ask
AGJENCIA E STATISTIKAVE TË KOSOVËS
AGENCIJA ZA STATISTIKE KOSOVA
KOSOVO AGENCY OF STATISTICS

Republika e Kosovës
Republika Kosova - Republic of Kosovo
Qeveria - Vlada - Government
Zyra e Kryeministrit -Ured Premijera -Office of the Prime Minister
Agjencia e Statistikave të Kosovës - Agencija Statistike Kosova - Kosovo Agency of Statistics

ReKos 2024

POPULATION AND HOUSING CENSUS IN KOSOVO

2024

POST ENUMERATION SURVEY

ask
AGJENCIA E STATISTIKAVE TË KOSOVËS
AGENCIJA ZA STATISTIKE KOSOVA
KOSOVO AGENCY OF STATISTICS

Republika e Kosovës
Republika Kosova - Republic of Kosovo
Qeveria - Vlada - Government
Zyra e Kryeministrit-Ured Premijera-Office of the Prime Minister
Agjencia e Statistikave të Kosovës - Agencija Statistike Kosova - Kosovo Agency of Statistics

ReKos 2024

POPULATION AND HOUSING CENSUS IN KOSOVO

2024

POST ENUMERATION SURVEY

Population and Housing Census of Kosovo 2024

Post Enumeration Survey: Main findings

Preface

The quality of population and housing census data is very important for many reasons, including building public trust in and understanding of the national statistical system. The purpose of census evaluation is to provide users with a measure of the level of accuracy and confidence when utilizing the data, and to explain any errors in the census results.

In recognition of the importance of census data, providing valuable information for analyzing changes in the structure and demographic and socio-economic profile of the population, it is widely recognized that for all censuses, overall quality of population and housing census results should be evaluated. Kosovo Agency of Statistics (KAS) conducted the Post Enumeration Survey shortly after the 2024 Population and Housing Census to estimate main quality indicators for the census results.

The second post-enumeration survey (PES) of Kosovo was conducted in 24 June-13 July 2024 using Dual System Estimation Method as recommended by the United Nations. Like the first PES in 2011, its main aim was to provide information on the completeness of census coverage and the quality of the basic data collected in the census. It involves an independent re-capture of a statistically designed sample of private dwellings and the people within them covered by the national census. The PES responses are linked to the census records to find out who was missed by the census, counted more than once, or counted in error.

The KAS is pleased to present in this document the PES results pertinent to evaluating the coverage and content errors of the 2024 Population and Housing Census. A detailed description of the adopted survey methodology is also incorporated.

The KAS would like to thank to the partner Ministries, Institutions, United Nations Development Programme for their respective great support and inputs throughout the process of implementing this survey. I am also equally grateful to the staff of the KAS, who tirelessly worked hard to ensure the survey is successfully implemented.

Highlights

A post-enumeration survey (PES) was conducted in Kosovo, following the 2024 Population and Housing Census, to estimate the completeness of census coverage as well as for estimating content errors for key census variables. The main findings of the 2024 PES were:

- Net coverage error rate for the 2024 Population and Housing Census was about 2.47 percent representing that the census has over counting of the resident population. Net coverage error rate was 2.65 for urban and 2.32 for rural areas;
- 4.84 percent of the resident population of Kosovo was over-counted, meaning that some people counted more than once or should not be included in the enumeration as they were not part of the resident population. Similarly, overcounting rate was 5.43 percent for urban and 4.47 percent for rural areas;
- Under-coverage rate was about 2.49 percent of the total resident population, while it was 2.92 percent for urban and 2.25 percent for rural areas;
- Content errors indicators related to completed age of each individual enumerated in both the PES and the census data showed very high consistency. Rate of agreement in the reported age is 99.2 percent and aggregate index of inconsistency is 1.15 percent.

It should be noted that all overcount and undercount estimates have a sampling error and should be considered within the 95 percent confidence interval.

1. Introduction

A population and housing census consists of a complex series of interrelated steps, and constitutes perhaps the single most extensive, complicated and expensive operation in national statistical system. In principle, it requires counting everybody in the country on census night. The census provides valuable information for analyzing changes in the socio-demographic profile of the population, and for monitoring, planning, and decision making at the national and local level, by government, business and the general community. The census results are used as a benchmark for research and analysis. Population projections are one of the most important analytical outputs produced based on census data.

It is universally accepted that a population census is not perfect and that errors can and do occur at all stages of the census operation. Errors in the census results are classified into two general categories - coverage errors and content errors. Coverage errors are the errors that arise due to omissions or duplications of persons or housing units in the census enumeration. Also, inclusion of people who are not part of the target population of the census, such as citizens living abroad can create coverage errors. Content errors are errors that arise in the incorrect reporting or recording of the characteristics of persons, households and housing units enumerated in the census.

Many countries have recognized the need to evaluate the overall quality of their census results and have employed various methods for evaluating census coverage as well as certain types of content error. Various statistical methods are recommended by the United Nations to check the accuracy of census coverage. These include: (a) checks against demographically derived estimates, (b) comparison of census figures with administrative records and other sources, and, (c) a post-enumeration survey (PES).

In order to assess the extent of coverage and content errors in the 2024 Population and Housing Census of Kosovo, the Kosovo Agency of Statistics (KAS) has conducted a Post Enumeration Survey (PES), as recommended by the United Nations Principles and Recommendations for Population and Housing Censuses Revision 3. The PES methodology is fully in line with the international guidelines on PES conducted with the principle of independence from the census process. Estimations have been based on the Dual System Estimation (DSE) methodology which is based on capture and recapture methodology. The methodology estimates the total population.

The 2024 PES questionnaire includes questions for determining whether each re-interviewed person belongs to the census target population and defining place of residency of each person at the time of the census. Furthermore, the questionnaire collects information about demographic characteristics of each individual to measure response error.

KAS conducted the PES in 24 June- 13 July 2024, shortly after completing the field enumeration of the 2024 census. PES involved an independent re-enumeration of a statistically designed sample of all dwellings and the people within them covered by the national census. This report describes the objectives of the 2024 PES, including its scope, methodology, the information gathered and estimates of the coverage and content errors.

2. PES objectives and methodology

PES objectives

Main objective of the PES is to estimate coverage and content errors in the 2024 Census of Kosovo.

It has the following specific objectives:

- To measure under-coverage and over-coverage of persons;
- To measure coverage errors at national level and by urban and rural areas;
- To measure levels of agreement for responses to questions on selected characteristics, such as age, highest level of completed education, marital status and ethnicity.

The PES survey population consisted of the household population living in Kosovo. In line with international statistical practices, the following population subgroups were excluded from the PES:

- people living in institutional places, and
- overseas diplomats, their families and people living with them.

PES Methodology

The PES plays a crucial role in providing accurate population and helps to identify any biases or errors in the census data. The main goal of this survey is to measure the coverage of the census by identifying any undercounts or overcounts of the population. The PES is a complete re-enumeration of a representative sample of a census population followed by matching each individual enumerated in the PES with information from the census enumeration[1]. It is based on dual-system estimation whereby data from the census and the PES are matched to estimate the population size.

The PES basically involves two samples, namely, P sample and the E sample. The population (P) sample: consists of a sample of Enumeration Areas (EAs) drawn from the same target population but independent from the census, for the purpose of estimating census omissions when compared to census records. The estimate of erroneous inclusion provides a correction factor needed in the Dual System Estimate (DSE) of the true population[2]. The E sample is an enumeration sample drawn from cases already enumerated in the census, but selected for independent re-interview for the purpose of estimating census erroneous inclusions when compared to original census records. It consists of the same EAs selected for the PES.

---

[1]The United Nations Principles and Recommendations for Population and Housing Censuses, Revision.3

[2] Post Enumeration Surveys, Operational Guidelines, UN, 2010
https://unstats.un.org/unsd/demographic/standmeth/handbooks/manual_pesen.pdf

DSE methodology which is based on capture and recapture methodology estimates the total population. The DSE model is conceptualized that each person has a probability of being either included in the census or not included in the census as well as either included or not included in PES. This can be described as in the following table.

	In census	**Out of census**	**Total**
In PES	N_{11}	N_{12}	N_{1+}
Out of PES	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	$\mathbf{N_{++}}$

This table illustrates the DSE. In this table, N_{11} represents the population captured in both the census and the PES, N_{12} represents the population captured in the PES nut not in the census, N_{21} represents the population captured in the census but not in the PES and so on. The subscript + denotes summation over possible values of the subscript. $\boldsymbol{N_{++}}$ **represents the total population.**

Dual System Estimate of the size of the total population is given by:

$N_{++} = (N_{+1})(N_{1+})/N_{11}$

The DSE raises the corrected census total (where erroneous enumerations are subtracted for the census population) by the total estimate of the number of people in the PES multiplied by the total number of people counted in the census divided by the estimate of the number that matched to the census.

Principle of independence

The DSE methodology is based on the assumption that the PES is an independent data collection from the Census. In order for the PES to achieve its objectives, its processes need to be independent from the census. To ensure this independence, the PES:

- Was defined and managed by a group of persons who had no responsibility regarding the census;
- Used different field staff than the census one in the selected EAs;
- Was conducted after the census field work was completed to avoid contact between census enumerators and PES interviewers; and
- PES enumerators visited all buildings/dwellings in the sample EAs independently from the enumerated buildings/dwellings in the Census.

The 2024 PES used more tightly controlled collection procedures, and more experienced and better trained field staff than the census.

3. Data collection and PES questionnaire

The PES survey was carried out during 24 June and 13 July 2024, following the completion of census fieldwork. The survey period was chosen to avoid overlap of census enumerators and PES interviewers in the field, while being close enough to census date (5 April 2024) to assist respondent recall. Data was collected by 50 specially trained interviewers using a tablet application for the household questionnaire, and their work was monitored by 9 supervisors. Information on occupants of the dwelling who satisfied the scope and coverage criteria was collected through a face-to-face interview.

In total, 16 601 people responded to the survey in the selected 50 enumeration areas. Non-response rate due to refusal and closed dwellings was relatively high compared to refusal and closed households observed in the census. The PES was conducted in one month after the census and it was very close to the end of the school term, therefore a number of households were not present in their residence address during the PES.

Personal details sought on the PES questionnaire included: name, sex, date of birth or age, ethnicity, completed level of education and address. Besides usual address at the time of the PES and on the census night, the survey also collected information on any other address where the person might have been included on any other census form. This was to help increase the chances of finding and matching any individual census forms for a particular person, and to help identify multiple counts.

4. Sample Design

A sampling frame that was used for the selection of the PES enumeration areas obtained from the 2011 Census of Kosovo which was updated during 2018, 2019 and 2020. For the purpose of the census, the sampling frame has been divided into the enumeration area (EA), which represent relatively small operational segments defined for enumeration.

There are 4673 EAs for Kosovo and these are used for stratification of samples for various statistical surveys. Geographically, Kosovo is divided into municipalities and settlements, which are then divided into cities or localities. Enumeration Areas are defined within the settlements.

The sample size for PES was 50 enumeration area (EA) or 1 percent of the total population.

The sample designed was stratified with two factors: “region” and “urban/rural” type of area. The primary sampling units were enumeration areas. Selected number of EAs by the sampling strata is presented in Table 1 below.

Table 1. Selection of sample EAs to sampling strata

Regjon	Total EAs in the sample	Urban EAs	Rural EAs	Households in the sample
Prishtinë	12	6	6	1,058
Mitrovicë	6	3	3	509
Pejë	6	3	3	555
Prizren	10	5	5	766
Ferizaj	6	3	3	645
Gjilan	6	3	3	464
Gjakovë	4	2	2	299
Total	50	25	25	4,296

5. Matching

The objective of matching was to determine if PES respondents were counted in the census at each address at which they stated that they had completed a census form, or at each address where a census form may have been completed for them (search address). This was achieved by comparing the information given by PES respondents with the information given on census forms. If an address given was different to the PES address, searching was carried out to locate the address in census before matching could be attempted.

Matching was performed in two phases: first automatic matching was implemented, after that for the remaining unmatched individuals, computer- assisted matching was implemented by comparing the PES and the census data at household levels.

Matching determined whether a person was counted or not counted in the census at each search address and later on other addresses within the same EAs. The most important variables available for comparison were:

- Personal Identification Number (PIN)
- Name
- Surname
- Father name
- Date (day, month and year) of birth (or completed age)
- Sex

Following stages were implemented for matching persons enumerated in the PES (P sample) and in the census (E sample):

First stage of matching: All individuals enumerated in the P and E samples were matched with unique PIN. Almost 90 percent of persons enumerated in the P sample was successfully matched using the PIN with persons enumerated in the E sample.

Second stage of matching: For persons who were not matched in the first stage with PIN, the following matching rules were used within an EA:

a. Name, Last Name, Year of Birth and Month of Birth;
b. Name and Date of Birth (Day, Month and Year);
c. Last Name and Date of Birth (Day, Month and Year); and
d. Father Name and Date of Birth (Day, Month and Year).

While matching persons enumerated in the P and E samples, enumeration status also assigned to matched individuals: non-movers, in-movers, out-movers and out of scope (for births and deaths occurred after the census reference day and persons who moved to another country after the census reference day).

Reconciliation visits

Following the initial matching phase, as recommended by international guidelines[3], reconciliation visits were carried out in sample EAs. Main reason for conducting the reconciliation visits was to check whether households that refused to participate in the PES were actually residing at that address in the census. Also, the reconciliation visits aimed to identify whether dwellings that were closed during the PES were enumerated during the census or not.

Such visits gave an opportunity to identify erroneous census enumerations and the resolution of doubtful cases in order to achieve a realistic and definitive match status for every person enumerated in the P and E samples.

For persons who appeared in the census records but not in the PES, reconciliation visits permitted:

a. determination of whether such persons were usual residents as of census date (i.e. correctly enumerated) or
b. whether they were erroneously enumerated.

In the case of persons appearing on the PES list but not on the census record, reconciliation visits helped in:

a. Confirming whether such persons were usual residents as of the census date, in this case non-movers or out-movers or omission
b. whether they arrived or were born after the census (in-movers)

The results of the reconciliation visits helped in making final decisions with insufficient information for matching.

[3] Post Enumeration Surveys Operational Guidelines, UN, 2010
https://unstats.un.org/unsd/demographic/standmeth/handbooks/manual_pesen.pdf

6. Analysis of coverage and content errors

The information collected from the PES (P) sample and the census (E) sample was estimated using weights to provide an estimated net coverage error and estimated undercount and overcount errors for the whole population. In 2024, a household's 'weight' was determined initially by their probability of selection into the PES sample.

Table 2 shows the coverage errors for the total of the country and for urban and rural areas. Overall, it was identified that the 2024 Census coverage rate was very high, and this reflects the cooperation and support of the Kosovo residents. The net coverage error rate for the total of the country was about 2.47 percent. This rate was slightly higher for urban areas which was about 2.65 percent compared to the rate of rural areas which was about 2.32 percent.

The 2024 PES estimated that about 4.84 percent of the residents of Kosovo were overcounted while the same population was undercounted for about 2,49 percent. For urban areas, the PES showed that while the over coverage rate was about 5.43 percent, the under-coverage rate was estimated as 2.92 percent for the same areas. Similarly, the overcounting rate for rural areas was about 4.47 percent and undercounting rate was 2,25 for the same areas.

It should be noted that all coverage errors were calculated using weighted results of matching the PES and the census data, therefore adding estimates of under and over coverage rate showed a slightly different rate compared to the net coverage error rate.

Table 2- Coverage rates at national, urban and rural levels

Net coverage error rate (%)		
Total	Urban	Rural
2.47	2.65	2.32

Under-coverage rate (%)		
Total	Urban	Rural
2.49	2.92	2.25

Over-coverage rate (%)		
Total	Urban	Rural
4.84	5.43	4.47

Estimation of content error uses unweighted "matched" persons that are identified in the census and PES data. For this report, content error analysis has been performed only for age information. Main aim is to compare the census responses with the PES responses and to calculate the measures of consistency. Content error for the variable "age" is measured using three indicators: Index of inconsistency (by age groups), Aggregate index of inconsistency and Rate of agreement.

Index of Inconsistency is the relative number of cases for which the response varied between the census and the PES. This index is calculated for each response category.

$$\hat{I}_i = \frac{\left(x_{\bullet i} + x_{i\bullet} - 2x_{ii}\right)}{\frac{1}{n}\left\{x_{\bullet i}\left(n - x_{i\bullet}\right) + x_{i\bullet}\left(n - x_{\bullet i}\right)\right\}} \times 100$$

For $i = 1,2,---,s$

Where X_{ii} =
number of cases where category i was given a s response in both the census and the PES.

s = total number of response categories

Aggregate index of inconsistency is calculated for the variable as a whole.

$$\hat{I} = \frac{\left\{n - \sum_{i}^{c} x_{ii}\right\}}{\left\{n - \frac{1}{n}\sum_{i}^{c} x_{\bullet i} x_{i\bullet}\right\}} \times 100$$

Rate of agreement is calculated for the variable as a whole, it is the number of consistencies between the census responses and PES responses relative to the total number of matched cases.

$$\frac{\sum_{i}^{s} X_{ii}}{n} * 100$$

Table 3 presents the results of the content error indicators calculated for Completed Age

	Age group (reported in the PES)			
Age group (reported in the CENSUS)	0-19	20-45	45+	Total (CENSUS)
0-19	**4468**	63	3	4534
20-45	2	**5804**	44	5850
45+	1	10	**5700**	5711
Total (PES)	4471	5877	5747	**16095**

Index of inconsistency	
Age groups	
0-19	1.06
20-45	1.60
45 +	0.79

Aggregated index of inconsistency	1.15
Rate of agreement	99.2

Table 3 indicated that the index of inconsistency was very low for all three age groups that were used for the content analysis of Age. A high consistency in reporting the age in the census and PES shows high quality of data for Age. It should be noted that completed age was calculated automatically with the data collection application using the information on date of birth including day, month and year. Therefore, this result actually indicates that the quality of reported date of birth is very high. Similarly, aggregated index of inconsistency of 1.15 percent shows a low degree of inconsistency in the PES and the Census for Age variable as a whole.

Comparison of the PES and the Census data showed very high consistency for completed age. The rate of agreement was 99.2 percent presenting that only 0.8 percent of persons declared different age in the PES and the Census data or captured differently in the PES and the census data.