916-7_6.
26. Reynolds C, Goucher L, Quested T, Bromley S, Gillick S, Wells V, et al. Consumption-stage food waste reduction interventions – what works and how to design better interventions. *Food Policy.* 2019;83:7–27. doi: 10.1016/j.foodpol.2019.01.009.
27. Dennison GJ, Dodd VA, Whelan FJ. A socio-economic based survey of household waste characteristics in the city of Dublin, Ireland. *Resour Conserv Recycl.* 1996;17(4):227–44. doi: 10.1016/S0921-3449(96)01153-4.
28. Ministerio del Ambiente del Ecuador. *Informe nacional sobre la gestión de residuos sólidos en el Ecuador*. Quito: Ministerio del Ambiente del Ecuador; 2020.
29. Hoornweg D, Bhada-Tata P. *What a waste: a global review of solid waste management*. Washington (DC): World Bank; 2012. (Urban development series knowledge papers; no. 15).
30. Arellano B, Rivera-Huerta ML, Rodríguez-Castellanos FE, Aguilar-Virgen Q. Methodology for characterization of municipal solid waste in small urban populations. *Waste Manag.* 2014;34(11):1967–74. doi: 10.1016/j.wasman.2014.07.012.