Population and Housing Estimation, 2022

METHODOLOGY







San José, Costa Rica
MARCH 2026

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS

The reproduction of this document, in whole or in part, for educational or non-profit purposes is allowed, provided that the source is acknowledged. The INEC will be grateful for receiving any document prepared based on this publication.

Prepared by:
Production Process of Statistical Products of the Population and Housing Census Area

Design and Layout:
Graphic Production Process

317.2 Costa Rica. Instituto Nacional de Estadística y Censos
C837p Population and housing estimation 2022 : methodology [recurso electrónico] / Instituto Nacional de Estadística y Censos. – San José, Costa Rica : INEC, 2026.
1 recurso en línea (34 páginas).

Título original: Estimación de población y vivienda 2022 : metodología.

ISBN: 78-9930-641-11-8

1. Estimaciones de población 2. Estimaciones de vivienda 3. Metodología

Clarification note: Original publication: 2023 (Spanish version). English version published in 2026.

Presentation

The National Institute of Statistics and Census (INEC) presents the methodology for the population and housing estimation 2022 with the aim of ensuring transparency in the processes implemented and enabling a better understanding of the results obtained.

The population and housing estimates 2022 are the result of a complex technical process in which three different methodologies were implemented, validated, and analyzed to determine the final data to be released.

The INEC extends its deep appreciation to the Latin American and Caribbean Demographic Center (CELADE), the Statistics Division of the Economic Commission for Latin America and the Caribbean (ECLAC), the United Nations Population Fund (UNFPA), and the Central American Population Center (CCP) for their technical assistance and support throughout this process, as well as to the various institutional teams that took part in this effort to provide the country with the key demographic data for 2022.

Ligia Bermúdez Mesén
President
Directing Council

Marta Elena Esquivel Villalobos
Manager

Table of contents

Page

Presentation 3
Symbols and acronyms 6
1. Introduction 7
2. Context in which the Census 2022 was conducted 8
2.1 Census 2022 planning 8
2.2 COVID-19 pandemic 9
2.3 Budgetary constraints on the use of resources 10
2.4 Shortage of census takers 10
2.5 Technological changes 12
2.6 Training process 13
2.7 Nonresponse and refusal 14
3. Coverage estimation of the Census 2022 15
4. Methodologies for population and housing estimation 17
4.1 Population estimation based on census information 17
4.2 Demographic population estimation based on administrative records 25
4.3 Bayesian model for population estimation 28
4.4 Validation of results and definition of final population estimation 33
4.5 Housing estimation 33
Bibliography 34

Table index

Page

Table 3.1 Costa Rica. Distribution of housing units according to coverage status, 2022 16

Table 4.1 Costa Rica. Distribution of unoccupied housing units according to type, 2000 and 2011 . . . 22

Table 4.2 Costa Rica. Distribution of reclassified unoccupied housing units, 2022 24

Figure index

Figure 4.1 Costa Rica. Distribution of housing unit points by interview result, San Sebastian District, 2022 23

Figure 4.2 Costa Rica. Estimation of the number of unoccupied housing units and the number of inhabitants 31

Figure 4.3 Costa Rica. Population structure estimation (men and women in five-year age groups) 32

Chart index

Chart 4.1 Costa Rica. Average occupants per housing unit according to statistical operation, 2000-2022 18

Chart 4.2 Costa Rica. Average occupants per housing unit according to planning region, 2000-2022 19

Symbols and acronyms

Symbols

(-)	Used in a table to indicate if the amount is null or zero.
(.)	Used to indicate decimals.
(,)	Used to indicate thousands.

Acronyms and initialisms

CCP	Central American Population Center
CCSS	Costa Rican Social Security Administration
CELADE	Latin American and Caribbean Demographic Center
ECLAC	Economic Commission for Latin America and the Caribbean
CONADECO	Confederation of Community Development Associations
COVID-19	Disease caused by the SARS-CoV-2 virus
DGME	General Directorate of Migration and Alien Affairs
INEC	National Institute of Statistics and Census
MEP	Ministry of Public Education
WHO	World Health Organization
UN	United Nations
RUP	Single Population Registry
TEC	Costa Rica Institute of Technology
UNFPA	United Nations Population Fund

1. Introduction

The XI National Population Census and VII Housing Census was conducted between June and September 2022; however, due to various circumstances detailed in this document, the final coverage obtained was partial.

This situation created the need to prepare a population and housing unit estimation in order to provide the country with key demographic indicators for decision-making at both the national and cantonal levels.

To this end, the INEC worked with the advisory support of the Statistics Division of ECLAC, CELADE, UNFPA, and the Central American Population Center (CCP) to implement different statistical methodologies for population estimation.

Three estimation methodologies were developed, namely: one based on census data; another based on demographic administrative records; and a third based on a Bayesian model for population estimation. Finally, a process of validation and comparison of the three methodologies was carried out using different sources and the same census data to determine the final estimation.

In this context, and in the interest of the transparency and technical rigor that has always characterized the INEC, this document presents the factors that influenced the partial coverage of the Census 2022, the final coverage estimation, and the detailed methodology of the three population estimations carried out, as well as the final housing estimation.

2. Context in which the census 2022 was conducted

2.1 Census 2022 planning

Since 2018, the INEC began the planning work for the XI National Population Census and VII Housing Census, which is, undoubtedly, the most complex statistical operation carried out in the country, as it involves collecting information from all housing units and usual residents in Costa Rica.

Therefore, a preparation stage was launched, which included the formation of a working team, the preparation of a budget, and the design of the project profile. In this first stage, the project was also registered in the Project Bank of the Ministry of National Planning and Economic Policy (MIDEPLAN).

The development of a project of such magnitude as a national census requires an interdisciplinary team responsible for the various processes involved, namely: cartography, methodological design, data collection, data processing and analysis, information technology, communication, and, of course, administration and planning.

Between 2018 and 2021, the various necessary processes were designed and implemented, working around an Annual Project Work Plan that established quarterly goals and included risk matrices, which were reviewed periodically. During the national census planning period, the following activities were carried out:

- Cartographic field update, in which all housing units in the country were georeferenced.
- Design of the census questionnaire, which included a consultation with users and experts, involving the participation of more than 400 people.
- Design of the data collection operation, where 34 teams called “Census Agencies” were responsible for managing resources and carrying out the necessary coordination throughout the country.
- Design of the training process, which, for the first time in the history of censuses, was conducted in a bimodal format and, therefore, required the design of a virtual course and various educational resources.
- Design of the various information systems, both those developed by the INEC and those outsourced. These include the Censista App, the Self-Census, the Field Operations Management System, the Processing System, and the Computerized Coding System.
- Design of a comprehensive communication campaign and all related communication products.
- Coordination of agreements with various national and local institutions to provide support during training and data collection.
- Management of the logistics for distributing and safeguarding census materials throughout the country.
- Procurement management of mobile data collection devices and other supplies and services required for the implementation of the field operation.
- Implementation of seven field tests, cognitive tests, and a Trial Census in which the entire canton of El Guarco, in Cartago, was enumerated.

All processes included specific deliverables and were evaluated by the institution's internal control system. Nevertheless, despite the extensive and detailed planning, there were factors that particularly affected data collection, which are detailed below.

2.2 COVID-19 pandemic

In March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. On the 16th of the same month, Costa Rica declared a national state of emergency due to the pandemic, which imposed a series of restrictions on institutions, businesses, and the general public, with the objective of safeguarding the population's health.

The main restrictions applied included the implementation of remote work in public institutions and private companies, the closure of businesses with sanitary permits, and vehicle circulation limits. Throughout 2020 and 2021, sanitary restrictions were either relaxed or tightened depending on the so-called "waves" of the virus, meaning periods with significant increases or decreases in cases.

At that point, the INEC had been working for two years on the planning of the national census, which was originally scheduled for 2020. However, the pandemic was one (though not the only) factor in the decision to postpone it—first to 2021 and then to 2022.

The changes implemented included reorganizing the tasks of the Census Agencies to prioritize those that could be carried out remotely; applying a cartographic update methodology that did not require field visits to complete pending districts; changing the data collection methodology to self-enumeration during the census operation, which was originally intended only for follow-up visits to unenumerated housing units; acquiring sanitary protection equipment (face masks and hand sanitizer) for all individuals participating in the operation.

The Trial Census was conducted in November 2020 in El Guarco canton, Cartago province. However, due to restrictions, it was not possible to hire dedicated personnel for this purpose, so the entire operation was carried out with personnel from the same Census Agencies. In addition, training—originally planned in a bimodal (virtual/in-person) format—had to be conducted 100% virtually.

Although the Trial Census met some of its objectives, such as testing the questionnaire and the data capture systems, the adjustments made due to the pandemic had an impact, as they prevented full testing of the conditions of the national census, such as the recruitment process or the mass loading of the capture application.

2.3 Budgetary constraints on the use of resources

Law No. 9694 of the National Statistics System, enacted in June 2019, provides the INEC with financial capacity to conduct national censuses. Therefore, the resources needed for the implementation of the census project were available. However, in 2020 and 2021, the INEC faced the restrictions established by Law No. 9635, known as the "Public Finance Strengthening" Law, which sets a fiscal rule limiting the growth of current expenditure in public institutions.

Due to uncertainty over the use of available resources to continue the project in 2020 and 2021, the decision was made to postpone the national census to 2022. Nevertheless, it became necessary to advocate for an amendment to Law No. 9694 to make it possible to budget the required resources.

The amendment was achieved in November 2021, near the end of the year, which left many aspects in limbo and very short timelines to start recruiting census takers, purchasing supplies such as mobile devices, and contracting services such as the technological infrastructure needed to implement the developed systems. It also delayed the signing of key agreements, such as the one required with the Ministry of Public Education (MEP) for the use of facilities and personnel to carry out training.

2.4 Shortage of census takers

The Census 2022 was the first in Costa Rica to be conducted using mobile data collection devices, and this significant change affected nearly all project processes, including the personnel for data collection.

Traditionally, census data takers were teachers from the MEP. This was possible because the data collection period lasted five days and, through an agreement, authorization was granted to suspend classes during that period so that MEP personnel could conduct the census. They received their regular salary plus an additional financial incentive from the INEC.

However, when the decision was made to incorporate mobile devices for data capture, it also became necessary to extend the collection period for several reasons: the limited number of devices that could be acquired, the logistics of installing the application, and the need for greater flexibility to respond to any eventualities. In light of this, the period was extended to 18 days, for which approximately 14,000 people were required, distributed among operational center coordinators, section supervisors, and census takers.

Given these changes and, in addition, the educational crisis the country faced due to all the adjustments the Ministry of Public Education (MEP) had to make during the COVID-19 pandemic, it was determined that for the Census 2022 it would not be possible to work with teachers—at least not for data collection—although it was planned for them to provide support as facilitators during the training process.

This decision meant that the INEC had to seek new alternatives for hiring field personnel, which was no easy task since there is little national experience with such large-scale hiring for such a short period (approximately one month). For this reason, the first step was to conduct research on the different options available.

Several alternatives emerged, each with advantages and limitations: volunteer work, hiring through professional services contracts, outsourced hiring, or direct hiring by the INEC through temporary positions. According to all analyses, the option with the least risk was the last one—creating temporary positions and having the INEC, which already had experience managing very large payrolls on a temporary basis, handle the recruitment and selection process.

To make this possible, it was necessary to approach the Ministry of Finance and the Technical Secretariat of the Budgetary Authority to obtain authorization for the INEC to create the 14,000 temporary positions. However, this request was not approved, and it was therefore necessary to opt for a riskier alternative.

This alternative involved finding an entity capable of managing the payroll and carrying out the recruitment, selection, and personnel administration process for the entire country. Initially, a market study was conducted to identify possible private companies that could provide the service; however, none were willing to take it on due to its complexity.

Given this situation, the possibility arose of working with the Confederation of Community Development Associations (CONADECO), the organization that brings together the various community development associations in the country through federations and has a presence in almost the entire national territory.

The advantage of working with CONADECO was precisely its extensive knowledge of the territories, its grassroots work experience, and its ability to mobilize communities toward a common objective. However, it had never managed such a large number of paid personnel before.

The methodology used consisted of the INEC contracting 14 federations, which together made up CONADECO, with each federation managing its own payroll. In addition, personnel were also hired to provide administrative and logistical support, such as an administrative coordinator, a financial coordinator, human resources professionals, and payroll assistants for each federation.

However, time became one of the main challenges, since by the end of 2021, due to uncertainty over whether the project would continue, all negotiations had been put on hold, and it was not until early 2022 that work with CONADECO was able to begin.

While there was a strong willingness to work in coordination, some federations needed to complete preliminary procedures before they could be contracted, which delayed certain planned activities, and some contracts were signed very close to the start date of the field operation.

Recruitment began with preliminary lists prepared by each federation, in which interested individuals could register. In this process, known as pre-recruitment, more than 20,000 people were registered, meaning that, at that stage, the established objectives were being met.

The formal recruitment process had to be carried out through a web form, which also served as the system for tracking each person during training, selection, and assignment of their workload. Although it was a simple system, it represented an initial obstacle for many, as it required an email address and basic computer skills. Even though candidate profiles had been defined—requiring at least basic knowledge of how to use a computer or mobile device—it became evident that some pre-recruited individuals did not meet this requirement, and the list began to reduce.

Another factor that reduced the list was that some people had applied for all three available positions (operational center coordinator, section supervisor, and census taker) but later, upon learning they did not meet certain requirements or realizing that a training process was mandatory, chose to withdraw from the recruitment. In addition, there were areas of the country—particularly tourist zones—where no applicants were interested in the job.

Ultimately, by the close of the recruitment period, the total required personnel had not been reached, especially for the census taker position, which had a deficit of approximately 2,000 people. This situation led to an extension of the recruitment process even after the field operation had begun; however, this required adjustments to other related processes, such as training. Therefore, at the end of the first week of the operation, it was decided to stop hiring and proceed with the personnel already contracted.

After the original data collection period ended (June 28), contracts were extended for approximately 1,500 people, who continued collecting information for an additional three weeks in the areas that had not yet been covered.

2.5 Technological changes

From the outset of the project, it was decided that the national census would be carried out by maximizing the use of technology in all processes—from cartographic updating, field operation design, and data collection to training, personnel management, and, of course, the processing and analysis of information.

With this in mind, the INEC sought to hire experts in the field to provide guidance in this area and to develop the application for data collection—both in the field and via the web—as well as the system for personnel management and coverage monitoring, which is often one of the most complex aspects of such a large statistical operation. After a market analysis process, Go Touch—an IT research initiative of the Costa Rica Institute of Technology (TEC) specializing in mobile application development—was contracted.

The objective was to have a series of integrated systems that would manage everything from recruitment and selection of personnel to workload assignments and subsequent coverage tracking during the operation. In this way, the same user account created during the recruitment process would also be used to log into the data collection application (called the Censista App) and the supervision module.

While this integration was intended to facilitate information traceability and ensure control, it created bottlenecks because some people did not remember the username they had created, had entered incorrect or inactive email addresses, had full inboxes, or even used someone else's email address—preventing them from accessing the rest of the processes, including their workload.

In addition to the developed systems, it was also necessary to obtain a mobile device management system to allow the INEC to control the devices, restrict access to non-work-related applications, and facilitate the installation of the Censista App. The system used was Google Workspace, which provided all these features but required each person to install an account on the device assigned to them.

Although the process was straightforward, in an operation of this scale even minimal complexity becomes significant, and therefore this installation also became a bottleneck, as not all people—despite receiving training and explanatory resources—were able to complete it within the required timeframe.

As a result of these bottlenecks, on the first day of the operation not all people had managed to access the application and therefore their assigned workload. This issue decreased during the first few days; however, some people were never able to log in throughout the operation and therefore had to work with paper questionnaires.

There were also certain border areas where the digital maps that should have been incorporated into the Censista App had coordinate issues; as a result, those areas also had to be enumerated using paper questionnaires, since the assigned census takers would not have been able to access those maps.

In addition to the complications associated with system use, another factor that further complicated the situation was that approximately two months before the start of the census, the Government of Costa Rica suffered a cyberattack that affected several institutions, causing their systems to be inoperative for several weeks. Although the INEC was not directly targeted, stricter cybersecurity measures had to be implemented, which, in turn, made it more difficult for users to access the systems.

2.6 Training process

For the first time in a national census, a hybrid or blended training program was implemented, consisting of a virtual self-paced course for each role within the field operation and two days of in-person training that included practice activities and installation of the Censista App.

This process was directly affected by the delays and difficulties that arose during personnel recruitment, since the extension of timelines also required modifying course closing dates and even creating a shortened course for those hired after the operation had already begun.

For the in-person component, the original plan followed a cascade methodology: an initial group of INEC officials would train teachers from the Ministry of Public Education (MEP), who would then facilitate in-person workshops for supervisors and census takers.

However, due to uncertainty about the project's continuation, it was not possible to sign the agreement with the MEP in time. Despite all efforts, the outgoing administration's leadership did not sign the agreement, and it was necessary to wait for the change of government for the new minister to sign it— only two weeks before the start of in-person training.

With so little time to prepare, and given the MEP's highly hierarchical structure, the information was not communicated to schools in time. As a result, only 50% of the required MEP personnel attended their training. This meant that more than one hundred INEC officials had to deliver training directly at the final level, and operational center coordinators had to assume this role in some locations. Contingency training sessions were also held, focusing on condensed key content, particularly the installation of the Censista App and practical exercises.

2.7 Nonresponse and refusal

Despite all the efforts made by the INEC to overcome the aforementioned challenges and to visit over 90% of the established enumeration areas, the obstacle that could not be overcome was the public's apathy, which led to refusal and nonresponse rates never before seen in a national census.

This attitude is part of a broader context that goes beyond the census itself and reflects a social process currently taking place in the country, which has also been evident in other significant events, such as voter abstention in electoral processes.

Political apathy is not usually associated with a lack of interest in participating in a statistical operation; however, when it comes to a national census as a process for generating data with political relevance, it is indeed related—particularly to the component of disaffection, that is, indifference and lack of interest in matters that, in one way or another, require citizen participation or action (Cebrián, 2011; Romano, 2015).

The causes of such apathy can be diverse, including distrust, dissatisfaction, lack of motivation, exclusion, inequality, among others, which may develop in the population through long-term processes or as a result of specific situations in a given context. In Costa Rica's case, an in-depth analysis of this apathy is certainly needed; however, it is evident that, for the Census 2022, certain specific events and corruption cases worsened these attitudes.

Costa Rica is not the only country facing this type of apathy; other countries have also observed, even in past census rounds, an increase in interviews that were not conducted or were refused. In fact, the Russell Sage Foundation published in 2006 the book *The Hard Count: The Political and Social Challenges of Census Mobilization*, which aimed to explore participation in the United States Census 2000 and evaluate the communication campaign for the Census 2010. The book also reviews a series of contextual elements—mainly political—that influence participation.

The book concludes that one of the factors associated with the lack of cooperation in the census is the public's distrust regarding how the government will use their data and with whom it will share them, which, in turn, implies distrust in the producing institution and in the confidential handling of the information. In addition, it found a relationship between household type and cooperation with the census: there is greater cooperation among conventional households (couples with or without children) than among non-conventional households (single-parent households, single-person households, or households composed of non-relatives). It is difficult to determine exactly why this is the case, but it may be related to lifestyle and time-use patterns (Hillygus, Nie, & al., 2006).

While it is necessary to further explore the causes of nonresponse and refusal to provide information, the fact remains that these factors significantly affected the results of the Census 2022. Although coverage implies visiting all defined areas, it is also a fundamental requirement to complete all interviews, and this was not possible in a significant percentage of housing units.

3. Coverage estimation of the Census 2022

For monitoring coverage during the Census 2022 data collection, the Processing Management System was used, along with the continuous transfer of interviews, which made it possible to track the operation's progress in real time.

As a result, coverage problems were identified early, particularly in areas where it had not been possible to hire the full number of required personnel, leaving enumeration areas without an assigned census taker.

In this context, the decision was made to extend the data collection period until September 30, 2022—almost three months beyond the original schedule (June 8 to 28). To cover this additional period, contracts were extended for approximately 1,500 people for three more weeks, and later, support was provided by volunteers and by the INEC itself.

In addition, the use of self-enumeration was promoted through the media, and an online portal was created to allow respondents to request a code without having to wait for a census taker to deliver it. Telephone interviews were also offered, with follow-up to encourage participation.

While the percentage of unvisited housing units decreased from 28% at the end of the regular period to 12% by the close of the extended period, the percentage of visited housing units that either refused the interview or could not be contacted for the interview did not decrease.

Ultimately, due to the administrative complexity of continuing the process and the lack of expected results, the data collection operation was officially closed on September 30.

Afterwards, the following actions were taken to determine final coverage:

- **Integration of data**: Although more than 95% of the data were collected using mobile capture devices, 32,463 self-enumeration interviews were received, and approximately 16,000 paper questionnaires were completed, which underwent review and digitization. In addition, data from 140 collective housing units—such as correctional facilities—were obtained from administrative records due to the complexity of conducting face-to-face interviews in these places. A process was therefore required to integrate these different sources into a single housing database.
- **Housing database cleaning**: During the cartographic field update, all housing units were georeferenced and included in the maps provided to census takers. The initial database contained 1,770,375 housing units. However, census takers updated this information during their visits, identifying housing units that were still in place, no longer existed, or were newly built. By the end of the operation, 203,125 new housing units had been added and 144,605 removed. Duplicate and erroneous points were also eliminated, resulting in a final database of 1,769,081 housing units. This final database served as the reference for housing unit coverage estimation.

Based on that information and on the cleaned data, it was determined that 88% of the country's housing units—1,550,168 in total—were visited. Of these, 1,074,258 were successfully enumerated (60.7%); 136,099 refused the interview (7.7%); and 339,901 (19.2%) were visited but the interview could not be completed. The total number of unvisited housing units was 218,823, representing 12.4%.

TABLE 3.1
Costa Rica. Distribution of housing units according to coverage status, 2022

Housing unit status	Absolute	Relative
Total	**1,769,081**	**100.0**
Visited	**1,550,258**	**87.6**
Enumerated	1,074,258	60.7
Refused	136,099	7.7
Pending	339,901	19.2
Not visited	**218,823**	**12.4**

Source: INEC-Costa Rica. Census 2022.

Due to the partial coverage, it was necessary to perform a population estimate to provide a figure suitable for use. The methodologies used for this purpose are detailed below.

4. Methodologies for population and housing estimation

For the implementation of the population estimation methodologies, the valuable technical support and guidance of the Statistics Division of ECLAC, the Latin American and Caribbean Demographic Centre (CELADE), the United Nations Population Fund (UNFPA), and the Central American Population Center (CCP) were instrumental. Thanks to this collaboration, it was possible to develop different estimation methods based on statistical techniques, which made it possible to produce a robust and high-quality final figure.

Three methodologies were developed for population estimation: one based on the census information itself and the average number of occupants per housing unit; another using demographic information from administrative records; and a third implementing a Bayesian predictive model for population estimation based on housing units and various covariates. Each methodology is described below.

4.1 Population estimation based on census information

The first methodology estimates the population, at the district level, using the average number of occupants from the housing units that were successfully enumerated. To do this, it was first necessary to analyze and make decisions regarding certain factors that could affect the estimation, since—due to the partial coverage—it was essential to ensure that the data used were adequate and yielded coherent estimates. The aspects analyzed were:

- Consistency of the average number of occupants per housing unit by district.
- Adjustment of the classification of unoccupied housing units.
- Classification of unvisited housing units.

The processes are detailed below:

4.1.1 Consistency of the average number of occupants per district

Due to the partial coverage of the Census 2022, it was necessary to validate and adjust the averages of occupants per housing unit, since this figure would be the key input for this estimation methodology.

For this purpose, a regression model was implemented which, through a linear function, seeks to model a variable Y in terms of a set of predictor variables X, using the following formula:

$$Y\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Where:

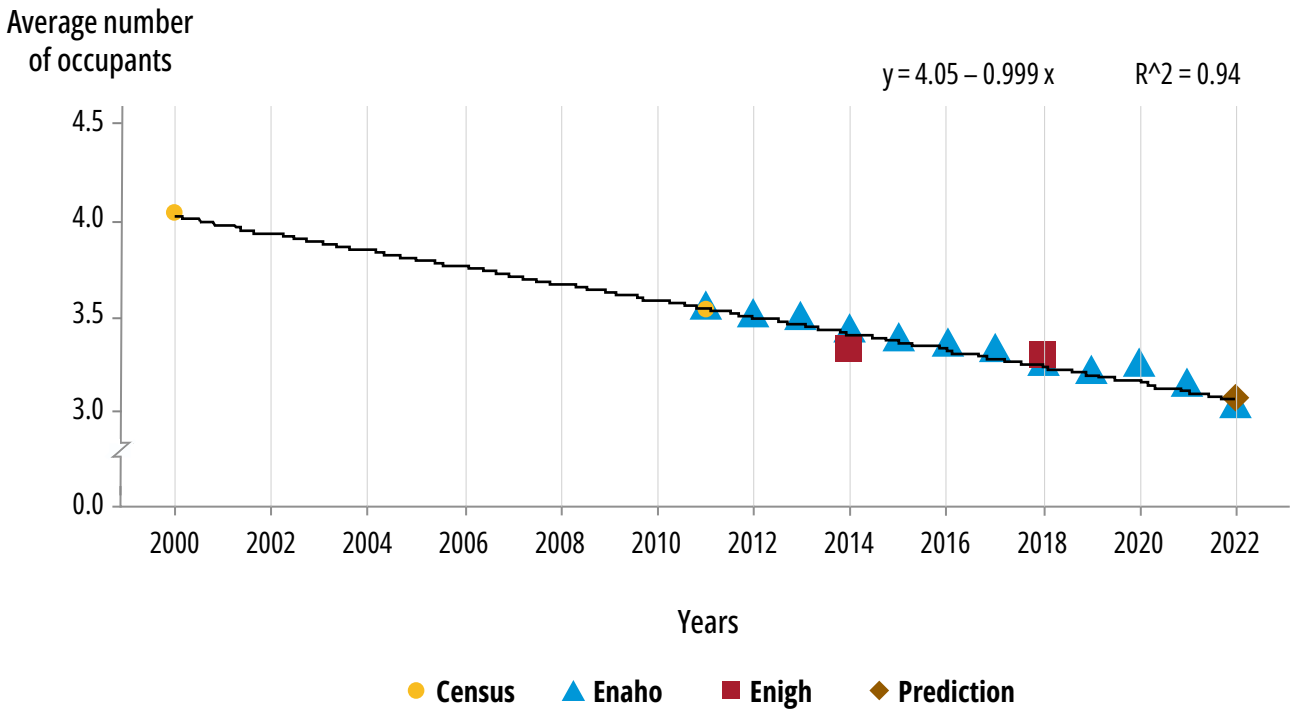
- β_i = 0, 1, 2, 3 are unknown parameters that are estimated from the data used for prediction. In this case, the observed averages of people from the 2014 and 2018 National Household Income and Expenditure Survey, the National Household Survey 2010-2022, and the Censuses 2000, 2011, and 2022 were used.
- ε represents the error, which is the distance between a response value and the conditional mean given a value of X_i.

To obtain the coefficient values of the regression model equation, the ordinary least squares technique was applied. Once the coefficients were obtained and the value of the specific year for which the prediction was desired was substituted, the estimated average number of people per housing unit was produced.

This methodology was implemented at both the national level and for each of the planning regions[1/], resulting in 7 regression lines: one for each region and one national estimate for 2022.

In Chart 4.1, it can be seen that, when applying the model at the national level, the result is a positive intercept and a negative slope—meaning that over the period analyzed, the average number of people per housing unit decreased by approximately 0.999 people per housing unit.

CHART 4.1
Costa Rica. Average occupants per housing unit according to statistical operation, 2000-2022

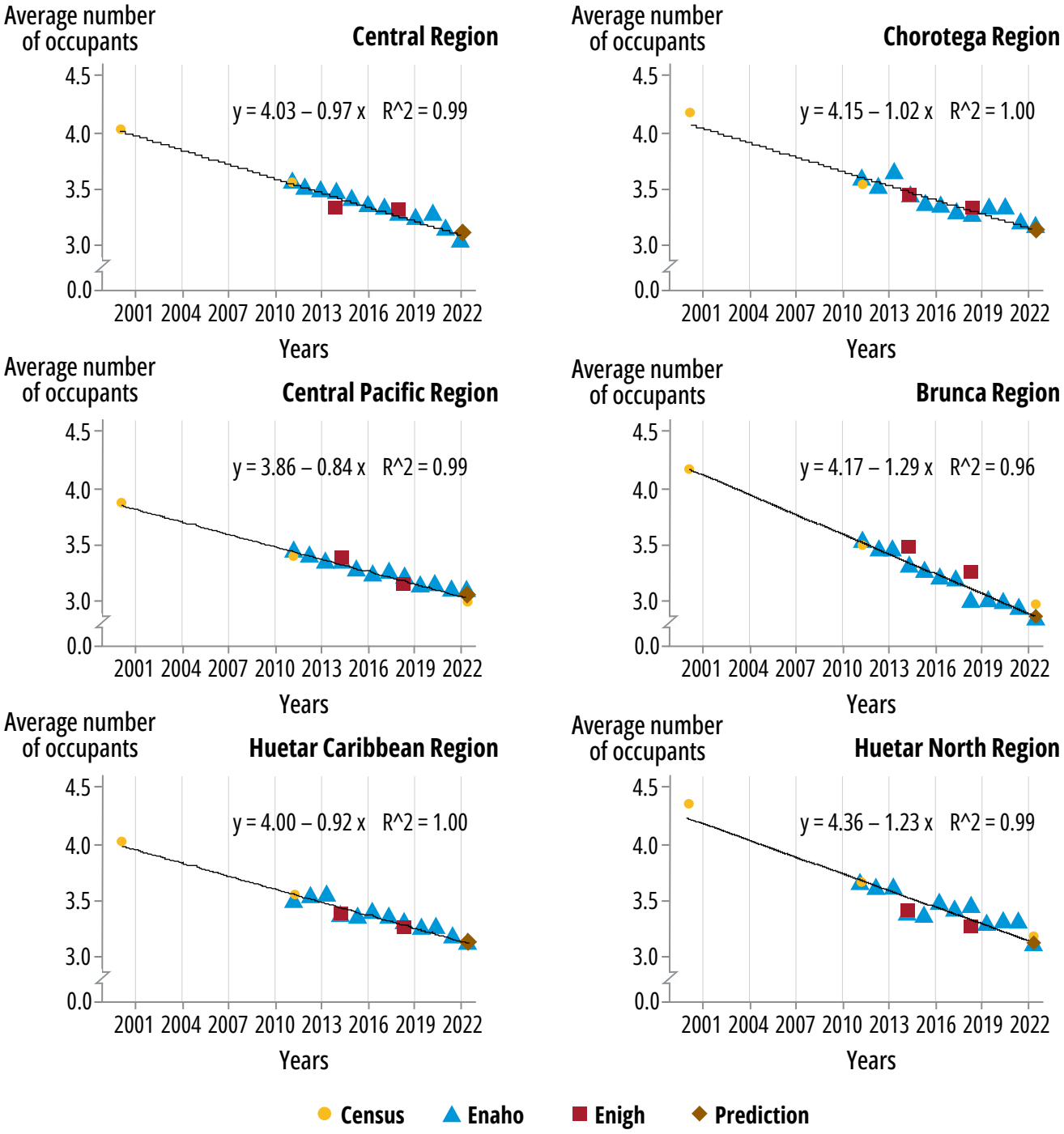


Source: INEC-Costa Rica. National Household Survey, 2011-2022. National Household Income and Expenditure Survey, 2014 and 2018. IX National Population Census and V Housing Census, 2000. X National Population Census and VI Housing Census, 2011. Census project, 2022.

1/ The planning region is the smallest geographic disaggregation allowed by the sample of the surveys analyzed.

Likewise, Chart 4.2 shows that each region has a negative slope, with the Brunca region showing the largest decrease over the years in the average number of people per housing unit.

CHART 4.2
Costa Rica. Average occupants per housing unit according to planning region, 2000-2022



Source: INEC-Costa Rica. National Household Survey, 2011-2022. National Household Income and Expenditure Survey, 2014 and 2018. IX National Population Census and V Housing Census, 2000. X National Population Census and VI Housing Census, 2011. Census project, 2022.

As can be seen, the averages obtained after the regression analysis are at the planning region level, as this is the smallest geographic disaggregation achieved by the surveys.

To disaggregate the averages at the canton level, the calculation of population, housing units, and average occupants per canton was performed using the exponential growth rate. For this, the cantons had to be divided into two groups, due to the constant updates in the administrative territorial division. The first group contained cantons that remained unchanged between 2000 and 2011, and the second group included those created after the year 2000.

For the first group, the exponential growth rate—which assumes continuous growth for each year elapsed—was applied using the following formula:

$$r = \frac{1}{k} * \frac{N_f}{N_i}$$

Where:

- k is the time elapsed between the two censuses (11 years).
- N_f is the final population (Census 2011).
- N_i is the initial population (Census 2000).

Subsequently, the projection for housing units, population, and the average number of people per housing unit for 2022 was calculated using the following formula:

$$N_{2022} = N_{2011} * e^{r*t}$$

Where:

- N_{2022} is the projected figure for the year 2022.
- N_{2011} is the figure from the Census 2011.

For the second group, the growth rate of the cantons from which they originated was assigned in order to calculate the updated population, housing units, and average number of people per housing unit for 2022. Since these cantons did not have records for population and housing units for both periods, the percentage of the population represented by each in the census project was calculated, and based on that figure, housing units and population from the originating canton(s) were proportionally allocated to obtain the projected figures.

With this information, it was assumed that the cantons behave similarly to the planning region to which they belong, and the following formula was applied:

$$f = \frac{\overline{X}_{regression_r}}{\overline{X}_{exponential_r}}$$

Where:

- $\overline{X}_{regression_r}$ **:** average number of people per housing unit obtained from linear regression.
- $\overline{X}_{exponential_r}$ **:** average number of people per housing unit estimated from the exponential growth rate.
- r *:* planning region.

Next, the final average number of people per housing unit was calculated using the following formula:

$$\overline{X}_{final} = \overline{X}_{exponential_c} * f$$

Where $\boldsymbol{c}$ corresponds to the canton.

4.1.2 Adjustment of the classification of unoccupied housing units

The Census 2022 recorded information for 795,860 occupied housing units and counted 476,000 pending or refused housing units, giving a total of 1,271,860 housing units with usual residents. In addition, 278,142 unoccupied housing units were identified, representing 17.9%. In the Census 2011, the percentage of unoccupied housing units was 10.8%.

The difference in unoccupied housing units between 2011 and 2022 was nearly 40%, which could be even higher if we consider that among the unvisited housing units there could also have been unoccupied units. Therefore, it was decided to conduct a more detailed analysis of unoccupied housing units to validate the data or identify whether there may have been misclassification.

According to the censuses 2000 and 2011, the largest proportion of unoccupied housing units were available for rent or sale, followed by temporary units for vacation, work, or study. While the category "Other unoccupied" (a residual category) had a significant percentage in 2000 (20.4%), it decreased to 18.1% in 2011. This category refers to other reasons for vacancy not otherwise specified, such as abandoned housing units.

For the Census 2022, the distribution of unoccupied housing units showed an unexpected pattern: the share of units for rent/sale and temporary use decreased, while the "Other" category increased. This trend is inconsistent with the country's real estate patterns, where there has been growth in both residential and vacation housing construction. Table 4.1 presents the distribution of types of unoccupied housing units for the last three censuses.

TABLE 4.1
Costa Rica. Distribution of unoccupied housing units by type, 2000 and 2011

Type of unoccupied housing unit	Census 2000		Census 2011		Census 2022	
	Absolute	Relative	Absolute	Relative	Absolute	Relative
Unoccupied	**98,650**	**100.0**	**147,204**	**100.0**	**278,142**	**100.0**
Unoccupied for rent or sale	31,559	32.0	51,426	34.9	92,797	33.4
Temporarily unoccupied (vacation, workers)	30,404	30.8	50,081	34.0	78,842	28.3
Unoccupied under construction or repair	16,541	16.8	19,004	12.9	31,319	11.3
Other unoccupied	20,146	20.4	26,693	18.1	75,184	27.0

Source: INEC-Costa Rica. Censuses 2000, 2011, 2022.

Given this, a more in-depth study was undertaken to identify possible causes for the increase in the "Other" category among unoccupied housing units. A comparative analysis of all vacancy categories was conducted using data from the censuses 2000 and 2011, the National Agricultural Census 2014, the urbanity index, and cartographic reference information, with the objective of correlating unoccupied housing unit data with possible reasons for their existence.

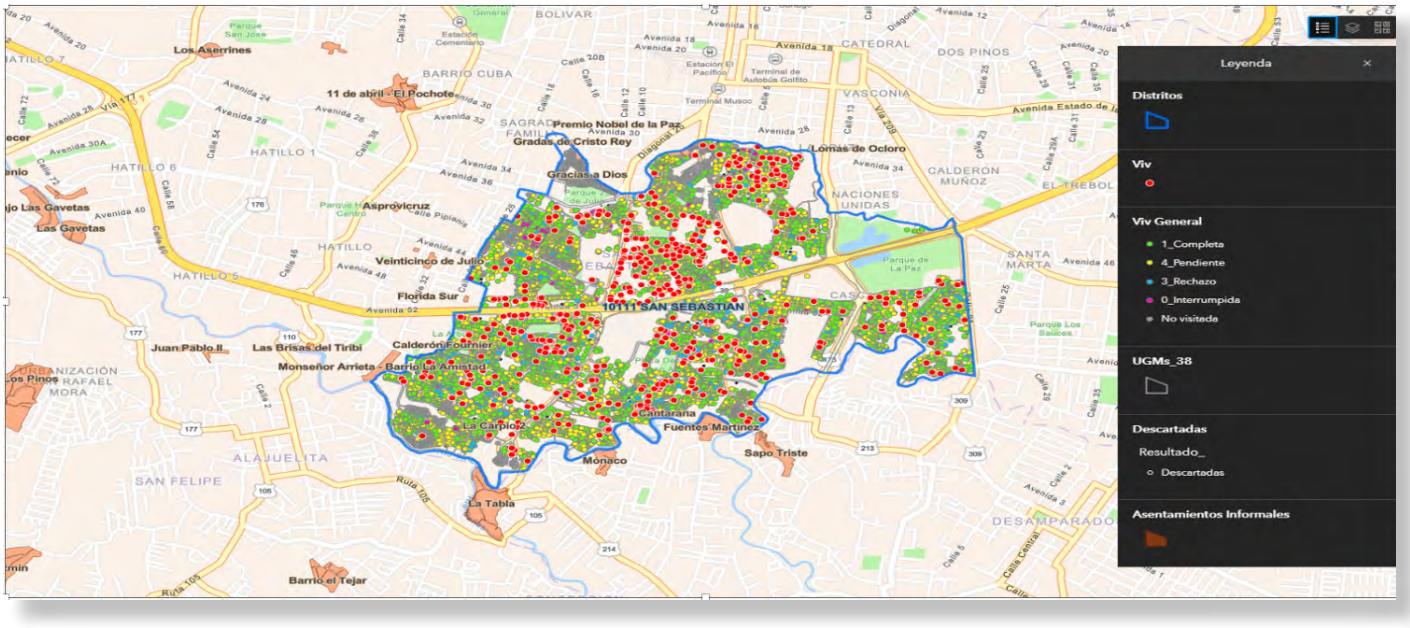
The results showed concentrations of unoccupied housing units for workers in areas where this would be expected, due to predominant agricultural or industrial activities (e.g., the Los Santos area). There was also coherence in the concentration of "student" unoccupied housing units in districts near universities. Additionally, expected patterns were observed for units for rent, sale, or under construction/repair in districts experiencing residential real estate growth, such as Carmen in San Jose or San Mateo in Alajuela.

However, the concentration of "Other" unoccupied housing units did not appear to be related to any particular economic activity and, in fact, appeared in areas where such a concentration would not be expected, such as informal settlements or indigenous territories. Examples include Los Guido, Telire, Boruca, and Brunka, where more than 50% of unoccupied housing units fell into the "Other" category.

To deepen the analysis, the distribution of unoccupied housing units was mapped, and for some atypical districts, a point-level georeferenced analysis was conducted specifically for "Other" unoccupied units. In addition, in districts with inconsistent results, concentrations of "Other" unoccupied housing units were identified to find possible explanatory patterns.

The map below shows an example of this analysis for the San Sebastian district. The red dots represent unoccupied housing units classified as "Other." As can be seen, there are areas with significant concentrations of this type of housing units, which is not to be expected in such a densely populated district, given that this is a residual category.

FIGURE 4.1
Costa Rica. Distribution of housing unit points by interview result. San Sebastián District, 2022.



Note: Image taken from the unoccupied housing unit viewer. The red dots represent unoccupied housing units classified as "Other."
Source: INEC-Costa Rica. Census 2022.

It is presumed that in these areas, census takers may have misinterpreted the use of this category. Therefore, it was concluded that the "Other unoccupied" category was erroneously used by census takers to classify housing units that were actually occupied but where no interview could be completed, and which should instead have been classified as "Pending."

Based on the analysis of unoccupied housing units, it was decided to reclassify all units listed as "Other unoccupied" as "Unvisited," and subsequently determine whether they were truly unoccupied or occupied.

In addition, unoccupied housing units located in Minimal Geostatistical Units (UGM) with more than 20 days of elapsed enumeration were also considered suspect, since after that period it was unlikely that unvisited unoccupied housing units would remain. These could also be cases of incorrect classification and, similar to the "Other unoccupied" category, could, in fact, be pending housing units.

Finally, both the unoccupied housing units and those identified in UGM visited after 20 days were adjusted by applying, as a factor, the proportion of unoccupied housing units in the canton. The housing units obtained through this process were categorized as unoccupied, and the remainder were reclassified as occupied housing units.

The following table shows the final distribution of reclassified housing units.

TABLE 4.2
Costa Rica. Distribution of reclassified unoccupied housing units, 2022

Type of unoccupied housing unit	Classification		
	Total	Occupied	Unoccupied
Other unoccupied	75,184	65,690	9,499
Unoccupied housing units: 20-day extreme	38,999	34,603	4,396

Source: INEC-Costa Rica. Censuses 2000, 2011, 2022.

After these adjustments, the final figure for unoccupied housing units was 214,654, representing 11.7% of the country's housing units.

4.1.3 Classification of unvisited housing units

For the estimation, it was necessary to classify unvisited housing units as either occupied or unoccupied. For this purpose, the same factor used in the previous stage (adjustment of unoccupied housing units) was applied—namely, the proportion of unoccupied housing units in the canton was used to allocate unoccupied housing units, and the remainder were classified as occupied housing units.

In total, there were 218,823 unvisited housing units, of which 193,155 were classified as occupied and 25,668 as unoccupied.

4.1.4 Population estimation

With the adjustments made and the validated averages of occupants per housing unit, the population estimation process was carried out. This consisted of adding all pending and refused housing units, along with the housing units reclassified as occupied, and multiplying them by the average number of occupants per housing unit for each canton.

Once this estimated population was calculated, it was added to the total number of people actually enumerated in both individual and collective housing units to obtain an estimated population by canton. The national population estimated using this method was 4,864,442.

4.2 Demographic population estimation based on administrative records

This methodology was based on the use of various administrative records, such as the electoral roll, the master birth file[2/], the master death file[3/], the birth and death databases of the INEC, and the entry and exit database of the General Directorate of Migration and Alien Affairs (DGME).

The demographic estimation was carried out by constructing a base population in different stages, as explained below.

4.2.1 National population estimation

The national population estimation was performed in two stages: first, estimating the population under 20 years of age, and second, the population aged 20 years and over.

The population under 20 years of age was estimated by reconstructing the birth cohorts that make up this population in 2022. For this, the INEC birth and death databases, corrected for late registration, were used, following the procedure below:

1. A Lexis diagram was created, through which the population aged 0 to 19 years was estimated as of January 1, 2022, and as of January 1, 2023.
2. To reconstruct this population as of January 1, 2022, using the Lexis diagram, births from the period 2002-2021 and deaths of individuals under 20 years old from the same period were used. For the population as of January 1, 2023, births and deaths of individuals under 20 years old from the period 2003-2023 were used.
3. To obtain the proportion of deaths in each year corresponding to each cohort, Glover factors were used.
4. Separate reconstructions were carried out for males and females, and the total was obtained by summation.
5. In order to obtain the mid-term population (June 30, 2022), the exponential growth formula was used.

$$P_f = P_i * e^{rt}$$

Where:

- P_f represents the population under 20 years old as of January 1, 2023.
- P_i represents population under 20 years old as of January 1, 2022.
- r is the growth rate between both periods.
- t is the period to which it is necessary to move population, which in this case is mid-year.

2/ Database grouping all the births historically registered in the country.
3/ Database grouping all the deaths historically registered in the country.

To apply this formula, first r had to be calculated, which is obtained by clearing the same formula:

$$r = \frac{LN\left(\frac{P_f}{P_i}\right)}{t}$$

The national population aged 20 years and over was estimated using the electoral roll, adjusted by removing deaths. This was done by linking the electoral roll with the master death file using the national identification number.

Finally, the total national population was obtained by summing the population aged 0-19 years and that aged 20 years and over. The estimated national population was 4,732,177 people.

4.2.2 Foreign population estimation

The foreign population was estimated in two stages: first, the Nicaraguan population—by far the largest foreign group in the country—was estimated; then the rest of the foreign population.

First, the ratio of Nicaraguan population to Costa Rican population was obtained for each age and sex group, both from the Census 2011 and from the Census 2022. These ratios were then averaged for each age-sex group. Next, these specific ratios were multiplied by the population in each age-sex group to obtain the total Nicaraguan population.

The same procedure was applied using the ratio of other foreign populations to obtain their total population. Finally, the two populations were summed to obtain the total foreign population, which was estimated at 461,547 people.

4.2.3 Costa Rican emigrant population estimation

To estimate the Costa Rican emigrant population, the DGME entry and departure base was used. Using the cycle methodology[4/], only coherent cycles were taken into account—where the last movement was an exit that had occurred 6 months[5/] or more before the reference date.

An important aspect to note is that the number of departures was limited to three, in order to filter cases of people who, having many movements, appear to be international tourists or entrepreneurs, rather than emigrants.

To this methodology were added single movement cycles (departure) which had to have been performed six months or more since the cut-off date.

4/ A cycle is defined as a sequence of two movements, for example: entry-departure or departure-entry.
5/ Regarding the cut-off date, which is June 30, 2022.

The departures of people over 100 years old were prorated, as well as cases with no date of birth.

Since the entry and departure databases do not include individuals under 18 years old, an estimate was made using the ratio of emigrant population to resident Costa Rican population, derived from the Census 2022, given that the Census 2011 did not contain information on emigrants by sex and age.

To evaluate the ratio from the Census 2011, the same ratio was estimated using data for emigrants aged 20 and over and the resident Costa Rican population from the electoral roll. This estimate was compared with that of the census, and the difference was calculated, that is, how much the census underestimated this ratio.

Finally, to estimate the emigrant population under 18 years old, the emigrant/resident ratio was adjusted using the figure for women aged 20-24, selected as the closest group and based on the assumption that most children migrate with their mother. The estimated emigrant population was 149,527 people.

4.2.4 Total population estimation

To obtain the total population, the balancing equation was applied using the following formula:

Total population = *national population + foreign population – emigrant population*

The final estimated population for June 2022 was 5,044,197 people.

4.2.5 Disaggregation of the population by province and canton

To determine the best way to distribute the population among provinces and cantons, various demographic sources were analyzed, including births, deaths, the electoral roll, current population estimates and projections, as well as the distribution derived from the application of the Bayesian model described later.

It was ultimately determined that the most suitable structure was obtained from the average of the distributions of births and deaths for 2019 (pre-pandemic) and 2022 (post-pandemic), the electoral roll as of June 30, 2022, and the distribution derived from the Bayesian model—producing a model structure.

This model structure was then applied as a factor to the national population to obtain the population disaggregated by canton. Provincial populations were obtained by summing the figures for their respective cantons.

4.3 Bayesian model for population estimation[6/]

Through weekly meetings, under the leadership and active participation of the INEC, the Statistics Division of ECLAC, along with CELADE and UNFPA-LACRO, carried out a series of studies on modeled populations that resulted in the definition of three statistical models. These models integrated different sources of information: pre-census, census, administrative records, and geospatial information[7/] that could help predict the spatial distribution and density of the population. The compilation of geospatial information was carried out respecting the boundaries of the Minimum Geostatistical Units (UGM), which served as census sectors and enumeration areas in this Census. The details of the models are as follows:

1. In the first stage, a statistical model was defined to predict the probability that a housing unit was occupied. Because in the census operation there was a large number of housing units that never answered the census taker's visit, refused to be interviewed, or were simply never reached, it was necessary to assign an occupancy probability to these units. The model used was a Bayesian mixed model with a binomial response and random effects for provinces, cantons, and districts, as described below:

$$U_{ij} \sim Binomial(p_{ij})$$

$$logit(p_{ij}) = x'_{ij}\beta + z'_{ij}\gamma$$

Where U_{ij} is a dichotomous variable referring to the occupancy status of the housing unit; p_{ij} is the probability that the *i*-th housing unit in the *j*-th UGM is occupied; $x'_{ij}\beta$ represents the fixed effects considered in the model; and $z'_{ij}\gamma$ represents the random effects. The prior distributions of the parameters in the Bayesian models were considered non-informative. This model makes it possible to estimate the number of unoccupied housing units in each subdivision of the country's geographic grid.

6/ The description of this methodology was taken in its entirety from the technical note prepared by ECLAC, CELADE, and UNFPA in order to avoid any distortion of the procedures carried out.

7/ Among the variables considered in the models are: area (urban, rural), level of danger, access issues, risks or threats, telecommunications coverage, construction permits, population registered in the health system of the Costa Rican Social Security Fund, population enrolled in primary education, proportion of urban land, proportion of cultivated land, informal settlements, elevation, territories belonging to indigenous groups, protected areas, average travel time to schools and churches, nighttime light intensity, among others.

2. In the second step, it was necessary to predict the number of people living in housing units that were not visited, were occupied but whose inhabitants were absent, or that refused to answer the census questionnaire. The Bayesian model used in this work is similar to the models from the WorldPop Project at the University of Southampton in England. The fundamental assumption of this model is that the number of people in a given UGM follows a Poisson distribution. A basic description of the model is as follows:

$$Y_{ij} \sim Poisson(N_{ij}, D_j)$$
$$log(D_j) = x'_{ij}\beta + z'_{ij}\gamma$$

Where Y_{ij} is the number of people living in the *i*-th housing unit of the *j*-th UGM, N_{ij} is the number of housing units in that UGM (known from the census and pre-census), and D_j is the average population density in the UGM, which is related through the logarithmic function to the corresponding fixed effects $x'_{ij}\beta$ and random effects $z'_{ij}\gamma$. Likewise, the prior distributions of the parameters in the Bayesian models were considered non-informative. This model makes it possible to estimate the number of unoccupied housing units in each subdivision of the country's geographic grid.

3. In the final step, modeling was carried out on the demographic structure of the UGMs at the provincial level. This structure is represented by 40 population subgroups, resulting from the cross-classification of sex and five-year age groups. Inference on the unobserved housing units is performed through a conditional prediction based on the results obtained in the previous step. Due to the nature of the phenomenon under study, a multinomial distribution was considered appropriate as the likelihood function for the count across the 40 groups. A description of the model is presented below:

$$G_i \sim Multinomial(\boldsymbol{p}_{G_i})$$
$$logit(\boldsymbol{p}_{G_i}) = x_i\beta$$

Where G_i is the count for each of the forty population groups, and $\boldsymbol{p}_{G_i}$ represents a 40-element vector containing the probabilities that a person residing in a housing unit belongs to each group.

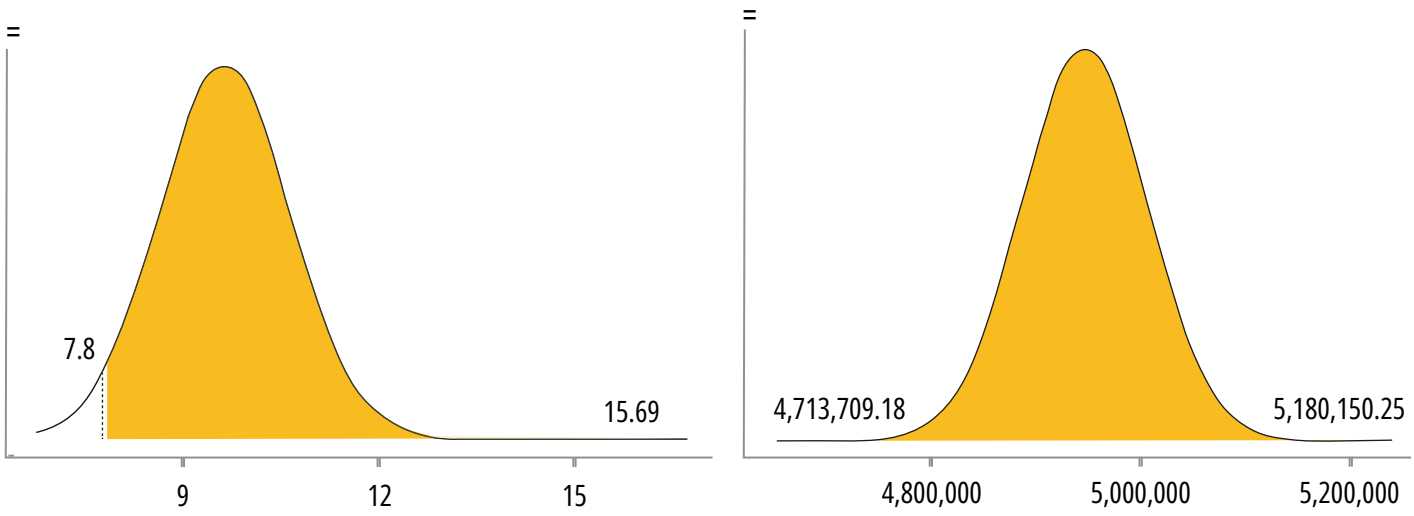
In defining each of the models, the following steps were taken:

- Data collection: This included the number of housing units observed in each UGM during the pre-census, census data from both the in-person and virtual operations, geospatial information, and administrative records.
- Data preparation: In this stage, the collected data were analyzed and assessed for quality and suitability for analysis. This included cleaning the data, standardizing formats, removing missing or inconsistent data, and transforming and standardizing the data for further processing.

- Consolidation of information types: In this step, each housing unit was classified into one of the following six categories, i.e., unoccupied (7.36%), occupied (43.7%), unoccupied enumerated on paper (0.08%), occupied enumerated on paper (1.32%), occupied without information on the number of inhabitants (26.89%), and with no information of any kind (20.66%). The prediction of the number of inhabitants (Poisson model) was applied to the last two categories, while the prediction of housing unit vacancy (Binomial model) was applied to the first category.
- Model definition: The Bayesian models describing the relationship between observed variables and covariates (fixed and random effects) were then defined. This involved selecting appropriate probability distributions to represent the uncertainty and the relationships between variables.
- Parameter estimation: Once the model was defined, its parameters were estimated using R and STAN software with Bayesian inference techniques (Markov Chain Monte Carlo, MCMC). These techniques make it possible to obtain parameter estimates and their credibility intervals.
- Model validation: At this stage, the quality and accuracy of the developed model were evaluated. This involved comparing the model's estimates with independent observed data and conducting sensitivity analyses to assess how different sources of uncertainty affect the results.
- Generation of population maps and uncertainty charts: Once the model was validated, population estimation maps were generated from the data and the estimated parameters. These maps represent the spatial distribution of the population and include charts showing the model's uncertainty.

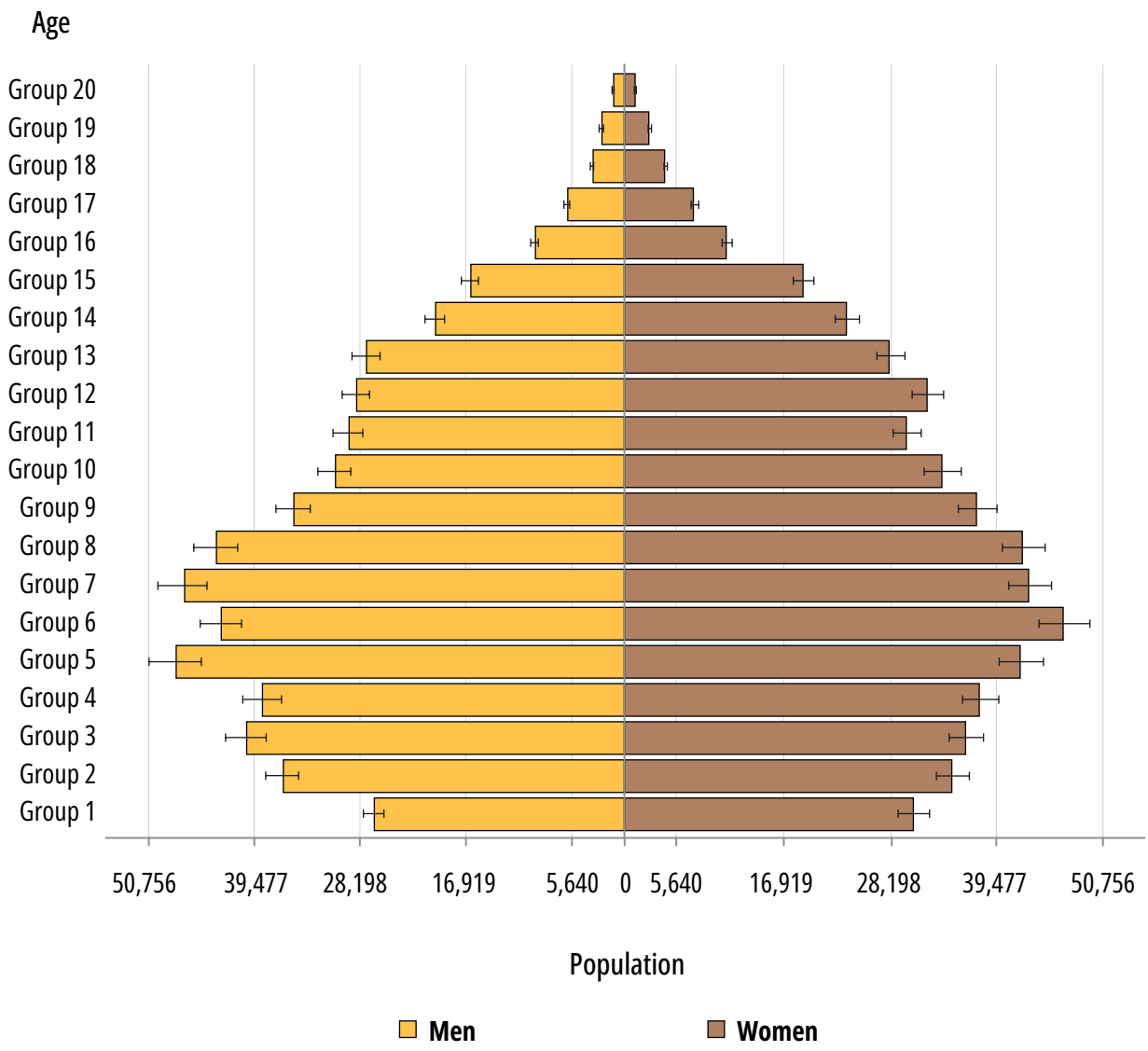
Using this type of modeling, an estimate of 170,596 unoccupied housing units was obtained, representing 9.65% of the total housing units in the country, with a credibility interval between 7.80% and 15.7% (Figure 4.2, left). In this joint effort with the INEC, the total number of inhabitants in Costa Rica was estimated at 4,946,930, with a credibility interval between 4,713,709 and 5,180,150 (Figure 4.2, right). Finally, the distribution by age and sex is shown, with corresponding credibility intervals for each of the 40 groups (Figure 4.3). Through this inferential approach, it is possible to disaggregate the estimates to the provincial, cantonal, and district levels.

FIGURE 4.2
Costa Rica. Estimation of the number of unoccupied housing units (to the left) and the number of inhabitants (to the right)



Source: Prediction model for the Costa Rica Census 2022.

FIGURE 4.3
Costa Rica. Population structure estimation (men and women in five-year age groups)



Source: Prediction model for the Costa Rica Census 2022.

The work and results of the developed models provide inputs that the INEC incorporates into its process for verification and comparison with the various estimates independently produced by its technical teams.

4.4 Validation of results and definition of final population estimate

The process of producing the population estimates was carried out in an integrated manner by three working teams, with the valuable support of CELADE and the CCP, whose experts advised the INEC team on the implementation of the calculations. In addition, the Bayesian model was developed in collaboration with experts from the Statistics Division of ECLAC and UNFPA. The INEC teams were made up of personnel from the Demographic Statistics Unit, the Cartography Unit, and the Census Project 2022. Throughout the process, there was constant validation of the partial results through working sessions, both internally and with international experts.

At the conclusion of the application of the methodologies, the three estimates were compared with the results observed in the Census 2022 to identify possible inconsistencies in the distribution of the population by canton, as well as by age and sex groups. The confidence intervals generated by the Bayesian model were also reviewed, evaluating at the cantonal level how the estimates behaved.

Although all three estimates are considered robust and define a range of estimated population figures, it was determined that the final population figure would be the one obtained through the demographic methodology based on records. This is because it relies on observed data from administrative records, which have already been previously evaluated and are known for their completeness and quality. Furthermore, this is the information that will serve as the basis for calculating future population projections.

4.5 Housing estimation

Once the estimated population was defined, an analysis of housing was carried out. For this purpose, the average number of occupants per housing unit, estimated and validated for the methodology based on census data, was used as a reference. As mentioned previously, these averages had been thoroughly analyzed and compared with various recent and historical sources, and it was determined that they were the most appropriate.

Based on this, the number of occupied housing units required for the total defined population was estimated using the calculated average occupancy rates. In this way, a total of 1,621,381 occupied housing units was estimated.

From the validation of the housing units recorded in the census and the analysis carried out on the distribution of unoccupied and unvisited housing units —as explained in the census-based estimation methodology— it was determined that there were 1,554,171 occupied housing units in the Census 2022, thus revealing an undercount of 67,210 occupied housing units.

For unoccupied housing units, the same set of housing units analyzed and classified in the census-based estimation methodology was used. This included housing units identified as unoccupied in the Census 2022, the reassignment of housing units classified as unoccupied for other reasons, and housing units not visited, as well as those located in UGM areas surveyed over more than 20 days.

In addition, in 9 cantons[8/] only the unoccupied housing units identified in the Census 2022 were retained, as they represented significant quantities that no longer required further adjustment. The total number of unoccupied housing units was 214,654.

Finally, the total number of housing units in the country was estimated at 1,836,291.

8/ Escazú, Santa Ana, Montes de Oca, Curridabat, Jiménez, Barva, Santo Domingo, Liberia y Carrillo.

Bibliography

Cebrián, E. (2011). *Desafección, apatía y participación. La desafección activa como forma de participación política*. *Revista de Gestión Pública y Privada*, 71-80.

Republic General Comptroller's Office. (July 2022). *CGR.go.cr*. Recovered from https://sites.google.com/cgr.go.cr/ipp2020-situacionyperspectivas/implementaci %C3 %B3n-de-la-regla-fiscal-en-costa-rica

General Directorate of Migration and Alien Affairs. (2022). *Base de datos de entradas y salidas migratorias*. Costa Rica.

Hillygus, Nie, & al., e. (2006). *The Hard Count.* New York: Russell Sage Foundation.

INEC. (2000). *Censo 2000*. Costa Rica.

INEC. (2010-2022). Continuous Employment Survey. Costa Rica.

INEC. (2010-2022). National Household Survey. Costa Rica.

INEC. (2011). Census 2011. Costa Rica.

INEC. (2014 and 2018). National Household Income and Expenditure Survey. Costa Rica.

Department of Vital Statistics. (June 2022). Electoral roll. Costa Rica.

Department of Vital Statistics. (Master as of June 2022). Deaths Database. Costa Rica.

Department of Vital Statistics. (Master as of June 2022). Births Database. Costa Rica.

Romano, A. (2015). *La apatía política en democracia*. Ecuador: FLACSO.

INEC, 450 meters west of La Bandera roundabout, Los Negritos street,
Ana Lorena Building, Mercedes, Montes de Oca, Costa Rica.
Email: solicitud.datos@inec.go.cr P.O. Box: 10163 - 1000 San José, C. R.
Telephone: (+506) 2527-1144, 2527-1145, 2527-1146, 2527-1147 y 2527-1124.