frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 27 February 2025
DOI 10.3389/fams.2025.1498513

OPEN ACCESS

EDITED BY
Artur Lemonte,
Federal University of Rio Grande do Norte, Brazil

REVIEWED BY
Kazim Onur Demirarslan,
Artvin Coruh University, Türkiye
Yanxing Yang,
Caldwell University, United States

*CORRESPONDENCE
Jorge L. Padilla-Vento
✉ jorge.padillav@upeu.edu.pe

RECEIVED 19 September 2024
ACCEPTED 10 January 2025
PUBLISHED 27 February 2025

CITATION
Padilla-Vento JL and Soria JJ (2025) Statistical approach for the evaluation of household solid waste generation in Peruvian households: 2014–2021.
Front. Appl. Math. Stat. 11:1498513.
doi: 10.3389/fams.2025.1498513

Statistical approach for the evaluation of household solid waste generation in Peruvian households: 2014–2021

Jorge L. Padilla-Vento* and Juan J. Soria

Unidad de Posgrado de Ingeniería y Arquitectura, Escuela de Posgrado, Universidad Peruana Unión, Lima, Peru

This study addresses the evaluation of the generation of domestic solid waste in Peruvian households using statistical techniques and the SEMMA and PCA data mining methodology. The objective is to explore how waste management, population and the *Per Capita* Generation index PCG index per capita influence the production of this waste in Peruvian departments. The sample was obtained from the database of annual reports submitted by district and provincial municipalities to MINAM through the Information System for Solid Waste Management (SIGERSOL), including data from the 24 departments of Peru, with a total of 14,852 records organized in 196 registration forms. Statistical techniques and the adaptation of the SEMMA methodology were applied together with the Principal Component Analysis (PCA) to examine the impacts of the accumulation of household solid waste in Peru. This study showed that the first component accounts for 80.2% of the inertia. Combining the first two components accounts for 99.8% of the total variation, suggesting that most of the meaningful information can be maintained using only two dimensions. Welch's ANOVA showed significant differences in domestic solid waste generation among Peruvian departments [F (6, 94.310) = 790.444; p = 0.0, p < 0.05]. In addition, a square Eta of 99.09% revealed a very large effect size, indicating that the amount of population explains 99.09% of the variation in the generation of this waste between the departments. The PCG index had a moderate effect, suggesting the need for further studies to explore the underlying causes of regional differences and assess the effectiveness of the waste management measures implemented. A positive relationship was found between the production of Domestic Solid Waste (DSW) and the number of inhabitants. Lima stood out with the highest average of DSW 13220.47 tons and the PCG index of 50%. Using Ward's method, three groups were obtained and PCA was applied to each group. In the Group, Lambayeque 5616.48 tons, Loreto 2946.44 tons and San Martín 1596.07 tons registered the highest DSW averages, while Amazonas 441.1 tons obtained the lowest. Ucayali 60%, Loreto 58% and San Martín 57% showed the highest PCG indexes. In Group b, Ayacucho 701.81 tons had the highest average DSW and Apurimac 497 tons the lowest. Tacna and Apurimac with 44% and Moquegua 43% registered the highest PCG indexes, while Huancavelica 42% and Pasco 41% had the lowest. In Group C Piura 4476.53 tons and La Libertad 3478.46 tons showed the highest DSW averages, while Huánuco 859.41 tons and Cajamarca 812.74 tons registered the lowest. Ica and Piura led with an average PCG of 48%, while Puno and Junín with 43% had the lowest values.

KEYWORDS

solid waste, waste management, solid waste management, municipal solid waste, waste disposal

1 Introduction

The generation and management of household solid waste is a global problem that is aggravated by factors such as population growth, economic expansion (1). Li et al. (2) they anticipate an increase in waste generation in the coming years, considering that the global population already exceeds 8 billion inhabitants. Li and Wang (3) they highlight that this phenomenon enhances the production of domestic solid waste. Cheng et al. (4) and Pillay et al. (5) have laid a theoretical foundation by dealing with issues associated with the influence of psychological, social and economic factors on waste segregation behavior. In addition, significant theories have been recognized, such as the critical evaluation of studies related to moral and environmental education (6), and the relevance of humility as an essential virtue for environmental sensitivity (7).

The Ministry of the Environment MINAM (8) points out that in Peru 70% of municipal waste is produced by households; the newspaper El Peruano (9) It reports that during the year 2021, approximately 8 million 214355.90 tons of municipal solid waste were generated, which is equivalent to an average of 22505.08 tons per day. This waste was composed of 56.70% organic matter, 20.94% inorganic, 12.66% non-usable and 9.71% corresponding to hazardous waste. The newspaper Gestión (10) reports that MINAN has instructed the local authorities of San Vicente de Cañete, in Lima, to develop an action plan to address the problem of solid waste (SW) collection in that district.

Despite the efforts made, gaps persist in the understanding of the relationship between the PCG (*Per Capita* Generation) index and the production of domestic solid waste (DSW). This knowledge gap leads to poor solid waste (SW) management and inadequate infrastructure (3). In addition, the pandemic has altered the composition of waste and increased the demand for plastics (11). Requena-Sanchez et al. (12) point out that the lack of essential information to understand the dynamics of SW management hinders the implementation of strategies (13). They highlight the importance of SW management and care for the environment.

In this context, this study aims to fill these knowledge gaps and provide a more complete and detailed view of the problem of the generation of domestic SW in Peruvian households. To achieve this, we have resorted to the statistical approach for data evaluation with the support of the SEMMA data mining methodology, in order to improve the interpretation of data and thus optimize decision-making in SW management. In this sense, the main contribution of this work is to offer a data evaluation scheme that allows visualizing the production of SW and it is expected to explain the influence of the PCG index and the population level on the generation of domestic SW in Peruvian households, constituting a valuable tool for the visualization and targeting of policies to support the management of SW. offering a comprehensive approach to observe the distribution of domestic SW.

This technique consists of the stages: Sample, Explore, Modify, Model and Evaluate, established by the SAS Institute, to perform the essential tasks that facilitate the understanding, structuring, implementation and maintenance of data mining projects (14). Its implementation allowed us to structure a robust analytical process that addresses the heterogeneity and considerable volume of data associated with the generation of DSW. We also use principal component analysis (PCA) is one of the unsupervised learning strategies used in exploratory data analysis (15). In the case of DSW, variables such as SW composition, population levels and per capita generation rates present intricate relationships that make direct analysis difficult. The PCA not only simplified the solid residual data set by reducing it to principal components that capture the greatest explanatory variance, but also allows for the identification of latent structures and trends in DSW generation. The choice of SEMMA and PCA responds not only to their methodological robustness, but also to their alignment with the strategic objectives of the study, which seeks to close the knowledge gaps on the relationship between the PCG index and the production of DSW. Previous research has pointed out the difficulty of transforming complex data into practical tools for waste management (3, 12). Minan is committed to the 2030 Agenda for Sustainable Development (16) and our research is specifically aligned with the thirteenth Sustainable Development Goal which consists of taking urgent action to deal with climate change and its impacts.

The rest of this study is organized as follows: Section 2 presents the various related works that precede this study. Section 3 explains the adaptation of the SEMMA methodology of data mining used for a more effective interpretation of domestic SW production data. For the evaluation stage, only the interpretation of the results was considered, clusters were generated by the Ward method and PCA was performed for a better observation of the totality of solid waste. Section 4 contrasts the discussion between the results obtained and other similar studies. Section 5 sets out the conclusions.

2 Related work

Zhong et al. (17) proposes a set of general cybersecurity event analysis models combined with cybersecurity visualization and the classic SEMMA analytical model to standardize the analysis process. Network asset segmentation, anomalous network event extraction, and attack event evolution are implemented through multi-view synergy. The results show the validity of the analytical model. In 2017, China decided to introduce new regulations for the categorization of urban SW. Through a text analysis and a Measured Policy Coherence (MPC) model, it was concluded that none of the 11 policies evaluated were of high or low quality, obtaining an average of 7.71 in the MPC index (18). Popli et al. (19) construct a prediction model of urban SW in Laos from 1995 to 2050 using multilinear regression (MLR) with six socioeconomic parameters. Models are created in four scenarios and evaluated for accuracy with R^2 for performance. The model includes population and GDP per capita, it is the most effective for estimating the rate of waste generation. Municipal solid waste is projected to reach 0.98 million tons in 2030, 1.26 million tons in 2040 and 1.52 million tons in 2050. Likewise, Karadimas et al. (20) used the Genetic Algorithm (GA) to identify optimal routes in the collection of Municipal Solid Waste (MSW). Considering that between 60 and 80% of waste management expenditure goes toward collection, a small improvement at this stage can lead to considerable savings.

Fontaine et al. (21) developed predictive models for end-of-life waste management of products, underlining the relevance of detailed prior analysis; a case study is presented in Gatineau. This model allows the environmental impact of waste management strategies to be assessed, the data revealed a true standard deviation 84% higher than expected, calculating measures such as asymmetry 0.072, which indicates a symmetrical distribution. Similarly, Rifaldi et al. (22) conducted a study in Bekasi Regency, using a qualitative descriptive approach and causal loop diagram (CLD) techniques to understand the waste management system. The results indicate that the pandemic affects variables such as consumption patterns, economic conditions and the volume of waste, factors such as the number of people, public welfare and health, public consumption, waste volume, landfill area, waste management costs, technology, economic conditions and government budgets were identified.

In their research, Requena-Sanchez et al. (12) used an online form to gather and generate information on the production and characteristics of domestic solid waste in Latin America. It was found that 43% of respondents reported an increase in the total volume of their waste. The COVID-19 pandemic has exacerbated the use and disposal of personal protective equipment (PPE), an analysis in Lima, Peru, revealed the presence of 138 items of PPE on 11 beaches over a 12-week period. Recreational beaches were the most affected, followed by surfing and fishing areas. This phenomenon could be associated with the high density of visitors in these places. Most PPE was discarded by beach visitors (23). Likewise, the Community Leadership technique in Urban Environmental Sanitation was used in a small community in Malawi to evaluate its technical, political and economic feasibility. The findings showed that although water quality and access were satisfactory, the amount of water available was insufficient and waste management needed improvement (24).

Maalouf Mavropoulos (25) define solid waste as the sum of non-liquid and non-gaseous materials that are discarded, such as packaging, paper, plastic, glass, metal and food scraps. In the same vein, MINAM (8) refers to the waste generated in tons by cleaning actions in commercial, institutional and public spaces. According to the Pan-American Center for Sanitary Engineering and Environmental Sciences (CEPIS), 70% of municipal solid waste comes from homes, while the remaining 30% belongs to non-domestic waste of the total municipal waste. SW management involves all actions from generation to final disposal of SW. MINAM (8) examines the generation of waste per capita in Peru. Patrao and Karnik (26) highlight the importance of recycling awareness and the negative effects of poor e-waste management. Mumtaz et al. (27) in Poland demonstrate that proper management of Municipal Solid Waste can have a significant economic impact, and suggest wet oxidation as an effective technique to manage large amounts of SR. The reuse of waste in geotechnical engineering is also proposed as a solution to the growing production of solid waste due to urbanization and population growth (28).

Fernández-Domínguez et al. (29) 28 substrate/digestate pairs were analyzed in terms of organic RS quality. Using principal component analysis and hierarchical clustering, the results indicated a 58.02% variability in accessibility and a decrease in accessibility between 16 and 66% and an increase in complexity between 34 and 98%. In addition, an increase in non-biodegradable compounds between 17 and 66% was observed. This information was used for a model that predicts organic carbon mineralization in soil. Kim and de Vasconcelos Barros (47) analyzed the effects of the National Solid Waste Policy (PNRS) on populous municipalities in Minas Gerais, Brazil, using PCA. Two components were found to account for more than 70% of the variability of the data. The components were related to the objectives and principles of the PNRS. Zhang et al. (48) applied a novel electronic nose (EN) to extract information from the headspace gas of municipal solid waste incineration leachate. Combining multiple machine learning and PCA techniques, the EN-based UMAP-XGBT model had the best classification performance, with an accuracy rate of 99.95% in the training set and 95.83% in the test set. Demirarslan and Yener (30) in Turkey, where rapid population growth and urbanization have led to increased environmental challenges. The research analyzed municipal solid waste (MSW) generation, composition, disposal methods, and per capita waste production across different regions. Findings revealed an average MSW generation of 26.74 Mton/yr., with regional variations strongly correlated to population size. The Marmara region produced the highest waste volume, followed by Central Anatolia, Aegean, Mediterranean, Black Sea, South East Anatolia, and Eastern Anatolia. Population demographics explained between 84 and 95% of MSW variation. Also Demirarslan and Çelik (31) presents the first solid waste characterization analysis for Artvin city center in Turkey's East Black Sea Region, examining waste composition across seven neighborhoods grouped by income levels (low, middle, and high). The research analyzed 11 waste categories, including organic, paper, plastic, glass, metal, ash, electronic, textile, garden, hazardous, and other materials. Using two-way ANOVA, researchers investigated seasonal variations in waste generation and its relationship with income levels. Results showed that organic waste dominated the annual composition at 61.06%, followed by paper (10.28%) and plastic (9%), with minimal electronic waste (0.037%). The study also evaluated three recyclable waste collection methods, concluding that building-based collection was the most economically viable option. The following questions were taken as questions for the study: What are the significant variations in the generation of household solid waste in the departments of Peru from 2014 to 2021? How does the population size influence the generation of household solid waste in the departments of Peru during this period? To what extent does the *Per Capita* Generation index influence the generation of household solid waste in the departments of Peru?

2.1 Hypotheses

There are significant spatial and temporal variations in household solid waste generation across the departments of Peru during the period 2014–2021. Population size significantly influences the generation of household solid waste in the departments of Peru during the period 2014–2021. The PCG index has a significant impact on the generation of household solid waste in the departments of Peru during the period 2014–2021.

2.2 Objective

To analyze the spatial and temporal variations in household solid waste generation across the departments of Peru during the period 2014–2021. To evaluate the extent to which population size influences solid waste generation in Peruvian departments. To investigate the impact of the PCG index on household solid waste generation in Peru.

3 Data mining methodology

The methodological framework of this research was structured around the SEMMA methodology because of its ability to guide data mining processes in clearly defined stages. This methodology was selected for its flexibility and applicability in the analysis of large and heterogeneous data, such as the records of the SIGERSOL system. By adopting SEMMA, we sought to ensure a systematic approach to data preprocessing, transformation and analysis, thus allowing us to extract meaningful patterns related to the generation of DSW. The application of Principal Component Analysis (PCA) responded to the need to reduce the dimensionality of the data set while preserving as much variability as possible. This technique is especially appropriate in contexts where multiple correlated variables can make direct interpretation of the results difficult.

3.1 Sample

This stage involves the selection of a representative subset of the study population to perform an exploratory analysis (32). The sample was obtained from the database from the annual reports submitted by the district and provincial municipalities to MINAM through the Information System for Solid Waste Management (SIGERSOL); available on the national open data platform, In compliance with D.S. 157-2021-PCM, it is available to civil society organizations, citizens, academia and the private sector. Each record, which represents a specific locality, was considered as a unit of analysis and contains information on the number of inhabitants and the generation of solid waste (SR) in that locality. The population of interest for this study was defined as all Peruvian households that produced solid waste between 2014 and 2021. To ensure the representativeness of the sample, data from all 24 departments of Peru were included. The following departments were included: Amazonas, Ancash, Apurímac, Arequipa, Ayacucho, Cajamarca, Cusco, Huancavelica, Huánuco, Ica, Junín, La Libertad, Lambayeque, Lima, Loreto, Madre de Dios, Moquegua, Pasco, Piura, Puno, San Martín, Tacna, Tumbes and Ucayali.

The annual reports submitted by district and provincial municipalities to MINAM through the SIGERSOL, Data preprocessing and cleaning procedures were executed through a three-step process to ensure data quality and analytical reliability. Starting from an initial dataset of 14,978 SIGERSOL records (2014–2021), the cleaning process involved specific exclusion criteria. First, 56 records from the Constitutional Province of Callao were eliminated due to their distinct administrative status, which could introduce biases in interdepartmental comparisons. Second, missing data were addressed by excluding 41 records that lacked information on waste generation, as these gaps would compromise analytical accuracy. Third, data validation identified and eliminated 29 records that contained logical inconsistencies, including negative values and implausible measurements that fell outside reasonable parameters. After implementing these cleaning protocols, the final data set comprised 14,852 valid records, which were subsequently organized into 196 record cards, with eight files assigned to each of the 24 departments. This structured approach to data cleaning and organization improved the integrity of the data set. While this choice reflects the limitations of the available data, it also ensures that the findings are based on information from provincial and district municipalities on solid waste generation per person per day.

3.2 Exploratory data analysis (EDA)

Figure 1 shows that the highest production of DSW is in Lima, exceeding 200,000 tons per year, while the lowest quantities are found in Huancavelica, Amazonas, Apurimac and Ayacucho; (b) the departments that had the highest number of registrations are Lima, Ancash and Cajamarca. Those with the lowest frequency are Pasco, Tacna, Moquegua, Ucayali, Tumbes and Madre de Dios (c) There are strong correlations between domestic SW, non-domestic SW and municipal SW; however, the correlation of these with the domestic PCG index is very weak.

Figure 2 shows a positive relationship between the average annual production of domestic SW and the size of the population of each department. The five departments with the highest annual production of domestic SW are Lima (13220.47 tons and 59813 inhabitants on average), Ucayali (5667.83 tons and 31756 inhabitants), Lambayeque (5616.48 tons and 33,710 inhabitants), Piura (4476.53 tons and 29579 inhabitants) and La Libertad

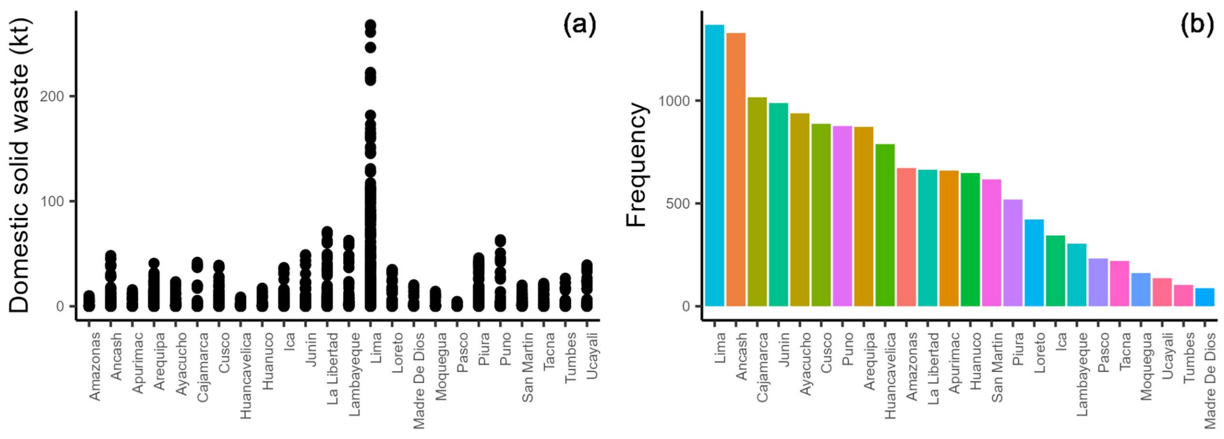

FIGURE 1
Characteristics of the SW in the period between 2014 and 2021: **(A)** Scatter graph of the number of Domestic SW by department; **(B)** Bar graph of the frequency of SW recording by department.

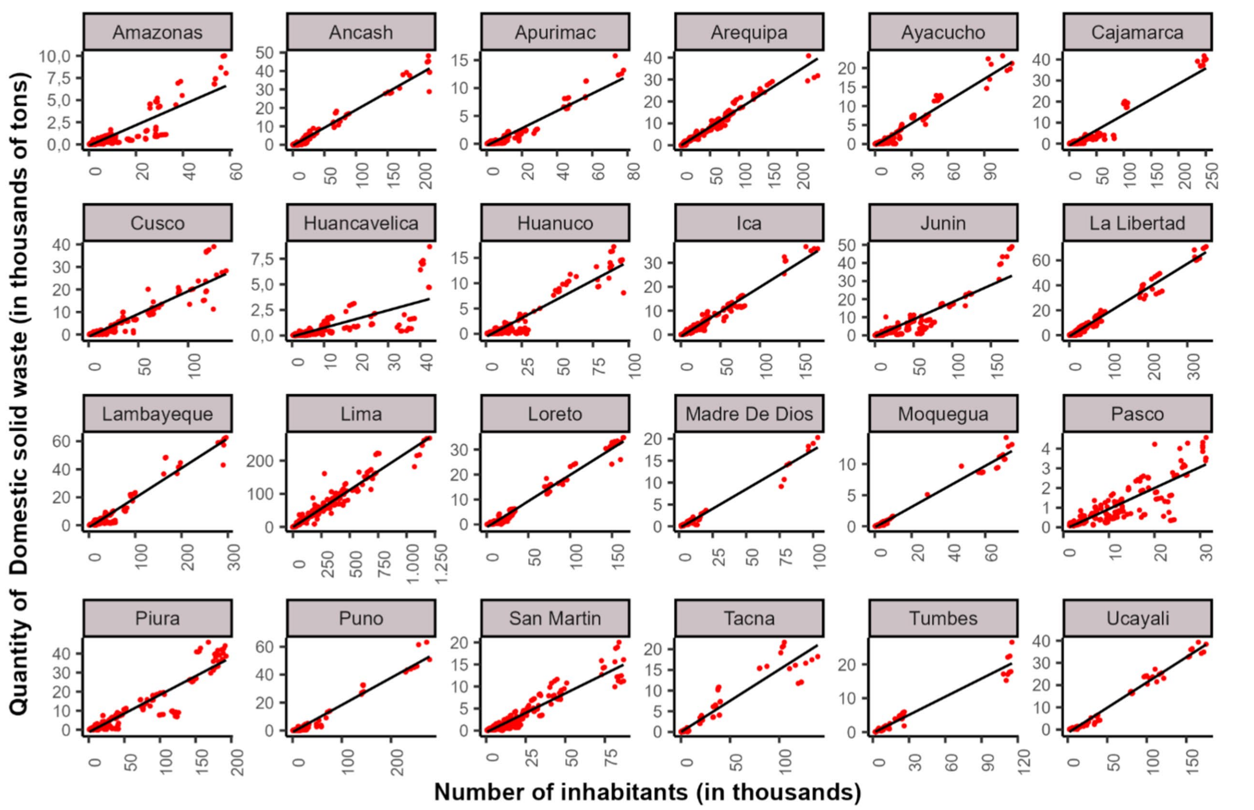

FIGURE 2
Scatter plot in quantity of domestic solid waste by number of inhabitants.

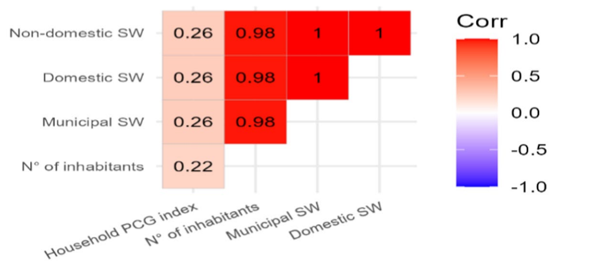

FIGURE 3
Correlation graph between SW variables. There are strong correlations between domestic SW, Non-domestic SW and municipal SW; however, the correlation of these with the Household PCG index is very weak.

(3478.46 tons and 23261 inhabitants). Figure 3 shows that household waste is strongly correlated with the number of inhabitants. On the other hand, the departments with the lowest annual production of domestic SW are Huancavelica (265.95 tons and 4,319 inhabitants on average), Amazonas (441.10 tons and 5,046 inhabitants), Apurimac (497.00 tons and 5,349 inhabitants) and Ayacucho (701.81 tons and 5,688 inhabitants).

Table 1 provides statistics on SW. The departments vary in population between 131 and 1,203,125 inhabitants, with an average of 16,540 and a median of 4,223 inhabitants. The *Per Capita* Generation of households varies between 0.01 and 2.00 kg/inhabitant/day, with a reasonably symmetrical distribution around the respective median and mean of these rates of 0.44 and 0.47 kg/inhabitant/day. The generation of domestic SW with an average of 2671.02 tons and a median of 207.98 tons, the production of domestic SW varies between 2.62 and 267875.78 tons, indicating a bias in favor of the areas with the highest production of SW. Similarly, non-domestic SW had an average of 1144.72 tons and a range of 1.12–114,803.91 tons, with 89.13 tons as the median. As for the municipal SW in total, the range goes from 3.80 to 382679.70 tons, with a median of 297.10 tons and an average of 3815.70 tons.

The group with the highest median in the PCG index per capita of Domestic SW is "46–50.99% kg/inhabitant/day," suggesting that this population group produces more Domestic SW. In contrast, the lowest median is detected in the "0–43.99% kg/inhabitant/day" group. In addition, less variability is perceived in the lower quartile of the latter group. The "46–50.99% kg/inhabitant/day" group also exhibits a balanced distribution and its interquartile range is higher compared to the other groups Figure 4.

A positive relationship between population size and the number of domestic SW is observed in Figure 5. The group with populations between 9,685,000 and 10,850,000 inhabitants show the highest median, while the group of 134,000–256,000 inhabitants exhibits the lowest median. In addition, the 9,685,000–10,850,000 group exhibits greater variability in quartiles and a wider interquartile range compared to other groups. A positive bias is observed in this population range, indicating that most of the inhabitants generate solid waste below the group average.

3.3 Modification

At this stage, the data was transformed in such a way that it facilitated the extraction of information. Out of a total of 14,852 registrations, 196 registration forms were organized. Seven groups were formed for the populations and four for the PCG index as shown in Table 2.

We transform data to improve visualization of meaningful comparisons. We use the PCA method to reduce the dimensionality of the data by changing the original set of 5 variables to a set (p) of 2 uncorrelated variables called principal components (q). The modified data show in Figure 5 a high positive correlation between the domiciliary SW, non-domiciliary SW and municipal SW and the number of inhabitants; it contributes greatly to Dim1. Lima is the department that has very high values in its production of domestic SW, having outliers. On the other hand, Piura, La Libertad, Arequipa and Lambayeque have high values in their production of domestic SW and lower values Madre de Dios, Huancavelica, Pasco and Moquegua. On the other hand, the departments with the highest percentage of PCG indexes are Ucayali, Loreto, San Martín, Lambayeque, Amazonas and Madre de Dios. The way in which data is being recorded may need to be reviewed as it contributes to Dim2 (Figure 6).

3.4 Modeling

This stage involves the application of various mathematical and statistical models to describe and predict patterns in the data that have been collected and prepared in the previous stages, for which the appropriate statistician for the data was selected and the assumptions for hypothesis testing were verified. Table 3 shows that all data have a normal distribution.

We verify the assumptions for the ANOVA: (a) the assumption that the average number of errors is zero was fulfilled. (b) The variance of errors is not equal between groups, since the Levene Test (p = 2e-16) and the Bartlett Test (p < 2.2e-16) show that this assumption is not true, evidencing heterogeneity of variances between the groups. The heterogeneity of variances detected by Levene's and Bartlett's tests justified careful consideration of the analytical approaches. While Welch's ANOVA was appropriately selected as a robust alternative to the traditional ANOVA, several additional statistical procedures could further strengthen the analysis. Box-Cox transformations could be applied to stabilize variances across groups (33), or weighted least squares regression could be implemented to account for unequal variances (34). For non-parametric alternatives, the Brown-Forsythe test could be considered, as it maintains robustness under heteroscedasticity while potentially offering greater power than Welch's ANOVA in certain conditions (35). (c) The errors are correlated, Residual autocorrelation was initially identified using the Durbin-Watson test, which yielded a statistic of DW = 0.848 with a p-value of 2.2e-16, indicating a significant positive autocorrelation in the residuals. This situation represented a critical violation of the assumption of independence of the errors, so the Welch ANOVA

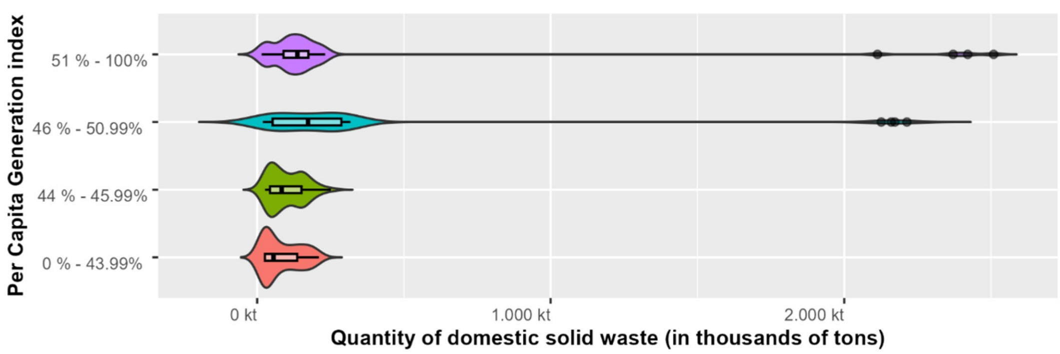

FIGURE 4
The *per capita* generation index of domestic solid waste.

TABLE 1 Solid waste management summary statistics.

	N° of inhabitants	PCG	Domestic SW (tons)	non-domestic SW (tons)	Municipal SW (tons)
Min.	131	0.01	2.62	1.12	3.80
1st Qu.	1,675	0.43	85.46	36.62	122.10
Median	4,223	0.44	207.98	89.13	297.10
Mean	16,540	0.47	2,671.02	1,144.72	3,815.70
3rd Qu.	10,919	0.51	767.23	328.81	1,096.00
Max.	1,203,125	2.00	267,875.78	114,803.91	382,679.70

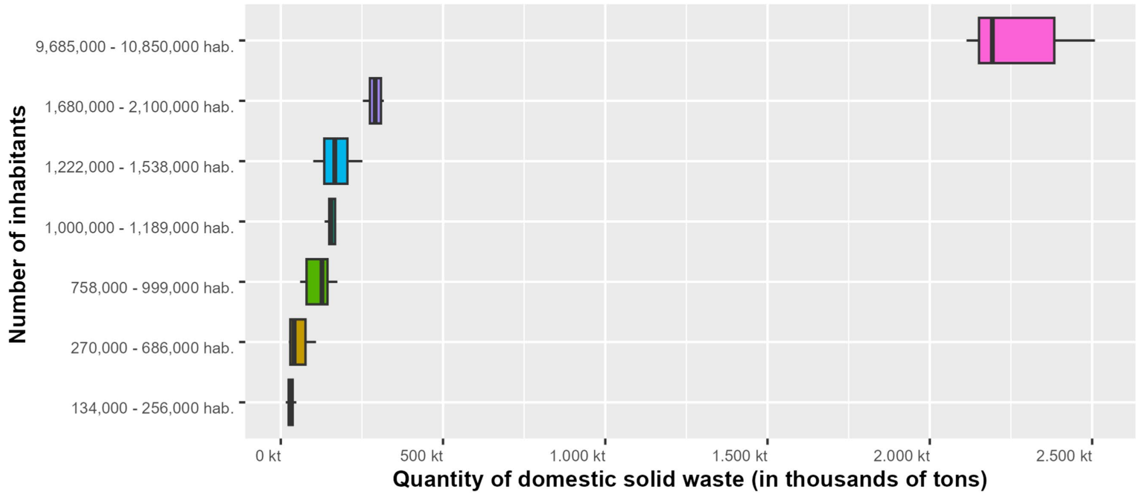

FIGURE 5
Graph of the box and whiskers, number of inhabitants and the amount of domestic SW.

analysis, known for its robustness to heteroscedasticity and certain forms of residual dependence, was chosen. (d) The errors are distributed according to a normal distribution, according to the results of the Shapiro–Wilk normality tests applied to the waste by department, where all the *p*-values are greater than the significance level of 0.05, meeting this assumption. Given the evidence of heterogeneity of variances between the groups and the presence of autocorrelation in the residuals, which violates two key assumptions for the traditional ANOVA, it was decided to perform the Welch ANOVA, which is more robust to violations of these assumptions instead of the classic ANOVA (36). It is an approach to dealing with heterogeneous variances in an analysis of variance over the means of the groups (37), p. 336. It is defined:

(1) K represents the number of groups or samples.
(2) W represents the weights associated with each group or sample.

(3) $$F = \frac{\frac{1}{k-1}\sum_{j=1}^{k} w_j \left(\overline{x}_j - \overline{x}'\right)^2}{1 + \frac{2(k-2)}{k^2-1}\sum_{j=1}^{k}\left(\frac{1}{n_j-1}\right)\left(1-\frac{w_j}{w}\right)^2}$$

(4) $$w_j = \frac{n_j}{s_j^2}$$

(5) $$w = \sum_{j=1}^{k} w_j \quad \overline{x}' = \frac{\sum_{j=1}^{k} w_j \overline{x}_j}{w}$$

(6) $$F \sim F\left(k-1, df\right)$$

(7) $$df = \frac{k^2-1}{3\sum_{j=1}^{k}\left(\frac{1}{n_j-1}\right)\left(1-\frac{w_j}{w}\right)^2}$$

Where F represents Welch's ANOVA test statistic, wj is the weights for each group, xj' is the sample mean of group j, x' is the overall weighted mean, nj is the sample size of group j, sj^2 is the sample variance of group j, and df is the degrees of freedom for the approximate F distribution.

To look for patterns and relationships in the data, the Cluster Average and Cluster Ward.D2 modeling techniques were used to identify groups of households that present similar patterns of DSW generation depending on the department where they occurred Figure 7. Cluster Average Method: This method calculates the average distance between each pair of objects in different groups. Cluster Ward.D2 Method seeks to minimize the sum of squared differences within each group.

3.5 Evaluation

The results of the statistical analysis support the hypothesis that the amount of population significantly influences the generation of domestic solid waste in the departments of Peru during the period 2014–2021. Welch's ANOVA [$F(6, 94.310) = 790.444$; $p = 0.0$, $p < 0.05$] and a squared Eta of 0.9909 reveal a very large effect of the Population Quantity on the generation of domestic SW between departments. *Post hoc* analysis using the Games-Howell test showed statistically relevant differences ($p < 0.05$) in the volumes of domestic SW produced among most of the pairs of population groups compared. However, a notable exception was found between the groups "1,000,000–1,189,000 inhabitants" and "1,222,000–1,538,000 inhabitants," where the comparison resulted in a *p*-value of 0.564 ($p > 0.05$), suggesting that there are no statistically relevant differences in the amount of domestic SW produced between these two strata. Student's *t*-test for independent samples revealed a statistically significant difference [$t(38) = -4.0833$; $p = 0.0001102$, $p < 0.05$] in solid waste generation between the groups "758,000–999,000 inhabitants" and "1,000,000–1,189,000 inhabitants." The confidence interval [-Inf, −23748.71] indicates that the "758,000–999,000 inhabitants" group produces less domestic solid waste than the

TABLE 2 Levels and summary statistics of the population and PCG.

Population groups	PCG group	*n*	Average	DE
134,000–256,000 hab.	0–43.99%	6	25,135.01	2,022.73
134,000–256,000 hab.	44–45.99%	6	34,037.52	3,817.47
134,000–256,000 hab.	46–50.99%	7	36,443.78	10,121.25
134,000–256,000 hab.	51–100.00%	5	21,831.59	6,752.18
270,000–686,000 hab.	0–43.99%	25	44,479.63	24,170.17
270,000–686,000 hab.	44–45.99%	12	48,863.24	16,127.07
270,000–686,000 hab.	46–50.99%	6	41,477.51	9,278.57
270,000–686,000 hab.	51–100.00%	13	73,287.75	31,087.42
758,000–999,000 hab.	44–45.99%	10	89,838.42	33,187.30
758,000–999,000 hab.	46–50.99%	4	131,059.28	47,258.41
758,000–999,000 hab.	51–100.00%	11	133,864.54	23,105.03
1,000,000–1,189,000 hab.	0–43.99%	2	150,650.96	6,447.90
1,000,000–1,189,000 hab.	44–45.99%	6	157,749.80	14,254.05
1,000,000–1,189,000 hab.	51–100.00%	7	156,576.32	10,120.70
1,222,000–1,538,000 hab.	0–43.99%	15	160,846.99	32,073.34
1,222,000–1,538,000 hab.	44–45.99%	14	154,336.29	47,829.82
1,222,000–1,538,000 hab.	46–50.99%	11	166,091.33	53,913.85
1,222,000–1,538,000 hab.	51–100.00%	8	213,426.19	15,427.38
1,680,000–2,100,000 hab.	46–50.99%	16	289,563.54	20,695.12
9,685,000–10,850,000 hab.	46–50.99%	4	2,168,068.49	36,029.00
9,685,000–10,850,000 hab.	51–100.00%	4	2,353,332.45	169,895.15

"1,000,000–1,189,000 inhabitants" group, with a size of Cohen's d-effect = −1.33359, suggesting a large effect. The Student's *t*-test for independent samples showed a statistically significant difference [$t(21) = -54.919$; $p = 2.2e\text{-}16$, $p < 0.05$] in solid waste generation between the groups "1,000,000–1,189,000 inhabitants" and "9,685,000–10,850,000 inhabitants." The confidence interval [-Inf, −2,038,507] indicates that the "1,000,000–1,189,000 inhabitants" group produces less domestic solid waste than the "9,685,000–10,850,000 inhabitants" group, with a size of Cohen's d-effect = −24.04328, This indicates a large effect.

On the other hand, the results support the hypothesis that the PCG index per capita influences the generation of domestic solid waste in the departments of Peru during the period 2014–2021. Welch's analysis of variance [F (3, 94.310) = 5.359; $p = 0.02$, $p < 0.05$] and a squared Eta of 0.0705 reveal the existence of a significant influence of the PCG on the production of domestic solid waste, explaining 7.05% of the total variance in production levels between departments, suggesting a close to medium or moderate effect on the volumes of household waste generation. The Games-Howell *post hoc* analysis identified statistically significant differences in solid waste volumes, specifically when comparing the "0–43.99%" and "45.98–50.53%" groups ($p = 0.017$), and between the "44–45.99%" and "46.0–50.99%" groups ($p = 0.03$). Student's *t*-test for independent samples revealed a statistically significant difference [$t(94) = -3.0795$; $p = 0.001359$, $p < 0.05$] in solid waste generation between the "0–43.99%" and "46.00–50.99%" groups. The confidence interval [−Inf, −116898.7] indicates that the "0–43.99%" group produces less domestic solid waste than the "46.00–50.99%" group, with a Cohen d-effect size = −0.6285902, suggesting a mean effect. Student's *t*-test for independent samples between the groups "44–45.99%" and "46–50.99%" [$t(94) = -2.8744$; $p = 0.002503$, $p < 0.05$] revealed a statistically significant difference. The confidence interval [-Inf, −99927.99] indicates that the "44–45.99%" group produces less domestic solid waste than the "46–50.99%" group, with a Cohen d-effect size = −0.586743, suggesting a mean effect (Figure 8).

The PCA revealed substantial insights into the relationship between domestic solid waste generation and demographic patterns across Peruvian departments. We form three groups: In Group (a) the first component explains 81.5% of the inertia and the first two components represent 99.4% of the total variance. Lambayeque (5,616.48 tons), Loreto (2946.44 tons) and San Martín (1596.07 tons) had the highest DSW averages, while Amazonas (441.1 tons) had the lowest. Ucayali (60%), Loreto (58%) and San Martín (57%) registered the highest PCG indices. Group (b) the corresponding figures are 79.2 and 95.6%. Ayacucho (701.81 tons) showed the highest average of DSW, and Apurimac (497 tons) the lowest. Tacna and Apurimac (44%) and Moquegua (43%) exhibited the highest PCG index, while Huancavelica (42%) and Pasco (41%) the lowest. Group (c) the first component collects 77.2% of the inertia and the first two 96.4% of the variance. Piura (4,476.53 tons) and La Libertad (3,478.46 tons) presented the highest averages of domestic SW, and Huánuco (859.41 tons) and Cajamarca (812.74 tons) the lowest. Ica and Piura led with an average PCG of 48%, while Puno and Junín (43%) registered the lowest values. In all cases, with only two dimensions, most of the essential information is preserved and a positive correlation is observed between the types of domestic SW and the number of inhabitants. The PCA visualization clearly identified Lima as an outlier with exceptionally high DSW production, while departments like Piura, La Libertad, Arequipa, and Lambayeque formed a distinct cluster with elevated waste

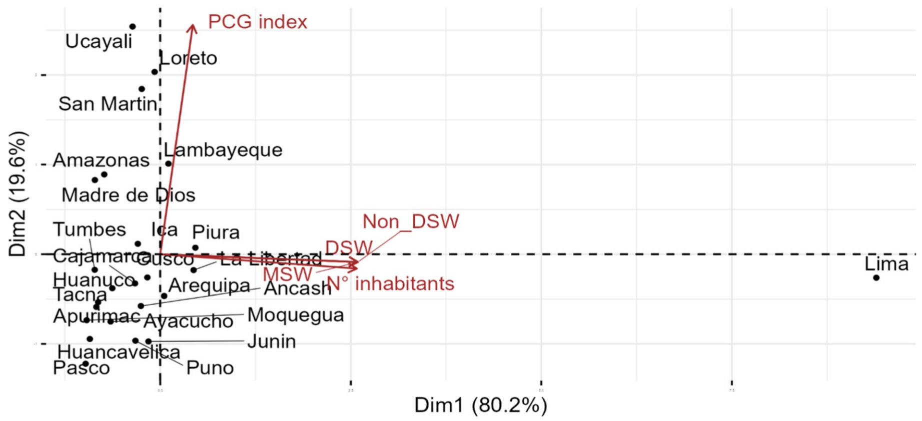

FIGURE 6
PCA graph shows the solid waste of the departments of Peru, Non-Domestic Solid Waste (Non_DSW), Domestic Solid Waste (DSW), Municipal Solid waste (MSW), Number of Inhabitants (N° Inhabitants). The first component explains 80.2% of the inertia, a significant percentage that shows that this main component collects much of the essential information in the data. The first two components in total account for 99.8% of the total variance, indicating that most of the relevant information can be preserved with only two dimensions.

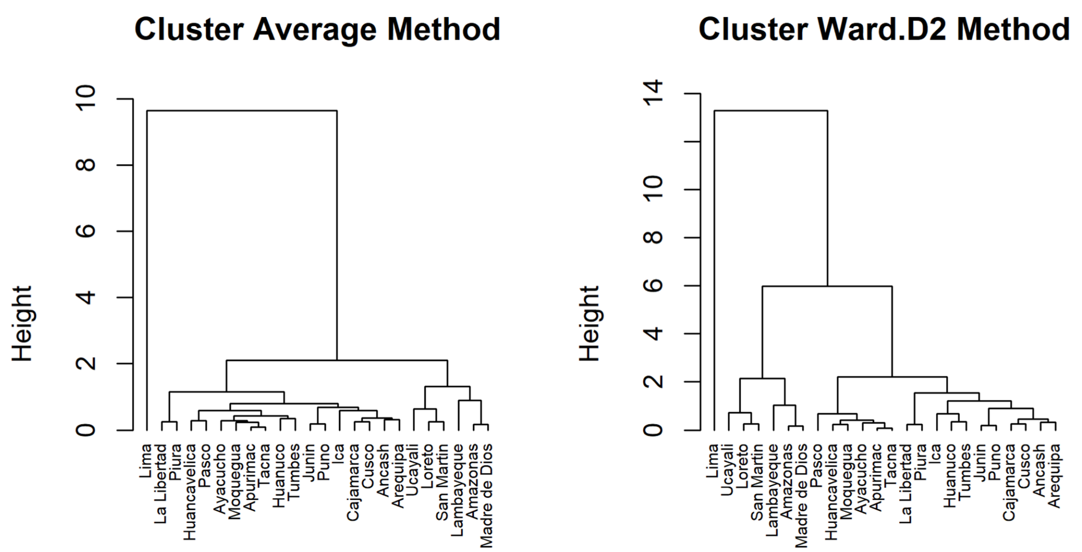

FIGURE 7
Dendrogram of the clusters.

TABLE 3 Normality tests—Shapiro–Wilk for home SW quantity.

Department	w	p-value	Department	w	p-value	Department	w	p-value
Amazonas	0.89	0.26	Huánuco	0.85	0.09	Moquegua	0.91	0.33
Ancash	0.92	0.41	Ica	0.86	0.13	Pasco	0.90	0.26
Apurímac	0.95	0.67	Junín	0.92	0.42	Piura	0.92	0.45
Arequipa	0.89	0.22	La Libertad	0.89	0.25	Puno	0.95	0.74
Ayacucho	0.86	0.12	Lambayeque	0.93	0.51	San Martín	0.89	0.25
Cajamarca	0.84	0.08	Lima	0.87	0.14	Tacna	0.93	0.49
Cusco	0.98	0.98	Loreto	0.87	0.14	Tumbes	0.88	0.19
Huancavelica	0.84	0.08	Madre de Dios	0.89	0.22	Ucayali	0.92	0.39

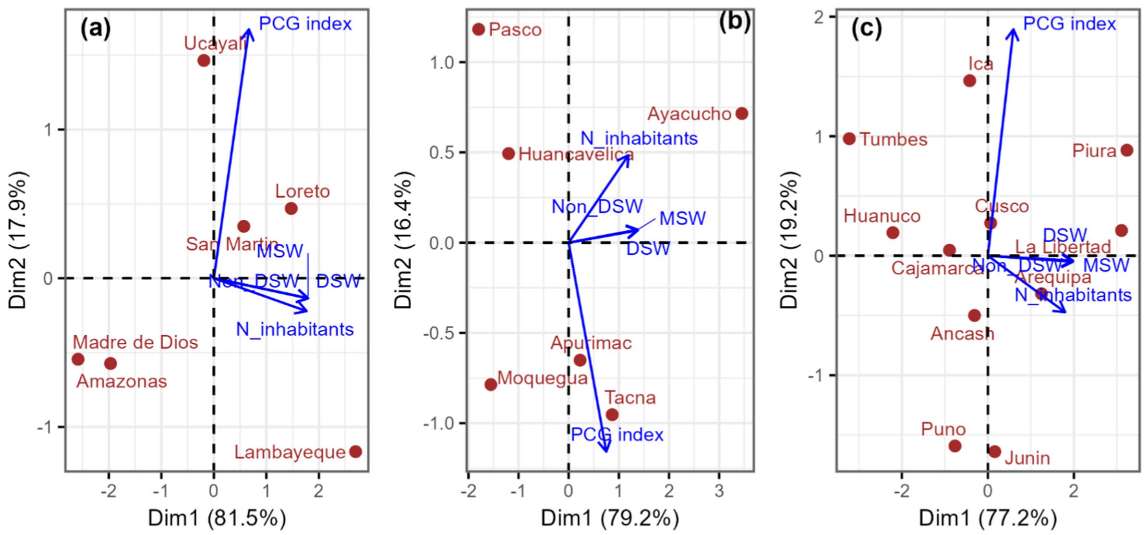

FIGURE 8
PCA of the Cluster Ward.D2 method of the departments **(A)** Ucayali, Loreto, San Martín, Lambayeque, Amazonas and Madre de Dios; **(B)** Pasco, Huancavelica, Moquegua, Ayacucho, Apurímac and Tacna; **(C)** Ancash, La Libertad, Piura, Ica, Huánuco, Puno, Junín, Cajamarca, Cusco, Tumbes and Arequipa.

generation levels. Notably, the analysis also revealed that departments such as Ucayali, Loreto, and San Martín exhibited higher PCG index, suggesting that factors beyond population size, such as regional waste management practices and socioeconomic conditions, may influence waste generation patterns in these areas.

4 Discussion

The analysis of the data collected and the use of statistical techniques corroborated that the number of population significantly influences the generation of DSW in the departments of Peru. Welch's ANOVA and the square Eta provided solid evidence of a very large effect of the amount of population on the generation of DSW between departments. These findings are consistent with existing literature suggesting a direct relationship between population size and solid waste generation. As the population increases, so does the demand for goods and services, which in turn leads to an increase in waste production (38). *Post hoc* analysis using the Games-Howell test revealed statistically significant differences in the volumes of domestic SR produced among most of the pairs of population groups compared. However, there was a notable exception between the "1,000,000–1,189,000 inhabitants" and "1,222,000–1,538,000 inhabitants" groups, where no statistically significant differences were found in the amount of domestic SW produced. This may be due to factors not considered in this study, such as variability in waste management practices or differences in consumption levels. In harmony Quispe et al. (39) in the district of Santa Rosa, Ayacucho carried out an analysis showed significant differences between the SW. A high correlation between zones A and B indicates that organic waste exceeds inorganic waste.

On the other hand, Welch's analysis of variance, Games-Howell's *post hoc* analysis, and Student's *t*-tests have provided solid statistical evidence that the PCG index per capita has a significant influence on the generation of domestic SW. The fact that the PCG accounts for 7.05% of the total variance in solid waste production levels between departments suggests that, although there are other factors that also influence waste generation, the PCG per capita has a moderate effect on household waste generation volumes. This statement coincides with what Lo Storto (40) found that the indicator was relative to the impact on eco-efficiency of variables that measure the size or density of DSW services, the indicators remain inconclusive.

Statistically significant differences in solid waste volumes were observed between the different groups of PCG. Groups with a higher PCG ("45.98–50.53%" and "46.0–50.99%") generated more solid waste than groups with a lower PCG ("0–43.99%" and "44–45.99%"). Accordingly, Li et al. (2) the amount of waste is expected to increase in the coming years. The results of the study corroborate this theory, as they show a positive correlation between the different types of solid waste domestic SW, non-domestic SW and municipal SW and the number of inhabitants in the departments of Peru.

The PCA analysis and Ward's minimum variance method revealed that Lima, being the department with the largest number of inhabitants, is also the one with the highest values in the production of domestic SW, constituting a group in itself. This observation suggests that solid waste generation is strongly influenced by population density. In agreement Cerna-Cueva et al. (41) due to the increase in population and consumption habits, the management of SW are environmental challenges in the world. In Group (a) the first component explained 81.5% of the inertia and the first two 99.4% of the variance. Lambayeque 5,616.48 tons, Loreto 2,946.44 and San Martín 1,596.07 had the highest DSW averages, while Amazonas 441.1 the lowest. Ucayali 60%, Loreto 58% and San Martín 57% registered the highest PCG indices. In Group (b) Ayacucho 701.81 tons showed the highest average DSW and Apurimac 497 the lowest. Tacna, Apurimac 44% and Moquegua 43% exhibited the highest PCG indices, while Huancavelica 42% and Pasco 41% the lowest. In Group (c) Piura 4,476.53 tons and La Libertad 3,478.46 with higher DSW averages, Huánuco 859.41 and Cajamarca 812.74 the lowest. Ica and

Piura led with PCG of 48%, while Puno and Junín 43% were the lowest. A positive correlation was observed between DSW and number of inhabitants. These findings are consistent with those reported by Fernández-Domínguez et al. (29), who used PCA and hierarchical clustering analysis to characterize the quality of organic waste and found a 58.02% variability in accessibility, as well as significant changes in complexity and non-biodegradable compounds after treatment. Likewise, Kim and de Vasconcelos Barros (42) applied the PCA to evaluate the effects of the National Solid Waste Policy in municipalities of Minas Gerais, Brazil, and managed to represent more than 70% of the variability in the data. In addition, the study by Zhang et al. (43) demonstrates the usefulness of PCA in combination with machine learning techniques for the classification of leachate from municipal solid waste incineration, achieving accuracy rates of over 95% in training and test sets.

DSW generation patterns observed in Peru show notable similarities and differences with international studies (44). The high concentration of waste generation in Lima (13,220.47 tons) reflects a phenomenon documented in other Latin American megacities. Like Mexico City in 2020 (45), DSW is an issue of great relevance due to its environmental and social impact. According to recent data, the city generates approximately 13,149 tons of solid waste daily. This average translates to around 1.40 kg per inhabitant per day. This concentration pattern in major metropolises is consistent with trends observed in developing countries. Filipenco (46) menciona que la producción anual de DSW es más de 2.010 millones de toneladas de basura sólida urbana en todo el mundo. Si las tendencias actuales continúan, se espera que esta cifra aumente drásticamente y posiblemente alcance los 3.800 millones de toneladas en 2050.

5 Conclusion

Through the application of statistical techniques and the adaptation of the SEMMA data mining methodology and Principal Component Analysis (PCA) (47), it has been possible to obtain a more detailed and accurate view of the generation of domestic solid waste in Peru. This study has revealed the significant influences of both the amount of population and the PCG index per capita with the generation of domestic SW. the existence of a high positive correlation between domestic SW, non-domiciliary SW and municipal SW.

The *t*-tests give evidence of the effect of population size on solid waste generation (48). Groups with larger populations produced significantly more residues than groups with smaller populations, which is reflected in the large effect sizes found.

The PCG per capita has a moderate effect, which gives rise to further studies to further explore the underlying causes of these regional differences and evaluate the effectiveness of the waste management measures implemented to date (49).

The study indicates a positive correlation between the generation of residential solid waste (SW) and the number of inhabitants in the departments of Peru, which underlines the impact of the demographic volume on the production of domestic SW. In particular, Lima, with an average of 13220.47 tons, 59,813 residents and a PCG coefficient of 50%, is distinguished by its high production of residential SW, representing a unique case. PCA was performed for three groups as a result of grouping them by Ward's minimum variance method, the results are detailed: In group (a) Lambayeque (5616.48 tons), Loreto (2946.44 tons) and Madre de Dios (2091.20 tons), in these the highest average in the production of domestic Solid Waste is shown. On the other hand, Amazonas (441.1 tons) and San Martin (1596.07 tons) exhibit the lowest averages. With regard to the per capita PCG index on average, Ucayali 60%, Loreto 58% and San Martín 57% have the highest values, compared to Madre de Dios 51%, Amazonas 52% and Lambayeque 53% have the lowest values. In group (b) Ayacucho (701.81 tons) presents the highest average of Domestic Solid Waste (DSW), while Apurimac (497 tons) registers the lowest number. In relation to the PCG indicator, the departments of Tacna with 44%, Apurimac with 44% and Moquegua with 43% exhibit the highest average percentages, in contrast to Huancavelica with 42% and Pasco with 41%, which display the lowest. In group (c), the departments with the highest average values of domestic solid waste are Piura (4476.53 tons) and La Libertad (3478.46 tons), while those with the lowest volume are Huánuco (859.41 tons) and Cajamarca (812.74 tons). With regard to the PCG, Ica and Piura lead with an average of 48%, as opposed to Puno and Junín, which have the lowest values with 43%.

The presence of Lima as a statistical outlier significantly influenced the analytical outcomes and subsequent interpretations. With an average DSW production of 13220.47 tons and a population of 59,813 inhabitants, Lima's values deviated substantially from other departments, necessitating careful consideration in the analysis. This anomaly's impact manifested in several key ways: Lima's exceptional DSW production values necessitated the implementation of Ward's hierarchical clustering method, resulting in three distinct groups and requiring separate PCA for each cluster. This methodological adaptation was crucial for maintaining the analytical integrity and preventing Lima's outlier status from distorting the overall patterns in the dataset. Lima's extreme values significantly influenced the dimensionality of the PCA, particularly in Dimension 1, where it emerged as the primary outlier. This affected the spatial distribution of other departments in the analysis, creating a distinct separation between high-production departments (Piura, La Libertad, Arequipa, and Lambayeque) and low-production regions (Madre de Dios, Huancavelica, Pasco, and Moquegua). Lima's unique position in the dataset highlighted important demographic-waste generation relationships. Despite having a moderate PCG index of 50%, Lima's total waste production far exceeded other departments. This finding particularly contrasts with departments like Ucayali, Loreto, and San Martín, which displayed higher PCG indices despite lower total waste production.

The study employs multiple statistical techniques Welch's ANOVA, PCA, Ward's method that complement each other, providing a comprehensive analytical framework. The use of SEMMA methodology adds structure to the data mining process. The high variance explanation 99.8% by the first two components of PCA demonstrates strong dimensional reduction effectiveness. The analysis provides robust statistical evidence of the influence of population on waste generation (Eta squared = 99.09%).

The extensive dataset (14,852 records from 196 registration forms) provides good coverage of all 24 departments of Peru. The study effectively captures regional variations in waste generation patterns. The study does not fully explore the underlying causes of the non-significant differences between population groups

(1,000,000-1,189,000 and 1,222,000–1,538,000 inhabitants). The moderate effect of the PCG index (7.05% variance explanation) suggests unexplored factors affecting waste generation, providing valuable insights into Peru's household solid waste generation patterns, particularly through its robust statistical approach. However, expanding the analysis to include more contextual factors and predictors would improve its practical utility for waste management planning and policies.

6 Recommendations

Domestic solid waste generation in Peru is highly correlated with population size, with Lima standing out as the largest daily generator of domestic solid waste, which requires specific management protocols. In regions such as Ucayali and Loreto, with high per capita generation rates of 60 and 58%, respectively, it is recommended to prioritize waste reduction programs. Analysis using Ward's method identifies distinct regional patterns, highlighting the need for management policies differentiated by region. Furthermore, although the per capita generation rate moderately explains the variability 7.05%, it is crucial to investigate other factors that influence waste generation. Finally, in the face of projected waste growth, long-term infrastructure planning that integrates reduction and recycling strategies is urged to promote sustainable management.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

JP-V: Investigation, Methodology, Writing – original draft. JS: Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2025.1498513/full#supplementary-material

References

1. Abushammala H, Ghulam ST. Impact of residents' demographics on their knowledge, attitudes, and practices towards waste Management at the Household Level in the United Arab Emirates. *Sustainability*. (2023) 15:685. doi: 10.3390/su15010685

2. Li CH, Lee TT, Lau SSY. Enhancement of municipal solid waste Management in Hong Kong through innovative solutions: a review. *Sustainability*. (2023) 15:310. doi: 10.3390/su15043310

3. Li W, Wang J. Feasibility study of the unit pricing system for household wastes sorting in the context of China. *Waste Manag Res*. (2022) 40:586–95. doi: 10.1177/0734242X20988783

4. Cheng KM, Tan JY, Wong SY, Koo AC, Sharji EA. A review of future household waste management for sustainable environment in Malaysian cities. *Sustainability*. (2022) 14:6517. doi: 10.3390/su14116517

5. Pillay R, Hansraj R, Rampersad N. Disposal of spectacles and contact lenses: optometrist and lens wearer perspectives. *Afr Vision Eye Health*. (2023) 82:784. doi: 10.4102/aveh.v82i1.784

6. Begum A, Liu J, Qayum H, Mamdouh A. Environmental and moral education for effective environmentalism: an ideological and philosophical approach. *Int J Environ Res Public Health*. (2022) 19:5549. doi: 10.3390/ijerph192315549

7. Bonnett M. Environmental consciousness, nature and the philosophy of education: matters arising. *Environ Educ Res Scopus*. (2023) 29:1377–85. doi: 10.1080/13504622.2023.2225807

8. MINAM. (2023). Ministerio del Ambiente. Available at: https://www.datosabiertos.gob.pe/group/ministerio-del-ambiente (Accessed January 17, 2015).

9. Peruano El. (2024). Situación actual de los residuos sólidos. Available at: https://elperuano.pe/noticia/216136-situacion-actual-de-los-residuos-solidos (Accessed January 17, 2015).

10. Gestión N. (2023). Residuos en Cañete | Minam: San Vicente de Cañete en emergencia por gestión y manejo de residuos sólidos | Gobierno | San Vicente de Cañete | Minam | PERU. Gestión; NOTICIAS GESTIÓN. Available at: https://gestion.pe/peru/residuos-en-canete-minam-san-vicente-de-canete-en-emergencia-por-gestion-y-manejo-de-residuos-solidos-gobierno-san-vicente-de-canete-minam-noticia/ (Accessed January 17, 2015).

11. Nayak SP, Uppal SB. Preliminary assessment of COVID-19 waste management scenario during lockdown in Chandigarh & nearby areas. *Scopus*. (2023) 1110:012067. doi: 10.1088/1755-1315/1110/1/012067

12. Requena-Sanchez N, Carbonel D, Moonsammy S, Demel L, Vallester E, Velasquez D, et al. COVID-19 impacts on household solid waste generation in six Latin American countries: a participatory approach. *Environ Monit Assess*. (2023a) 195:155. doi: 10.1007/s10661-022-10771-9

13. Zulkipli F, Nopiah ZM, Jamian NH, Basri NEA, Kie CJ. Sustainable public awareness on solid waste management and environmental care using logistics regression. *J Kejuruteraan*. (2022) si5:139–48. doi: 10.17576/jkukm-2022-si5(2)-15

14. Tapia JAC. Modelo de Aprendizaje Profundo para identificar plagas en la producción de quinua. *Rev Ingeniería*. (2024) 8:20. doi: 10.33996/revistaingenieria.v8i20.116

15. Martínez C. (2018). RPubs—Análisis de componentes principales (PCA). Available at: https://rpubs.com/Cristina_Gil/PCA (Accessed January 17, 2015).

16. SIAR. (2024). MINAM contribuye al conocimiento y puesta en marcha de la Agenda 2030 a través de nueva publicación sobre los Objetivos de Desarrollo Sostenible | SIAR Cajamarca | Sistema de Información Ambiental Regional de Cajamarca. Available at: https://siar.regioncajamarca.gob.pe/novedades/minam-contribuye-conocimiento-puesta-marcha-agenda-2030-traves-nueva (Accessed January 17, 2015).

17. Zhong Y, Wang S, Wu H, Cheng Z, Li X. SEMMA-based visual exploration of cyber security event. *J Zhejiang Univ Sci*. (2022) 49:131–40. doi: 10.3785/j.issn.1008-9497.2022.02.001

18. Xu J, Zhang Z, Xu Y, Liu L, Pei T. Quantitative evaluation of waste sorting management policies in China's major cities based on the PMC index model. *Front Environ Sci*. (2023) 11:1065900. doi: 10.3389/fenvs.2023.1065900

19. Popli K, Park C, Han S-M, Kim S. Prediction of solid waste generation rates in urban region of Laos using socio-demographic and economic parameters with a multi linear regression approach. *Sustainability*. (2021) 13:3038. doi: 10.3390/su13063038

20. Karadimas NV, Papatzelou K, Loumos VG. Genetic algorithms for municipal solid waste collection and routing optimization. *IFIP*. (2007) 247:223–31. doi: 10.1007/978-0-387-74161-1_24

21. Fontaine L, Legros R, Frayret J-M. Solid waste generation prediction model framework using socioeconomic and demographic factors with real-time MSW collection data. *ISWA*. (2024) 26:734242X241231414. doi: 10.1177/0734242X241231414

22. Rifaldi M, Zid M, Sumargo B. Causal loop diagram as approaching model analysis in increasing the waste volume at the COVID19 pandemic period. *J Kesehatan Lingkungan*. (2021) 13:180–5. doi: 10.20473/jkl.v13i3.2021.180-185

23. De-la-Torre GE, Rakib MRJ, Pizarro-Ortega CI, Dioses-Salinas DC. Occurrence of personal protective equipment (PPE) associated with the COVID-19 pandemic along the coast of Lima, Peru. *Sci Total Environ*. (2021) 774:145774. doi: 10.1016/j.scitotenv.2021.145774

24. Mtika WM, Tilley E. Environmental sanitation planning: feasibility of the CLUES framework in a Malawian Small town. *Front Environ Sci*. (2020) 7:204. doi: 10.3389/fenvs.2019.00204

25. Maalouf A, Mavropoulos A. Re-assessing global municipal solid waste generation. *Waste Manag Res*. (2023) 41:936–47. doi: 10.1177/0734242X221074116

26. Patrao G, Karnik A. Identifying drivers and hindrances to the disposal of used Mobile phones: a study of user behavior in the UAE. *SAGE Open*. (2023) 13:6757. doi: 10.1177/21582440231196757

27. Mumtaz H, Werle S, Sobek S. A waste wet oxidation technique as a solution for chemical production and resource recovery in Poland. *Clean Technol Environ Policy*. (2023) 26:1363–82. doi: 10.1007/s10098-023-02520-4

28. Liu C-H, Hung C. Reutilization of solid wastes to improve the hydromechanical and mechanical behaviors of soils—a state-of-the-art review. *Sust Environ Res*. (2023) 33:1–21. doi: 10.1186/s42834-023-00179-6

29. Fernández-Domínguez D, Patureau D, Houot S, Sertillanges N, Zennaro B, Jimenez J. Prediction of organic matter accessibility and complexity in anaerobic digestates. *Waste Manag*. (2021) 136:132–42. doi: 10.1016/j.wasman.2021.10.004

30. Demirarslan KO, Yener I. Municipal solid waste in Turkey and its relationship with the population. *Res Ecol*. (2021) 3:31–40. doi: 10.30564/re.v3i1.2895

31. Demirarslan KO, Çelik BY. Urban solid waste characterization in the east part of Black Sea region. *Global J Environ Sci Manage*. (2018) 4:167–82. doi: 10.22034/gjesm.2018.04.02.005

32. Shin SS, Park SD, Kim G. Applicability comparison of GIS-based RUSLE and SEMMA for risk assessment of soil Erosion in wildfire watersheds. *Remote Sens*. (2024) 16:50932. doi: 10.3390/rs16050932

33. Box GEP, Cox DR. An analysis of transformations. *J Royal Stat Soc Series B*. (1964) 26:211–43. doi: 10.1111/j.2517-6161.1964.tb00553.x

34. Carroll RJ, Ruppert D. Transformation and weighting in regression. New York, NY: Chapman and Hall/CRC (2017).

35. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc*. (1974) 69:364–7. doi: 10.1080/01621459.1974.10482955

36. Zaiontz C. (2024). Prueba de Welch | Estadísticas reales usando Excel. Available at: https://real-statistics.com/one-way-analysis-of-variance-anova/welchs-procedure/ (Accessed January 17, 2015).

37. Howell D. C. (2010). Statistical methods for psychology (seventh edition [7 ed.]). Available at: www.cengage.com/permissions (Accessed January 17, 2015).

38. Bhada-Tata, D. (2012). *Open Knowledge Repository*. Available at: https://openknowledge.worldbank.org/entities/publication/1a464650-9d7a-58bb-b0ea-33ac4cd1f73c

39. Quispe K, Martínez M, da Costa K, Romero Giron H. Solid waste Management in Peru's cities: a clustering approach for an Andean district. *Appl Sci*. (2023) 13:646. doi: 10.3390/app13031646

40. Lo Storto C. Measuring ecoefficiency of municipal solid waste management in Apulia to account for governance heterogeneities. *Cleaner Waste Syst*. (2024) 7:100131 doi: 10.1016/j.clwas.2024.100131

41. Cerna-Cueva AF, Rosas-Echevarría C, Perales-Flores R, Ataucusi-Flores PL, Cerna-Cueva AF, Rosas-Echevarría C, et al. Predicción de la generación de residuos sólidos domiciliares con machine learning en una zona rural de Puno. *Tecnia*. (2022) 32:44–52. doi: 10.21754/tecnia.v32i1.1378

42. Kim VJH, de Vasconcelos Barros, R T. (2023). National Solid Waste Policy effects on urban solid waste management in the most populated municipalities ofMinas Gerais: A multivariate analysis. *Urbe*, 15. Scopus. doi: 10.1590/2175-3369.015.e20230017

43. Zhang Z, Qiu S, Zhou J, Huang J. (2022). Monitoring of MSW Incinerator Leachate Using Electronic Nose Combined with Manifold Learning and Ensemble Methods. *Chemosensors*, 10(12). Scopus. doi: 10.3390/chemosensors10120506

44. México (2020). INVENTARIO de RESIDUOS SÓLIDOS de la ciudad de MÉXICO 2020. Available at: https://biblioteca.semarnat.gob.mx/janium/Documentos/Cecadesu/Libros/CD007179.pdf (Accessed January 17, 2015).

45. Filipenco D. (2024). World waste: statistics by country and brief facts. DevelopmentAid. Available at: https://www.developmentaid.org/news-stream/post/158158/world-waste-statistics-by-country

46. Al Mamun A, Hayat N, Masud MM, Makhbul ZKM, Jannat T, Salleh MFM. Modelling the significance of value-belief-norm theory in predicting solid waste management intention and behavior. *Front Environ Sci*. (2022) 10:906002. doi: 10.3389/fenvs.2022.906002

47. Chowdhury T, Chowdhury H, Rahman MS, Hossain N, Ahmed A, Sait SM. Estimation of the healthcare waste generation during COVID-19 pandemic in Bangladesh. *Sci Total Environ*. (2022) 811:152295. doi: 10.1016/j.scitotenv.2021.152295

48. He R, Sandoval-Reyes M, Scott I, Semeano R, Ferrao P, Matthews S, et al. Global knowledge base for municipal solid waste management: framework development and application in waste generation prediction. *J Clean Prod*. (2022) 377:134501. doi: 10.1016/j.jclepro.2022.134501

49. Taherdoost H. (2022). What are different research approaches? Comprehensive review of qualitative, quantitative, and mixed method research, their applications, types, and limitations (SSRN scholarly paper 4178694). Available at: https://papers.ssrn.com/abstract=4178694 (Accessed January 17, 2015).