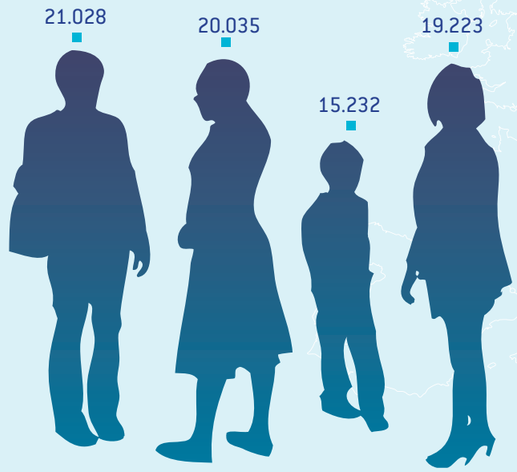
AGJENCIA E STATISTIKAVE TË KOSOVËS
AGENCIJA ZA STATISTIKE KOSOVA
KOSOVO AGENCY OF STATISTICS

Kosovo Population and Housing Census 2011

FINAL RESULTS

QUALITY REPORT



Census Project Multi-Donor Trust Fund

NOTE TO THE READER

The population and household census was carried out by door to door enumeration in April 2011, with as reference date the 31 March 2011, at midnight. It was based on a specific Law on the Census (03/L–237) approved on 7 October 2010 by the Kosovo Assembly. One of the most important provisions of the law regards data protection and confidentiality of personal information. This law also defines the obligation to conduct a census post enumeration survey immediately following the census enumeration.

Kosovo census is compliant with international methodology recommendations for the 2010 censuses of population and housing prepared by United Nations Economic Commission for Europe in cooperation with the Statistical Office of European Communities (Eurostat). All topics covered and data collected refer to internationally agreed definitions.

All the data reported in this document refer to the 2011 Census. The Census results include data from 34 municipalities. Due to objective reasons the enumeration could not be carried out in the northern municipalities, which has been recognized by the Census Trust Fund Steering Committee. The Kosovo Agency of Statistics will publish estimates for these municipalities later on.

Kosovo Population and Housing Census 2011

FINAL RESULTS

QUALITY REPORT

Technical assistance to the Kosovo Agency of Statistics for the preparation of a Population Census

An EU-funded project managed by the European Union Office in Kosovo implemented by Istat, CIRPS-Sapienza University of Rome, ICON-INSTITUT and Rrota

EUCEP 2011 CENSUS SUPPORT PROJECT

Table of Content

Foreword

As by the European Union Regulation (EU) No 1151/2010 (9 December 2010), there is a requirement to transmit together with census results a quality report containing a systematic description of the data sources used for census purposes as well as a report of the quality of the census results produced from these sources. Kosovo 2011 population and housing census is based on questionnaire-based field data-collection and its data sources have been thoroughly described in a metadata report, which complements the present one and is available for the large public.

This census quality report has as main objective to provide our users with all information required to assess the reliability of census results. It describes all steps used before, during and after the census enumeration in view of increasing data reliability and consequently its usefulness. It also presents a detail comparison analysis of census outputs with the ones obtained from other statistical surveys or administrative information. Together with the report on census coverage and content errors estimates (PES report), this report provides fully transparent and complete information on the Kosovo census outputs.

In this document, readers will find all information they require to understand what was the initial quality of data gathered by the canvassing operation, and to what extent data had to undergo editing procedures to become a coherent and representative data set to be used for further analyses and decision making.

Regarding adequacy, questionnaire-based census has been the only viable option in our Country, due to the current lack of administrative registers, on which most of the developed countries tend to base their censuses for cost-effectiveness reasons.

Completeness quality has been unfortunately hampered by the non-availability of census data for a geographical part of Kosovo – representing 3 out of 37 municipalities - where despite preparedness of the materials, insistence of international community and persuasive dialogue, it has been impossible to obtain the participation of the resident population and their municipality authorities. Future total population estimates should help overcoming problems with this census weakness.

Census accuracy is mainly measured by the errors found in the raw data and the editing and imputation process they underwent. It is clearly explained in the part 3 of this report. Accuracy is also measured with derived methodology by comparing census data with the ones obtained from alternative data sources.

Together with metadata report and post-enumeration survey (PES) report, I am confident that users have in their hands all necessary information to make a proper and valid use of our census data.

I take this opportunity to thank all persons, experts, institutions and organizations from Kosovo, from other countries and from international bodies – mainly of the European Union - who joined their efforts to ensure to Kosovo census highly reliable outputs. They are too many to be mentioned one by one, but they will all recognize themselves in the descriptions of the three main census qualitative reports elaborated by our Statistical Agency. My deepest gratefulness goes also to the principal participants and at the same time main beneficiaries of this giant undertaking: Kosovo residents who provided complete and reliable answers to the census questions. Cooperation has been the key of our census success, and it paved the way for a new area of doing and disseminating official statistics in our country.

It is also now perfectly understood that communication and service-oriented production are new standards to develop. I sincerely hope that further statistical data collection operations will be understood by all Kosovo residents and by the bodies and media in charge of informing them, because objective quantitative data are a real infrastructure for democratic societies on which each leading force can rely to argument its requirements to the Government bodies and to develop its own development strategies.

Isa Krasniqi
Chief Executive Officer

Introduction

Because of the complexity of the census, it is likely that errors of one kind or another may arise at any stage in the census operation. Census errors can arise because of many reasons such as poor planning/ implementation of demarcation of enumeration areas, varying interpretation of questions by enumerators and/or respondents, unwillingness or inability to give correct answers, non-response, data capture errors and coding/editing errors. While such errors are unavoidable to some extent, the value of census outputs can be enhanced if the quality of the data is properly understood.

The main objective of studying census evaluation is to provide both the producer and the user of the data valuable information regarding the quality of census data. Eventually census evaluation can contribute to building public trust and to better understanding of the value and limitations of census data which may minimize misinterpretation of census results.

The quality of a census depends on preventive measures aimed at quality assurance. Good training and organizational procedures for data collection or procedures and techniques to limit inconsistency errors during the data capture can contribute to a better data quality. These aspects are covered in the two first chapters.

Ex-post, when data capture is completed error localization methodology is applied to detect systematic and random inconsistencies that could not be detected during the data capture process. This leads to a long and complex process of data editing and imputations, whose final goal is to get a census database with consistent and coherent micro-data. This phase is described in the third section.

Finally, census evaluation is carried out to identify coverage and content errors with various methods. The ones applied to Kosovo census are two: the post enumeration survey (called PES) and demographic techniques.

A separate report is dedicated to census error estimates based on the PES since this methodology implied specific statistical methodology, which deserves its own description. The quality of census outputs assessed with demographic techniques is described in the last chapter.

CHAPTER 1

Measures taken to ensure quality data collection

Before the census took place, a set of measures have been developed and introduced by The Kosovo Agency of Statistics (ASK) in order to guarantee the quality of the most important phase of the census: gathering complete and trustful data by all Kosovo inhabitants.

1.1 Organizational measures

1.1.1. Enumeration method

Census in Kosovo was conducted according to traditional method (door to door) where enumerators trained on purpose filled-in the census forms. Due to the complexity of some questions and in order to achieve harmonized and coherent answers, allowing questionnaires being directly filled-in by the respondents would have increased the risk of incorrect responses and would have had as a consequence a loss of data quality.

1.1.2. Selection of the data-collection staff

The Kosovo Agency of Statistics (ASK) defined specific selection criteria for the enumerators, the controllers as well as for the supervisors. These were based on the education level, experience, knowledge of local context, ethnic and gender fair representativeness. According to the Census Law, the enumerators were selected by the Municipal Census Commissions (MCCs) set-up in each municipality while controllers proposed by the MCCs were selected in cooperation with ASK. Notification on the competition was published in three (3) main newspapers (in official languages including also the Turkish language) as well as in electronic media, while the competitions' announcements were done in each municipality.

Since the enumerators were the contact persons with the households, ASK recommended to the MCCs to select field enumeration staff favouring those who were acquainted with the area they had to enumerate and speaking the language of its inhabitants. In Enumeration Areas (EA) inhabited by mixed communities, the number of enumerators was in proportion to the various languages spoken by the residents. Each enumerator was provided with census material including maps, questionnaires in the EA languages, ID card as well as other materials to facilitate his/her recognition as official census staff by the public (census bag; census marked apron vest).

1.1.3. Trainings of the field staff

Trainings and training modules were given particular importance. It was understood that the performance of field work and the quality of data collection was depending from the quality and care provided to the whole training procedures.

Trainings were organized in a top-down cascade form.
First trainings were held for a group of 21 State Instructors who were mainly statisticians from ASK playing a major role in the census content supervision and in the training of the census field staff. A first training session of instructors lasted three (3) days at the beginning of February 2011 and a second session organized mid-February 2011 lasted four (4) days.

Training for Municipal Census Commissions (MCCs) - were held in nine (9) training centres (across Kosovo regions) from 01-03 February 2011. Trainings were held in both Albanian and Serbian languages.

Training of Supervisors took place during the period from 7 to 11 March 2011. These trainings were held in Prishtinë/Priština in four (4) training centres. Trainings were held by State Instructors.

Trainings of controllers were conducted from 14 to 18 March 2011. These trainings were conducted in 30 training centres (usually in their respective municipalities).

Trainings of enumerators were held from 21-25 March 2011, in their respective municipalities. They necessitated 161 training centres (classes). Usually, trainings for enumerators were held in schools, while a part of them was done in other municipal facilities. All training centres have been equipped with projectors and other auxiliary equipment. Trainings for enumerators initially planned to last three (3) days finally lasted five (5) days, to ensure a better mastering of the tasks by the enumerators. In the last part of trainings enumerators practiced to fill-in the census forms. Trainings were conducted all over Kosovo and were all done with the same training materials, instruments and methodology; each training centre was equipped with projectors.

1.1.4. Measures to ensure the coverage of all residents

Although it was proposed in the instruction manuals that the enumerators make at least three (3) trials to get in contact with the inhabitants of a household, during the trainings it was recommended, if needed, to visit the households more times in case of absence. Also, before the census start, each enumerator visited its enumeration area to get acquainted with its geographical limits and to update the existing building and dwellings existing. During this preliminary visit the enumerators were given as a task to leave at each dwelling a notification letter informing the residents on the census

process. During the census, if a dwelling was found empty, enumerators were instructed to leave a note with their contact telephone number asking the eventual inhabitant to contact them and agree about the timing for the interview.

1.1.5. Control measures to ensure quality enumeration

All municipalities had provided dedicated working space where the checking of census material returning from the field was done daily. So, since the first day the enumerators met with the controllers to check the materials or to discuss about the possible problems in the field. Later, (during the day) controllers had to meet the supervisors and after supervisors met with the Municipal Census Commission members. This procedure was followed in all municipalities till the end of the field work. These meetings helped the process by offering harmonized solutions for different problematic cases encountered. The control procedures also allowed monitoring closely the enumeration process and fostering the inclusion of all the households in the enumerated areas.

ASK had established a support group in Prishtinë/Priština which, within several hours, was able to support the work in the field, providing methodological clarifications needed, taking decisions or sending additional materials requested.

Working groups including the coordinators met every evening in Prishtinë/Priština to report on the fieldwork progress and inform about the further steps. The municipalities also reported daily to ASK about the field coverage.

ASK had taken all necessary steps and implemented many organizational measures with the purpose of having a qualitative and comprehensive census.

1.2 Census data collection forms

The most delicate and important element to ensure census quality has been the proper design of the questionnaire. The highest care was taken to ensure that questions were well understood, that they were clear to the interviewed person and not ambiguous. Structure of the questionnaire and its design were also given the utmost attention in order to facilitate an easy and correct reporting of the information provided. Questions addressed only to some persons were clearly identified, for example questions on fertility that were to be asked only to women of a certain age. All questions were numbered and logical paths of the interview were precisely indicated by arrows referring to a specific further question. For instance, if a household responded having no land in ownership, it did not have to answer on all questions regarding the land use; these questions were skipped by a clear instruction indicated on the form.

To facilitate the enumeration, forms have been divided in three categories: questionnaires on the housing unit; questionnaires about the household; and individual questionnaires. Each of these forms was printed on a specific coloured background, for an easy identification.

During 2006 and 2008, ASK had realized two pilot censuses aiming to identify which questions needed to be explained or modified. After these pilot censuses a number of questions were adjusted to increase their intelligibility and some other questions were added to the forms. Experts in censuses also revised the various linguistic versions of the questionnaires to ensure they were perfectly harmonized and that the translations did not distort any question.
In addition to the census forms, additional forms were also used in order to collect data on the progress of the enumeration and produce preliminary census results. During the census the enumerators had to fill-in the following additional forms:

Buildings List (BL): where enumerators reported for each building of the EA the use of the building (for habitation or not), its number of floors and other useful information.

Daily Summary form (DS1): this form was prepared for the enumerators in order to report daily about the enumeration progress in their Enumeration Area. By using this form they could capture information on each dwelling visited, such as its status (occupied or vacant); the number of households living in it; and the residents found in each household, by gender. In case of refusal of participation, enumerators were asked to indicate the refusal in a specific part of the form reserved for comments. It appeared after the census that not all enumerators followed the latter instruction. Indeed, in some areas having face refusals of participation from certain communities, enumerators did not report on the refusal, not wanting to bear responsibility for the concerned to be eventually fined as there were by law administrative sanctions foreseen for these persons.

Record and summary list for enumerators (RS1): on this form each enumerator compiled total numbers of each type of units he/she enumerated during the census such as: number of buildings in his/her Enumeration Area, number of inhabited/not inhabited dwellings, number of households, number of persons by gender male/female. This list at the same time was as a controlling list for other basic forms and these lists were a great contribution for the controllers, supervisors. They based the production of the census preliminary results.

These summary data were further compiled in a cascade way on other forms totalizing data at Controller level, Supervisor level up to municipality level (respectively forms RS2; RS3 and RS4).

RS forms have been the main control instrument for monitoring the quality of the enumeration progress.

Numerous examples in the various instruction manuals (*provided often in illustrative form easy to understand*) also participated to increasing the quality of data collection on the field. ASK did not receive any complaint about the legibility, readability or about other forms of questionnaires and manuals.

1.3 Census mapping

Given that settlements vary considerably regarding the territorial extension and the population density and size, it was impossible to use settlements as enumeration units assigned to one interviewer. Other methods were required to define an enumeration area. Kosovo lacks administrative data such as population registers and it has no physical address system, which would facilitate the identification of locations on the field. As a consequence ASK, since its first census preparations in 2005, started an ambitious project to establish a Geographic Information System (GIS). Building such a system was a challenge in itself as it required means and the development of appropriate skills. The system has been using satellite images of the territory.

So, by mid of 2005, ASK built a solid GIS database where were divided settlements into smaller areas (for census needs only), which were called Enumeration Areas (EAs). Depending on buildings' and dwellings' density and on their distribution over a settlement, enumeration areas were divided in three (3) categories: urban, rural and hilly/highland.

During the two pilot tests it was acknowledged that one enumerator could cover the entire assigned area in two (2) weeks.

For data collection purposes enumeration areas were divided into:
Urban EAs which included approximately from 80-120 households (dense habitat);
Rural EAs including from 40-60 households (medium habitat); and
Hilly/mountain EAs including up to 40 households (scattered habitat).

Changes in the administrative territorial division in the period 2007-2010 (creation of new municipalities) obliged ASK to review the initial EAs definition in the GIS system and to modify orthogonal-coordinates, objects codification, borders of Census areas, legend descriptions on the maps, and many more time-consuming operations. ASK also updated satellite images and field information on the buildings and population characteristics during this period, which greatly facilitated the selection of the personnel, as well as the distribution of material based on ethnicity of people living in each area.

According to the latest data, the territory of Kosovo was divided into 4,643 EAs.

Mapping achievements are well appreciated by the EU and the ASK Cartography & GIS Unit was officially admitted with full rights as a new member in the European Geo-statistics forum; an achievement that the regional countries did not yet reached.

All census field staff has been further equipped with colour satellite maps covering the EA of interest, showing the borders and buildings present in it. EAs were respecting the cadastral boundaries so that no EA was covering two distinct settlements, two rural EAs could however not respect this rule.

For the census several types of maps were printed:
Municipal Census Commissions maps (map of municipalities), map for supervisors, map for controllers map for enumerators, coordinators and other staff map and map for the International Monitoring Operation.
A working group of ASK GIS department provided the field staff with additional maps whenever needed during the enumeration process.

1.4 Share of responsibility

In the Census Law the role and responsibilities of institutions involved in the census project were specified. Thus, with the approval of the new Census Law (No. 03/L-237), the Central Census Commission (CCC) became the highest institution for overseeing the census.

This commission was elected by the Kosovo Assembly, on the proposal of the Government. According to the Law, the CCC was a special body established to administrate and supervise the preparation, organization and conduction of the census. In addition, pursuant to the Law the CCC members representing various institutions were determined.

The role and duties of the commission were:

1. To adopt regulation of work for the enumeration process; adopt questionnaires and census documentation, proposed by the competent institution of the official statistics;
2. To directly supervise all stages of the preparation and census conduct, and advice the competent authority for official statistics on the information, required by public authorities, scientific institutions and civil society. Ensure coordination of work between public authorities, scientific institutions and civil society;
3. To approve members of municipal census commissions of census proposed by the respective municipalities, as well as to define their duties, responsibilities and other tasks.

CHAPTER 1 Measures taken to ensure quality data collection

In the middle of the second week of January 2011 the CCC approved the census forms.

At the end of January 2011, municipal census commissions were established in 34 municipalities of Kosovo (out of 37). Also, in January 2011, two meetings took place with representatives of municipalities, to make them familiar with their duties and their role in the Census process.

With the approval of the new Census Law (No. 03/L-237) members of the CCC were changed. This change was made at the end of March 2011 (one day prior the start of the census) and of course new members did not have adequate access to the Census process.

CCC's role was unique, because it gathered representatives from key institutions in Kosovo, including local and international organizations.

Several weeks prior the census some CCC members participated together with other ASK staff in the public awareness campaign. They also participated in various interviews, gave public statements, took part in media debates, etc. During the enumeration period, CCC received short daily reports from ASK about the census progress.

Municipal Census Commissions (MCC) established in 34 municipalities in the end of January 2011 were composed of 9 members: the Mayor, 6 other representatives of the municipality (departments or sectors), one representative of local NGOs and one representative of ASK.

MCCs were entrusted with the duty to organize, implement and realize the field part of the census. The role of each MCC member was also clearly defined and linked to the usual responsibility and competence field of the concerned person (coordination, logistics, information, administration, methodology, etc.). As by the recommendations made by the CCC, in many municipalities MCC staff was in the field to closely monitor field staff performance. During the census, daily exchanges of information and communication were established between the MCCs and the ASK census headquarters in Prishtinë/Priština

Recommendations from the ASK census office were supported by MCCs all along the enumeration time. ASK census staff has been operating outside of regular working hours and during weekends, with the aim of sending-receiving different information, as well as to monitor the process.

Integrating the best experiences of other countries, Kosovo census ensured a census organization likely to ensure a data-collection of the best possible quality.

1.5 Public awareness campaign

ASK, with the support of international experts who have provided support and information awareness component prepared a strategy for public awareness and information, of which the main objectives were to sensitize and inform the public about the purpose of census and the enumeration method, disclosing key messages about the interest of the census for individuals as well as for local, central and international bodies. Through this strategy, despite the political developments in Kosovo (parliamentary and presidential elections[1]), it was hoped to foster the large participation of Kosovo residents in the census operation.

In order to inform about the population Census, awareness campaign used all possible forms of communication means such as television, radio, posters, billboards, brochures, leaflets, debates, interviews, articles, etc. In addition, ASK had established an internal working group, which only dealt with the awareness campaign. During the field campaign, its members also visited the municipalities of Kosovo, including a large number of rural settlements. MCCs and local NGOs have been precious partners in the public awareness campaign, using various channels and methods to inform the population and encourage positive views on census participation.

Some informative leaflets, questions-and-answers and other large public documents have been produced and largely disseminated in the country.

The awareness campaign was divided into three phases:

- The first phase, which had a slower dynamics, was conducted during the period September to December 2010. This phase was a preparatory one, characterized by meetings, workshops, and several debates. During this stage, the basic awareness campaign documents were prepared;
- The second phase of the awareness campaign included the period January-February 2011. During this period instruments necessary for sensitization were finalized and different groups were established, which born responsibility of the awareness campaign. Also, during this phase the distribution of awareness materials was done.
- The third phase of the awareness campaign included the month preceding and covering the census period from March to April 2011. During this period were hosted various television debates, interviews, articles and meetings, sensitizing the population to participate in the registration.

During the awareness campaign, a group of staff from ASK and external collaborators, who were involved in awareness campaign were focused on outreaching activities with citizens.

[1] Parliamentary elections in Kosovo developed in the mid of December 2010, while the new Government of Kosovo was formed in the second part of February 2011.

They conducted over 80 meetings in various centres of Kosovo. At these meetings, citizens got information about the census process.

After open tendering procedures organized by the census Trust Fund manager UNOPS, the awareness campaign was entrusted to a private operator, specialized in communications. Its role was to define and prepare the awareness campaign, campaign tools, format, text contents and promotional articles, preparation of video and audio clips, design, printing and distribution of campaign materials of awareness.

In all its activities, the communication company had to follow the suggestions and receive approvals from ASK and census trust fund contributors.

As an additional part of awareness and information campaign a bus tour has been organized (with inscriptions about Census - REKOS 2011), that visited 90 villages, announcing the time of census and disseminating some basic information. During this tour various gifts (T-shirts, footballs, key rings, brochures, etc.) containing census identification were distributed to the people met. The bus always carried eminent local and international personalities called “ambassadors of goodwill”, mainly known by the large public as famous scene artists, journalists or sports people.

In order to encourage minorities to participate in the census 13 non-governmental organizations working with specific ethnic communities have also been involved to participate in an outreach campaign, with a special focus on sensitizing and informing the Serb community. They represented six ethnic communities and seven mixed communities. They were selected after a grant tender managed by UNOPS.

Internet was part of the awareness campaign as well. Website REKOS 2011 was within the ASK portal on the Internet at an early stage of census preparations, providing basic information about the methodology and allowing people to ask questions about the census.

For all households, ASK had prepared basic information on Population Census in the form of informative letter distributed to all households by the enumerators during the pre-census visits.

Despite all efforts done, the success of the information campaign has been somehow mitigated as some persons – especially from the Serbian community - persisted in their unwillingness to take part in the census.

Part of the failure of the campaign probably lies in the channels used to reach the population: Kosovo Serbs are still mainly linked to Serbian origin media, while the main media used for the awareness campaign – even in Serbian language – were mainly accessed by Kosovo Albanians. Another potential cause of the difficulty to persuade the Kosovo Serbs in participating in the census was the identity-oriented shape given to the campaign by both the awareness responsible company and the media who covered the campaign mainly with political desk specialists instead of subject-matter ones. This might have instilled and supported the erroneous idea that census and Kosovo status were somehow linked.

In addition, in the enumerated areas, Kosovo-Serbs do not have strong local leaders, as they against here identify themselves more to the ones from Serbia (artists and sports figures are strong in Serbia, and they are the identity models for the majority of Kosovo Serbs). Despite the efforts and commitment of Mayors from Serbian-majority municipalities, not all its residents accepted to participate in the census.

CHAPTER 2

Quality of data capture

Quality control of the forms sent back from the field has been thoroughly prepared and applied, and many steps before, during and after the electronic data-capture were prepared in view of ensuring the data quality, integrity and confidentiality. This chapter provides a quick overview of these procedures.

2.1 Predefined controls of errors during data capture

The data entry system was designed based on the requirements and recommendations of the IMO steering committee. The basic principle adopted was that captured information had to be exactly as the one provided on the paper questionnaires. Yet, if inconsistencies were found, they were flagged. The system had the ability to track and store errors of inconsistency (logically not valid values between questions), missing errors (the questions were it was supposed to be an answer but were left empty) and Over-ride Skip Errors (the error type that occurred because of the non-respect of SKIP instructions performed moving between questions). It also was developed to enable real-time reporting on these errors. This feature of the system enabled to monitor the work done by enumerators and operators as it was progressing. In this way it was possible to correct wrongly entered data during the data entry process in case it was necessary, and at the same time identify errors coming from wrong enumeration.

The logical checks (edit rules) and other data validation checks implemented into the system were prepared by subject-matter specialists. Except the logical data checks the system included another monitoring tool that was developed in VISUAL STUDIO .NET 2010, that enabled to make some special checks and corrections based on the error reports. These checks were done by data entry controller. The Data Entry Controller checked the total figures of each EA and compared them with the figures given by Municipalities, manual counting figures, RS1 for dwellings, households and persons.

The controller systematically checked the compliance of entered data with paper form of all R3 questionnaires having more than 5 errors. These cases were checked manually to ensure that these errors stemmed from fieldwork; in the contrary cases the errors were classified as typing errors and were corrected.

The application for the controller ensured that all questionnaires present in an EA box were entered, and checked the totals against manual counts and RS1 data.

In case of discrepancies between those sources the controller had to do again the manual counting and find out where was the cause of the discrepancy; the latter was recorded in the application.

These checks were done for all the enumeration areas end fully ensured the data quality during the data entry processing.

The following tables presents the errors found and flagged (and later-on corrected) during the data capture process.

Error summary statistics

Date: 01/20/2012

Individual questionnaires R3

R3	No errors	R3	With errors
1076170		660329	
Total R3:	1736499		
61.97 %		38.03 %	

Total R3 with errors			
Total R3 Questionnaires with errors	660329		
R3 with less than 5 errors	592295		
R3 with errors from 6 to 44	68034		
Total Errors			
Total Errors in those questionnaires	1680231		
	955910	56.89 %	Missing
	215728	12.84 %	Un respected skip questions
	508593	30.27 %	Inconsistency between questions

Household questionnaires R2

R2	No errors	R2	With errors
127620		169463	
Total R2:	297083		
42.96 %		57.04 %	

Total Errors			
Total Errors in R2 questionnaires	401747		
	68355	17.01 %	Missing
	26594	6.62 %	Un respected skip questions
	306798	76.37 %	Inconsistency between questions

Dwelling questionnaires R1

R1	No errors	R1	With errors
362639		50753	
Total R1:	413392		
87.72 %		12.28 %	
Total Errors			
Total Errors in R1 questionnaires	96839		
	35358	36.51 %	Missing
	361	0.37 %	Un respected skip questions
	61120	63.12 %	Inconsistency between questions

Note: The above summary statistics are produced based on the error table information that was stored during the data entry verification phase.

2.2 Documentation of the census materials management system

The data entry processing operation was organized and managed in a way to fully ensure the quality of data capture and a complete compilation of paper questionnaires.

The steps below explain the main procedures taken to implement this operation.

- The data processing started from archiving. The archivist worked as first involved in the materials processing chain: the archivist first task was to register the EA box in the database and assign it a specific physical place in the Archive room.
- When receiving the boxes from the field, archivists count them manually and checked whether all of them were back in the office, based on GIS data.
- The archivists used a special application developed in VB.Net for registering the shelf ID number and the status of each box.
- Status of the box could be: Received, Manually counted, RS1 entered, content Coded, content captured (data-entry), Verified by Controller.
- No box could reach the next step unless the previous one was fully completed following the order above.
- The only person authorized to updating the status of the EA and assign it to the next phase was the archivist.
- Data Entry Controller checked the total figures of each EA and compared them with the figures given by Municipalities, manual counting figures, RS1 for dwellings, households and persons.
- A report comparing the figures from the boxes, manual counting and total figures from RS1 was generated, underlining the differences between those three figures.
- In case that the figures did not match between three sources the controller checked again the manual count trying to find out the reason of those differences.

- Precise instruction was given to the controller not to delete any data from the RS1 forms, just verifying that the manual count was done in the right way and that all the cases reported on the RS1 were physically in the box.
- The application for the controller ensured that all questionnaires in the box were entered, and checked the totals with manual count and RS1
- In case of discrepancies between those sources the controller had to perform again manual counting and find out where is the cause of the discrepancy; the latter was then recorded in the application.

CHAPTER 3

Localization of errors and correction of the inconsistencies

3.1 The principles of data editing and imputation

Data collected by statistical offices may contain errors occurred during the enumeration phase or during the data entry. These errors have to be corrected before reliable statistical information can be published. This is done with processes called editing and imputation, familiarly called "data cleaning", which are based on the most proven and recognized statistical methods[2]. Editing and imputation techniques are not only applied to censuses, they are also largely used to ensure consistency of data collected by sample surveys.

Thanks to a project of technical assistance financed by the European Union[3] ASK staff was trained to these methods and to the use of specialized software in the performance of data quality assessment and corrections. The exercise resulted above all in coherent and consistent final census data, which is essential for the objectives of data relevance, accuracy and coherence, three of the main criteria for statistical data quality required by the European Statistics Code of Practice.

3.2 General outline of data cleaning process

Errors in data can be detected by specifying several constraints (edits rules) that have to be satisfied by the values observed for each individual/respondent; for discrete variables, they are represented by a set of logical conditions, while for continuous variables they are represented by sets of linear equalities or inequalities.

Usually these edits derive from the structure of the questionnaire (for example when there are fields that represent the total of different answers) or are introduced by using some external information (the minimum/maximum of some variables; or a logical compatibility).

When a record fails an edit, it is considered erroneous; but this does not mean that all its values have to be considered not valid; the problem is to find the minimum number of fields to change in order to ensure that the record satisfies all the edits at the same time. This problem is strictly related to the need of keeping as much as possible the information given by the respondents. The

[2] For more details, see bibliographic references at the end of this report.

[3] "Technical Assistance to the Statistical Office of Kosovo for the preparation of a population census", an EU-funded project managed by the European Union Office in Kosovo; Implemented by Istat, CIRPS-Sapienza University of Rome, ICON-INSTITUT and Rrota

editing and imputation methodologies that are used in such context refer to Error Localization Problems (ELP).

The general workflow of data cleaning process is the same for all questionnaires and can be summarized in the following main phases:

DATA EXPORT

Detection and elimination of duplicate records as record with same values for key variables such as name, surname, date of birth, place of residence of the mother at respondent's birth, sex and marital status.
Export of Census data from the MySQL census raw data-base in ASCII tab delimited format.
Export and trim of special characters eventually presents on textual variables of data.

DATA MANAGEMENT

The first step was the analysis of raw data; it consisted in the creation of frequency tables for all categorical variables using ADaMSoft[4]. The distinctions between stochastic and systematic errors was done at this stage, by deterministic and probabilistic approach to error localisation; the IF-THEN rules, and the edit rules were then written in normal form; they were followed by introducing, with the apposite software, the main techniques for data imputation (by donor and by model).

The imputation algorithms used have been the following:

- **Deterministic imputation**: this procedure analyses each field previously identified as requiring imputation to determine if there is only one vector of possible values which would satisfy the original edits;
- **Donor imputation**: this procedure uses a nearest neighbour approach to find, for each record requiring imputation, a valid record that is most similar to it and that will allow the imputed recipient record to pass the edits;
- **Vertices imputation**; this procedure uses the solutions found during the error localization step to find the correct values to impute; its results, however, don't consider the relationship between the variables as they are observed in the data set.

The preliminary treatment of data consisted in the exclusion of "empty" records and the concomitant addition of new variables (numeric ID codes, flag of errors, variables from different form). The lists of empty records have been manually checked in order to verify if they are really to be rejected or they have to be reinserted. Other lists to be manually checked have been, for example, records concerning

[4] In the last two years CASPUR, an Italian inter-university research group developed the statistical software ADaMSoft - in a strict collaboration with Istat. The software contains a series of procedures aimed at identifying the minimum number of fields that should be corrected in statistical data to attribute them more suitable values.

units not to be enumerated: For instance, referring to R1 form “Other housing unit without inhabitants”; or referring to R3 form “Children who were born after 31/03/2011”.

In this step, the ad hoc programs were written in R Language.

SYSTEMATIC ERRORS EDIT & IMPUTATION

Automatic deterministic check programs to eliminate systematic errors were run. In this step, the *ad hoc* programs were written in R Language.

RANDOM ERRORS EDIT & IMPUTATION

Edit and imputation of random errors were done using CONCORD-SCIA[5] software for logical variables and ADaMSoft software for quantitative data.

Generation of statistical tables to compare cleaned data and raw data was performed in order to verify the obtained results looking for errors remaining in data (edit rules missing) and over cleaned variables (to high frequencies of imputation). If necessary, modification of edits and new run of the process was done consequently.

POST-CHECK DATA MANAGEMENT

Finally, a process of data cleaning of variables linked with imputed ones was done, e.g. if missing age values are imputed, corresponding values of date of birth have to be computed.

The last step has been to create frequency tables for all cleaned categorical variables.

3.3 Technical tools used for data editing and imputations

The analysis of the two questionnaires confirmed that most of the editing and imputation activities can be carried out by using the probabilistic approach and the SCIA module in the CONCORD system. Yet a significant part of them, in particular in the individual questionnaire, required other approaches and instruments. In particular, some *ad hoc* programs had to be developed by IT staff.

Moreover, the first analysis of the questionnaire on households showed that the whole part related to agricultural questions require still a probabilistic approach, but a different instrument than SCIA, as this module is not able to handle continuous variables and mathematical constraints. In this case, the use of ADaMSoft was recommended and adopted.

[5] Istat developed a particular generalised instrument (CONCORD) that contains also specialized modules for Census data validation (in particular the module SCIA). This application is an open-source.

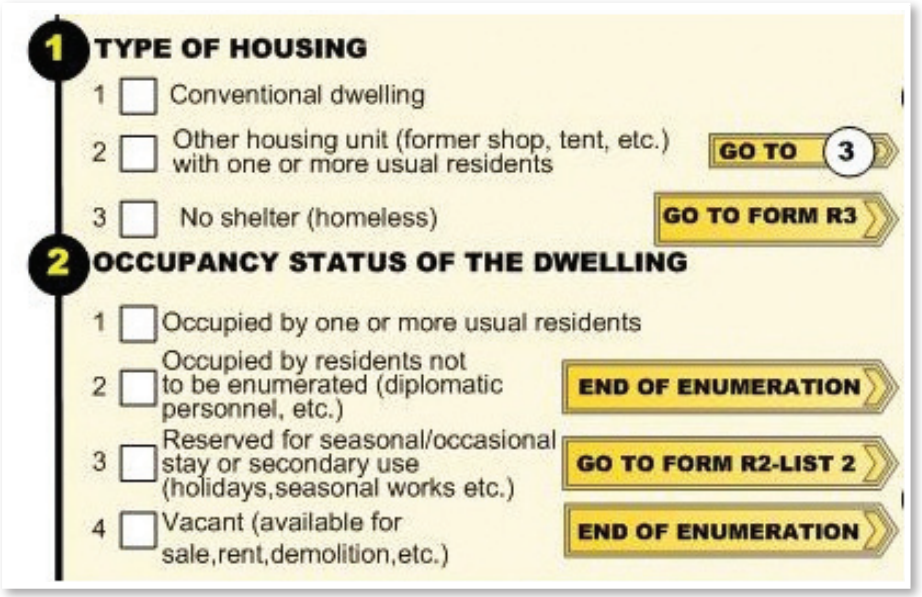
A very delicate question regards the issue of the connection between the phase of the data entry and that of the data editing and imputation. An effort has been done in order to align one the one hand the control rules adopted by the CsPro procedures in the data entry applications with the ones used in CONCORD and ADaMSoft applications; and on the other hand the format of output dataset of data entry applications, with the format required by editing and imputation software applications.

3.4 Editing and imputation rules used and their results

3.4.1 Data on dwellings and building

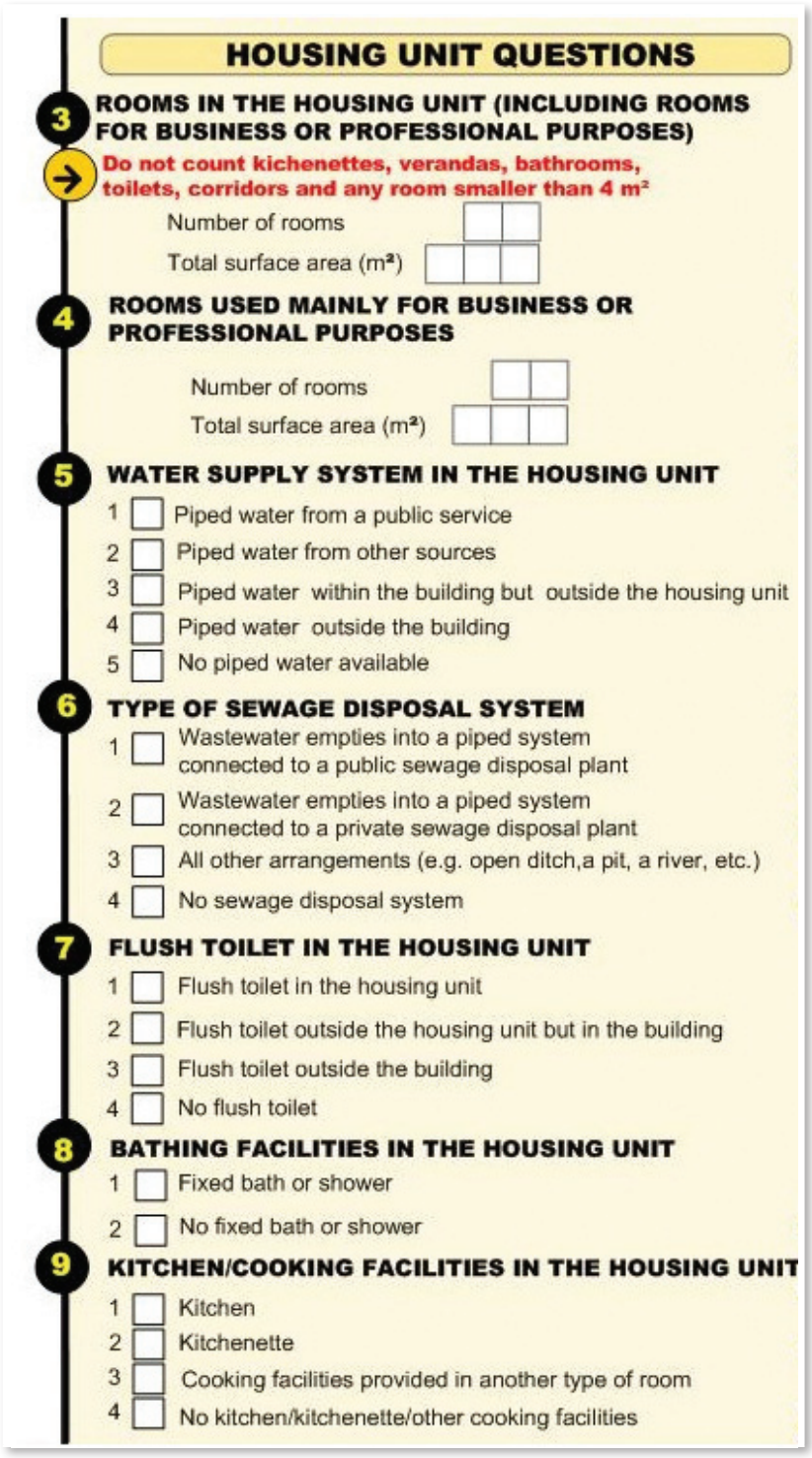
The Form R1 – population and housing census - dwelling questionnaire was the first one to be processed for E&I.

Dwelling questionnaire (Form R1) is the one used for recording all the housing units that can be: *a)* all conventional dwellings; and *b)* all other housing units used as the usual residence of one or more persons at the time of the census, as it is shown below.



The questionnaire is further composed by questions about housing unit (Q3-Q11), questions and about building (Q12-Q15) and, then two summary questions concerning the number of households and the number of persons resident in the housing unit (Q16 and Q17).

Questions on housing unit characteristics:



HOUSING UNIT QUESTIONS

3 ROOMS IN THE HOUSING UNIT (INCLUDING ROOMS FOR BUSINESS OR PROFESSIONAL PURPOSES)

→ **Do not count kichenettes, verandas, bathrooms, toilets, corridors and any room smaller than 4 m²**

Number of rooms ☐☐

Total surface area (m²) ☐☐☐

4 ROOMS USED MAINLY FOR BUSINESS OR PROFESSIONAL PURPOSES

Number of rooms ☐☐

Total surface area (m²) ☐☐☐

5 WATER SUPPLY SYSTEM IN THE HOUSING UNIT

1 ☐ Piped water from a public service
2 ☐ Piped water from other sources
3 ☐ Piped water within the building but outside the housing unit
4 ☐ Piped water outside the building
5 ☐ No piped water available

6 TYPE OF SEWAGE DISPOSAL SYSTEM

1 ☐ Wastewater empties into a piped system connected to a public sewage disposal plant
2 ☐ Wastewater empties into a piped system connected to a private sewage disposal plant
3 ☐ All other arrangements (e.g. open ditch,a pit, a river, etc.)
4 ☐ No sewage disposal system

7 FLUSH TOILET IN THE HOUSING UNIT

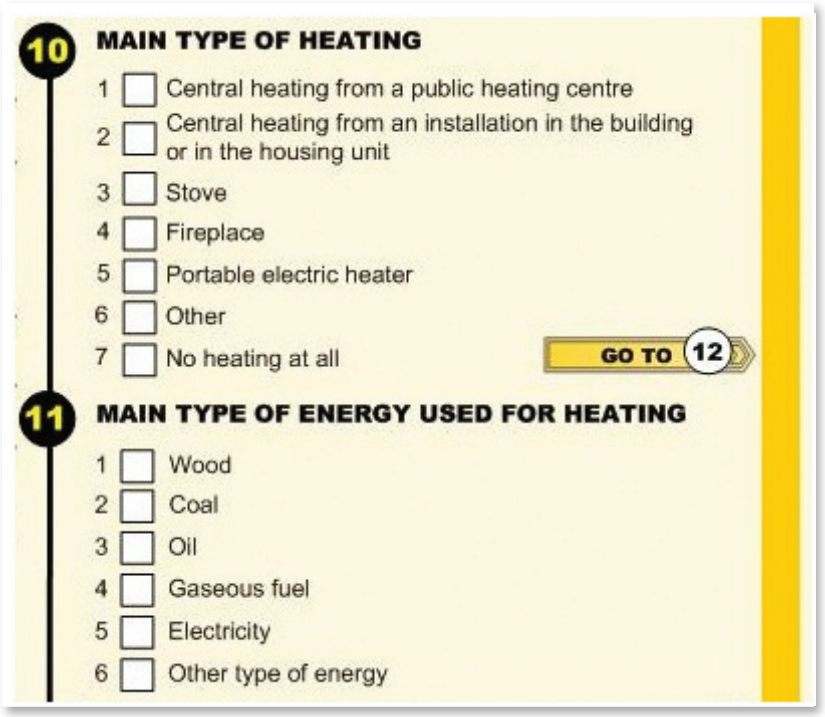
1 ☐ Flush toilet in the housing unit
2 ☐ Flush toilet outside the housing unit but in the building
3 ☐ Flush toilet outside the building
4 ☐ No flush toilet

8 BATHING FACILITIES IN THE HOUSING UNIT

1 ☐ Fixed bath or shower
2 ☐ No fixed bath or shower

9 KITCHEN/COOKING FACILITIES IN THE HOUSING UNIT

1 ☐ Kitchen
2 ☐ Kitchenette
3 ☐ Cooking facilities provided in another type of room
4 ☐ No kitchen/kitchenette/other cooking facilities



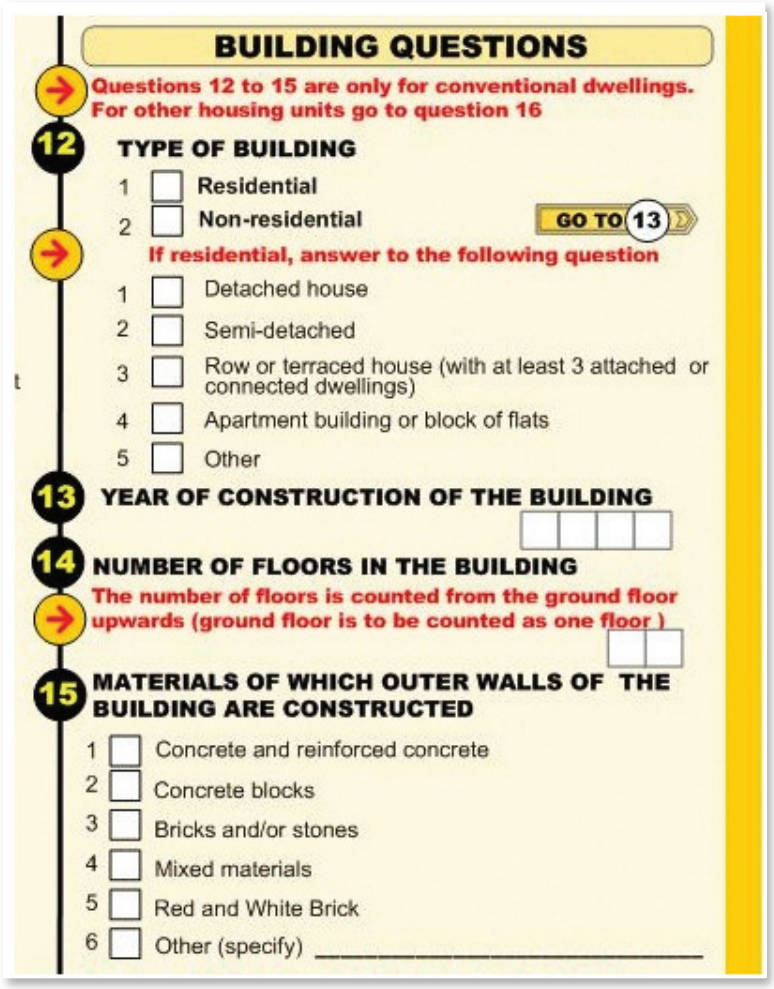
10 **MAIN TYPE OF HEATING**

1 ☐ Central heating from a public heating centre
2 ☐ Central heating from an installation in the building or in the housing unit
3 ☐ Stove
4 ☐ Fireplace
5 ☐ Portable electric heater
6 ☐ Other
7 ☐ No heating at all → GO TO 12

11 **MAIN TYPE OF ENERGY USED FOR HEATING**

1 ☐ Wood
2 ☐ Coal
3 ☐ Oil
4 ☐ Gaseous fuel
5 ☐ Electricity
6 ☐ Other type of energy

Questions on building characteristics:



BUILDING QUESTIONS

Questions 12 to 15 are only for conventional dwellings. For other housing units go to question 16

12 TYPE OF BUILDING

1 ☐ Residential
2 ☐ Non-residential — GO TO 13

If residential, answer to the following question

1 ☐ Detached house
2 ☐ Semi-detached
3 ☐ Row or terraced house (with at least 3 attached or connected dwellings)
4 ☐ Apartment building or block of flats
5 ☐ Other

13 YEAR OF CONSTRUCTION OF THE BUILDING ☐☐☐☐

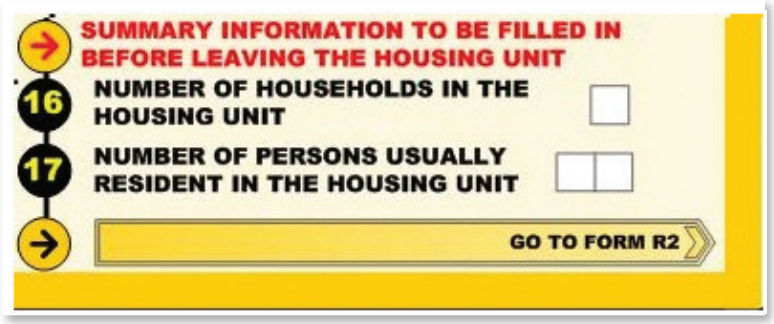
14 NUMBER OF FLOORS IN THE BUILDING

The number of floors is counted from the ground floor upwards (ground floor is to be counted as one floor) ☐☐

15 MATERIALS OF WHICH OUTER WALLS OF THE BUILDING ARE CONSTRUCTED

1 ☐ Concrete and reinforced concrete
2 ☐ Concrete blocks
3 ☐ Bricks and/or stones
4 ☐ Mixed materials
5 ☐ Red and White Brick
6 ☐ Other (specify) _____

Questions on households and persons occupying the housing unit:



SUMMARY INFORMATION TO BE FILLED IN BEFORE LEAVING THE HOUSING UNIT

16 NUMBER OF HOUSEHOLDS IN THE HOUSING UNIT ☐

17 NUMBER OF PERSONS USUALLY RESIDENT IN THE HOUSING UNIT ☐☐

GO TO FORM R2

A problem arose from the structure of the form because questions regarding the building are repeated for all occupied dwellings. For that reason, variables related to building questions were treated separately. It was also found interesting to compare information of Q12 (type of building) and Q14 (number of floors) questions with data recorded in the building list.

Records without information related to questions from Q1 to Q17 - that had only ID codes recorded – have been deleted. This problem was stemming from ID codes wrongly inserted in the data-entry phase.

Preliminary treatment

The goal of this phase was to prepare all variables necessary for data cleaning. The structure of output record was described in a specific EXCEL file.

Steps performed by the software application:

Reading R1 data on dwellings;
Elimination of duplicated records;
Preparation of a unique code for R1 data;
Sort data by the ID codes (Each form had a unique ID code, composed of a series of codes relating to: Municipality, settlement, EA, building, entrance, and dwelling);
Add new variables as error flags, unique code, other flags; Read R3 individual data (they serve to control R1 data);
Writing a flag of presence of dwelling codes within data of other questionnaires;
Writing a list of empty records (probably caused by errors during data-entry phase);
Writing a list of other housing units not to be enumerated;
Writing R1 files that had only valid records.

Systematic errors processing

A systematic error may be defined as a non probabilistic error. It can be caused by some defects in the surveying structure (e.g. question wrongly specified, not respected routing rules,). Deterministic imputation is considered more suitable for correction of systematic errors and consists in assigning to the erroneous variables a value considered ‘true’ by experts.

In the following there is the list of deterministic imputations performed for dwelling questionnaire.

ERROR CODE	CONDITION	ACTION
E100	*Empty record* – Questions 1 to 17 missing	Delete record
E101	*Q1 - Type of housing – Vacant dwelling:* Type of housing is missing or different from 1 (conventional dwelling) and the occupancy status is 4 (vacant dwelling) and all others questions of the form are missing and the ID codes of dwelling are not present in any other questionnaire	Q1 - Type of housing equal to conventional (Q1 = 1)
E102	*Inhabited dwelling without corresponding R2 or R3 forms:* If the dwelling is conventional (Q1=1) and is inhabited (Q2=1) and the number of HH and of usual residents is at least 1 (Q16>0; Q17>0) but there is no data in other forms (N_presence=0)	an error is signalled but no imputation is done
E103	*Vacant Dwelling* - Q1 equal to 1 (conventional dwelling) and Q2 different from 1 (missing or not conventional dwelling) and Q16 = 0 and Q17 =0 (no usual residents) and N_presence = 0 (no data in other forms)	Vacant dwelling: Q2=4 (Occupancy = vacant) and Q3-Q17 codes are put NULL (skipped)
E104	*Dwelling for temporary/seasonal stay (Routing rule)*: Q1 equal to 1 and Q2 equal to 3 (Dwelling reserved for seasonal stay) and no resident in R2 and R3 dataset but some other questions of R1 are answered is a not respected routing rule	The code of Q3-Q15 is put null (NA)
E105	*Number of rooms in total and for business*: Q3 (total number of rooms) is less than Q4 (the number of rooms for business or professional purposes)	Q3 and Q4 are both swapped
E106	*Total surface for business and for housing*: If Q3B (total surface of Housing Unit) is less than Q4B(the surface for business or professional purposes) then Q3B and Q4B are both swapped	Q3B and Q4B are both swapped
E107	*Homeless Skip*: If the answer to Q1 is "no shelter (homeless) and the answers Q2 to Q15 are indicated missing (code -1)	The code of Q2-Q15 is put null (NA)
E108	*Standardization of Q16*: Q16 = -1 or 0 and Q1 = 1 and Q2=3 (dwelling reserved for occasional stay) and no data in R2 list1 and R3 individual	Q16 = 1
E109	*Standardization of Q16*: Q16 = -1 and Q1 = 1 and Q2 !=3 and no data in R2 list1 and R3 individual	Q16 = NA AND Q2 = 4
E110	*Preparation of Question 4 for Concord*: change of NULL and -1 with 0 for Total surface area (Q4B) and number of rooms (Q4A) used for business and professional purposes	Q4A= 0 Q4B=0
E111	*Q1 – Homeless:* According to ASK interactive checks, There are no real homeless in Kosovo	Homeless code is changed in conventional dwelling code

Random errors processing

1. Qualitative variables on housing units – Questions 1-11

Random errors on variables (Q1 - Q11) relating to housing units were corrected using the CONCORD - SCIA software.

The defined edit rules are listed in the following table.

N.	Error Condition	SCIA EDIT
	ROUTING RULES	
	Q1 – Type of housing: if conventional dwelling (1) then Q2 has to be answered; if other housing unit (2) go to Q3;	
1	Q1 (Type of housing) is equal 3 (homeless) and some other question is answered; COMMENT	*DW01(3) SKIP1<)
2	Q1 (Type of housing) is equal 1 (conventional dwelling) and Q2 (occupancy status) is not answered;	DW01(1) DW02( )
3	Q1 (Type of housing) is equal 2 (other housing unit) and Q2 (occupancy status) is answered;	DW01(2) DW02<)
	Q2 - routing rule for dwelling NOT occupied by usual resident: skip housing unit questions SKIP2: DW03A DW03B DW04A DW04B DW05 DW06 DW07 DW08 DW09 DW10 DW11	
4	Q2 (occupancy status) is different from 1 and questions concerning housing unit and building (Q3 to Q15) are answered;	DW02(2-4) SKIP2<)
5	Q2 (occupancy status) equals 2 (residents not to be enumerated) or 4 (vacant) and Q16 (Number of household in the housing unit) is different from 0 or missing;	DW02(2,4) DW16< 0,)
6	Q2 (occupancy status) equals 2 (residents not to be enumerated) or 4 (vacant) and Q17 (Number of persons resident in the housing unit) is different from 0 or missing;	DW02(2,4) DW17< 0,)
7	Q2 (occupancy status) equals 3 (temporary/seasonal use) and Q17 (Number of persons resident in the housing unit) is different from 0 or missing.	DW02(3) DW17<0,)
	Q10 = no heating at all -> Q11(type of energy for heating) = blank	
8	Q10 (main type of heating) = 7 (no heating at all) and Q11 (main type of energy used for heating) different from missing	DW10(7) DW11<)
9	Q10 (main type of heating) = 1-6 (some heating) and Q11 (main type of energy used for heating) equals missing	DW10(1-6) DW11( )

	Q2 = other housing unit skip building questions(go to q16) Q12_15 = 1 means at least one of questions DW12A DW12B DW13 DW14 DW15A is answered	
10	Q1 (type of housing) = 2 (other housing unit) and building questions different from missing;	DW01(2,3) Q12_15(1)
	Building Questions (12-15) are mandatory for conventional dwellings	
	If conventional dwelling : Q1 (Type of housing) = 1 and Q2 (Occupancy status) = 1 building questions (Q12-Q15) are mandatory	DW01(1) DW02(1) Q12_15(0)
	Q2=2 questions Q3 TO Q10 AND Q16 AND Q17 has to be answered SKIP4: DW03A DW03B DW05 DW06 DW07 DW08 DW09 DW10 DW16	
12	Q1 (Type of housing) = 2 (other housing unit) and Q3 to Q11 (housing unit questions) and Q16 and Q17 (summary questions) equals missing;	DW01(2) SKIP4( )
	Housing unit questions are mandatory for conventional building SKIP5: DW03A DW03B DW04A DW05 DW06 DW07 DW08 DW09 DW10 DW12A DW13 DW14 DW15A DW16 DW17	
13	Q1 (Type of housing) is equal 1 (conventional dwelling) and Q2 (occupancy status) is equal 1 (usual residents) and Q3 to Q11 (housing unit questions), Q12A to Q15 (building questions) and Q16 and Q17 (summary questions) equals missing;	DW01(1) DW02(1) SKIP5( )
	SUBSTANTIAL RULES	
	EDITS ABOUT WATER SUPPLY SYSTEM, FLUSH TOILET AND SEWAGE DISPOSAL SYSTEM	
14	Q7 = 2 (flush toilet outside the housing unit but in the building) and Q5 (water supply system) is different from 1, 2 or 3 (piped water inside the building);	DW07(2) DW05<1-3)
15	Q6 (type of sewage) = 1 or 2 (wastewater into a piped system) and Q5 (water supply system) is equal 5 (no piped water available);	DW06(1,2) DW05(5,)
16	Q7 (flush toilet) = 1 (in the housing unit) and Q5 (water supply system) is equal 5 (no piped water available);	DW05(5,) DW07(1)
	EDITS ABOUT WATER SUPPLY SYSTEM AND BATHING FACILITIES	
18	Q8 (bathing facilities) = 1 (fixed bath or shower) and Q5 (water supply system) equal 5 (no piped water available);	
	EDITS ABOUT TYPE OF HEATING AND ENERGY USED FOR HEATING	
	CHANGED RULE - If in the Q10 is marked modality (3) then it follows that Q11 should have one of the modalities 1 or 2 or 3 or 4	
20	Q10 (main type of heating) = 3 (stove) and Q11 (main type of energy used for heating) can't be equal to 5 (Electricity) or 6 (Other type of energy)	DW10(3) DW11<1,2,3,4)
21	Q10 (main type of heating) = 4 (fireplace) and Q11 (main type of energy used for heating) different from 1 (wood);	DW10(4) DW11<1)
22	Q10 (main type of heating) = 5 (portable electric heater) and Q11 (main type of energy used for heating) different from 5 (electricity);	DW10(5) DW11<5)

2. Final Data management for dwelling and preparation of building data

Steps performed by the software:

Reading of CLEANED R1 data on dwellings;
Cleaning of building information according routing rules for type of housing;
Writing of dwelling final data (questions 1-11, 16, 17);
Preparation of building file. When building questions Q12-15 had different values for the dwellings it contains the choice was performed with the following criteria: If unique, the value with maximum frequency of answers; if the same maximum frequency is attained at different values, random choice between the modal values; Question 13 (Year of construction): the minimum value.
Writing of building file.

3. Qualitative variables on buildings – Questions 12-15

Finally, random errors E&I phase for buildings. Information was requested only for occupied conventional dwellings.

An inconsistency was defined when:

N.	Error Condition	SCIA EDIT
	Q12 = NON RESIDENTIAL GO TO Q13	
1	Q12A (type of building) = 2 (non residential building) and Q12B different from missing;	DW12A(2) DW12B <)
2	Q12A (type of building) = 1(residential building) and Q12B equal to missing;	DW12A(1) DW12B( )
	IF DETACHED OR SEMI-DETACHED HOUSE AND NUMBER OF FLOOR >= 7 THEN ERROR	
3	Q12B (type of building) = 1 or 2 (detached or semi-detached house) and Q14 (number of floor) >= 7;	DW12B(1,2) DW14(7-99)

4. Post-Check Data Management

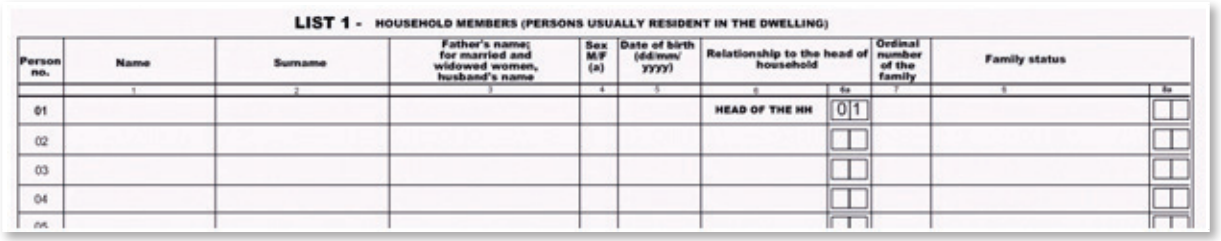
Final tables were produced and a merge of dwelling and building data was achieved. Annex 1 presents the results of editing and imputation of R1 questionnaires, by confronting the initial (or raw) data with the ones obtained by the cleaning process.

3.4.2 Data on households and families

The household questionnaire R2 comprises information of household compositions as well as data on household living characteristics, among which detailed data on land ownership and on its use for agricultural activities.

Household composition

Data on household composition were captures with the List 1 illustrated below.



LIST 1 - HOUSEHOLD MEMBERS (PERSONS USUALLY RESIDENT IN THE DWELLING)

Person no.	Name	Surname	Father's name; for married and widowed women, husband's name	Sex M/F (a)	Date of birth (dd/mm/ yyyy)	Relationship to the head of household		Ordinal number of the family	Family status	
	1	2	3	4	5	6	6a	7	8	8a
01						HEAD OF THE HH	01			
02										
03										
04										

Most of the individual data of this list could be compared with the same ones collected in individual questionnaires R3. This was the case for age, sex and names.
In this List 1, column 6 provides the household composition, in relationship with one unique reference person per household called “head of household”. Columns 7 and 8 provide the family ordinal number to which each household member belongs, as well as its family status inside this family nucleus.

Relationship to head of household and family status could be:

Household members: persons who belong to this private household (see definition below) and have lived in the dwelling for a continuous period of at least 12 months or have moved into the dwelling during the twelve months before Census Day with the intention of staying for at least one year.
N.B. Household members usually resident in this dwelling but temporarily absent for a period of less than 1 year on Census Day should be included.

Relationship to the head of household

01 Head of the household
02 Husband-wife with marriage certificate
03 Husband-wife without marriage certificate
04 Second wife
05 Son / Daughter
06 Son-in-law / Daughter-in-law
07 Father / Mother
08 Father-in-law / Mother-in-law
09 Brother / Sister
10 Grandson / Granddaughter
11 Other relative (grandfather, grandmother, uncle, aunt, etc.)
12 Not a relative

Family status

01 Husband/wife with marriage certificate
02 Husband/wife without marriage certificate
03 Second wife
04 Lone parent without a partner in the family
05 Lone parent with a partner living outside Kosovo
06 Child of both parents
07 Child of male partner
08 Child of female partner
09 Child of lone parent
10 Person belonging to no family nucleus

Within the sets of variables and data the census collected, household arrangements and family compositions have been the most complex and the less well understood by the enumerators. Indeed, the statistical concept of a family – also called family nucleus – is much more restricted that the concept of family people refers to in their everyday life. For example, if a married couple and their young children are living together with the older parents of one of the spouses, they all are considered as one family in popular and societal point of view; on the contrary, they are counted as two distinct families in statistical terms: the first family is composed by the older parents of one spouse; and the second family is composed of the couple with their young children. Indeed, a family is defined as "Two or more persons living together linked with family ties of first grade, such as a couple with or without children or a lone-parent with at least one child; in the census definition, a grand-mother with a grand-child, or two brothers without parents are not considered a family. Similarly, grand parents living with their married son, daughter in law and grand-children are counted as two families."

Identifying who were the families within Kosovo households has been quite difficult for the field staff, not only because of the difficulty to understand the definitions but additionally because of the complexity of family living arrangements in Kosovo. One of the systematic errors found was that the persons were not given a family ordinal number in the column 7 of the List 1 but a serial ordinal personal number as member of the household. By consequence, the identification of families within households had to be done ex-post, and family reconstitution procedures have been very complex.

On the other hand, relationship to the head of household also showed many reporting errors. In a quantity of cases the relationship has been reported in reverse direction. For example, if the head of household was a grandfather, instead for the grandchild to report as "grand-son" of head of household, cases were found where the answer was "grandfather", because the respondent answered who the reference person was to him (he is my grand-father) instead of responding who he was for the reference person (I am his grand-son). These cases were identified in relation with age, name and family status, combined with the marital status reported in individual questionnaire. Another frequent error was due to the fact that in Albanian language there is a unique word to define a grand-son/daughter and a nephew/niece. As a consequence, editing and imputation procedures for these cases have been extremely complex and required lot of iterative processes, which also caused delays in the initial operational plan for data cleaning.

Among all cases, roughly half million individuals had some reporting (non-logical) error in family status or in relationship to the head of household: slightly less than 30% of all population. Corrections of inconsistencies have been based on a set of rules indicating what the potentially proper combinations of households (relationship with head of HH) with the ones of families (based on family status and family ordinal number) were. Based on this information, which constituted the basic editing rules, specific programmes have been developed in ADaMSoft to identify the closest

model to each erroneous case in order to apply a proper donor-based imputation. Several constraints on age; sex; and marital status have been added to the procedures, in order to ensure internal data consistency.

For what concerns the changes made by the procedures, there were situations in which it was not possible to choose a valid marital status or a valid relationship with the head of household. This happened for 6,000 individuals (1,000 families). These cases underwent a case-by-case computer assisted editing.

When age has been considered as a valid data, in some cases it had been found not compatible with, for example, the marital status (example a too young person declared married as family status, but the "partner" could not be found in the household). In these situations the adopted solution has been to change the marital status to "never married" and, without the possibility to identify the correct husband between the different families of the household, to impute this person in a non-family member whose relationship to the head of household was 11 (i.e. other relative). These cases concerned 1,000 persons roughly.

In all other cases (something like more than 500,000) the procedure gave good results. For more than 300,000 family nuclei the procedures didn't change any variable but only attributed the ordinal family number.

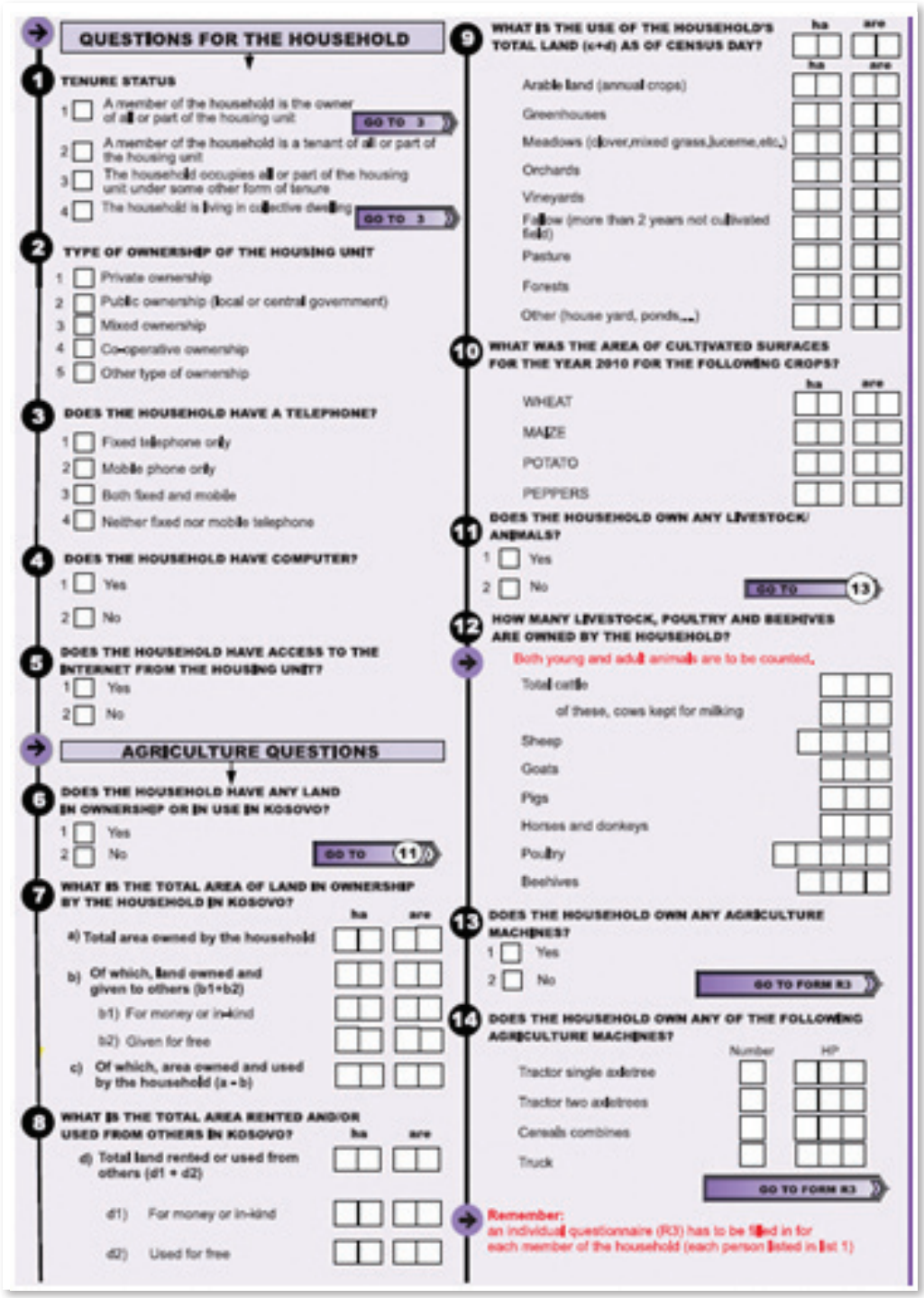
Household characteristics

Regarding household characteristics all the analyses were made on the information collected in the part of the questionnaire reported hereafter.

This information can be divided in five groups; in particular:

1. **General questions:** *Tenure status, type of ownership of the housing unit, does the household have a telephone, does the household have a computer, does the household have access to the internet from the housing unit*;
2. **Agriculture questions:** all the questions from **(6)** to **(9)**;
3. **Cultivated surfaces in 2010:** question **(10)**;
4. **Animals owned:** questions **(11)** and **(12)**;
5. **Agriculture machines:** questions **(13)** and **(14)**.

For each question in these groups some rules were introduced in order to identify errors in questionnaires (i.e. incoherencies in the values). All these rules are described in the next paragraph.



QUESTIONS FOR THE HOUSEHOLD

1 TENURE STATUS
1 A member of the household is the owner of all or part of the housing unit — GO TO 3
2 A member of the household is a tenant of all or part of the housing unit
3 The household occupies all or part of the housing unit under some other form of tenure
4 The household is living in collective dwelling — GO TO 3

2 TYPE OF OWNERSHIP OF THE HOUSING UNIT
1 Private ownership
2 Public ownership (local or central government)
3 Mixed ownership
4 Co-operative ownership
5 Other type of ownership

3 DOES THE HOUSEHOLD HAVE A TELEPHONE?
1 Fixed telephone only
2 Mobile phone only
3 Both fixed and mobile
4 Neither fixed nor mobile telephone

4 DOES THE HOUSEHOLD HAVE COMPUTER?
1 Yes
2 No

5 DOES THE HOUSEHOLD HAVE ACCESS TO THE INTERNET FROM THE HOUSING UNIT?
1 Yes
2 No

AGRICULTURE QUESTIONS

6 DOES THE HOUSEHOLD HAVE ANY LAND IN OWNERSHIP OR IN USE IN KOSOVO?
1 Yes
2 No — GO TO 11

7 WHAT IS THE TOTAL AREA OF LAND IN OWNERSHIP BY THE HOUSEHOLD IN KOSOVO?

	ha	are
a) Total area owned by the household		
b) Of which, land owned and given to others (b1+b2)		
b1) For money or in-kind		
b2) Given for free		
c) Of which, area owned and used by the household (a - b)		

8 WHAT IS THE TOTAL AREA RENTED AND/OR USED FROM OTHERS IN KOSOVO?

	ha	are
d) Total land rented or used from others (d1 + d2)		
d1) For money or in-kind		
d2) Used for free		

9 WHAT IS THE USE OF THE HOUSEHOLD'S TOTAL LAND (c+d) AS OF CENSUS DAY?

	ha	are
Arable land (annual crops)		
Greenhouses		
Meadows (clover,mixed grass,lucerne,etc.)		
Orchards		
Vineyards		
Fallow (more than 2 years not cultivated field)		
Pasture		
Forests		
Other (house yard, ponds,...)		

10 WHAT WAS THE AREA OF CULTIVATED SURFACES FOR THE YEAR 2010 FOR THE FOLLOWING CROPS?

	ha	are
WHEAT		
MAIZE		
POTATO		
PEPPERS		

11 DOES THE HOUSEHOLD OWN ANY LIVESTOCK/ ANIMALS?
1 Yes
2 No — GO TO 13

12 HOW MANY LIVESTOCK, POULTRY AND BEEHIVES ARE OWNED BY THE HOUSEHOLD?
Both young and adult animals are to be counted.

Total cattle	
of these, cows kept for milking	
Sheep	
Goats	
Pigs	
Horses and donkeys	
Poultry	
Beehives	

13 DOES THE HOUSEHOLD OWN ANY AGRICULTURE MACHINES?
1 Yes
2 No — GO TO FORM R3

14 DOES THE HOUSEHOLD OWN ANY OF THE FOLLOWING AGRICULTURE MACHINES?

	Number	HP
Tractor single axletree		
Tractor two axletrees		
Cereals combines		
Truck		

GO TO FORM R3

Remember:
an individual questionnaire (R3) has to be filled in for each member of the household (each person listed in list 1)

1. Rules defined in order to validate each household questionnaire

The rules that were defined take into account the type of the variables to which the questions refer. All the variables in the group (1) are considered qualitative; all those in group (2), (3), (4) and (5) as quantitative.

For each group different controls were introduced; in particular:

General questions:

Tenure status: this variable cannot contain missing values;
Type of ownership of the housing unit: this variable can contain missing values only when the *Tenure status* has a value of 1 or 4;
Tenure status and *Type of ownership of the housing unit*: when the *Tenure status* is equal to 1 or 4 the value of the *Type of ownership of…* must be blank;
Does the household have a telephone: this variable cannot contain missing values;
Does the household have computer: this variable cannot contain missing values;
Does the household have access to the internet from the housing unit: this variable cannot contain missing values;

Agriculture questions:

Does the household have any land in ownership or in use in Kosovo?: this variable should assume a value of *1* when there is a numerical values in one of the fields related to the questions from (7) to (9); on the contrary the variable must have a value of *2*;
Questions from (7) to (9): in this case 18 new variables were defined in order to transform the original two values (HA and Are) of each answer in a single field (obtained by considering Ha*100+Are). These new variables were: *D7a* (Total area owned by the household), *D7b* (area owned and given to others), *D7b1* (area owned, given to others for money or in-kind), *D7b2* (area owned, given to others for free), *D7c* (area owned and used by the household), *D8d* (total land rented or used from others), *D8d1* (total land rented or used from others for money or in-kind), *D8d2* (total land rented or used from others for free), *D9* (total land really used by the household), *D9a* (land used, arable), *D9b* (land used, greenhouses), *D9c* (land used, meadows), *D9d* (land used, orchards), *D9e* (land used, vineyards), *D9f* (land used, fallow), *D9g* (land used, pasture), *D9h* (land used, forest), *D9i* (land used for other purposes). According to what was specified in the questionnaire, the following checking rules were adopted: *d7a>=0; d7b<=d7a; d7b=d7b1+d7b2; d7c=d7a-d7b; d8d=d8d1+d8d2; d8d>=d8d1; d8d>=d8d2; d9=d7c+d8d; d9=d9a+d9b+d9c+d9d+d9e+d9f+d9g+d9h+d9i*. A record that did not satisfy all these rules at the same time was considered erroneous.

Cultivated surfaces in 2010: *in this case no control was introduced because the time to which the variable refers is different to the one that refers to all the agriculture questions.*

Animals owned: a control was introduced that was used in order to correct the systematic error due to a misleading of the question *Total cattle*, intended often as the *Total of the animals owned*. In particular it should be observed that the field *Total cattle* is followed by the question *Of these, cows kept for milking*. In this case the following rule was considered to check each individual questionnaire: If (*Total cattle*>0) AND IF (Total cattle=(of these cows kept for milking)+sheep +goats+pigs+(horses and donkeys)+poultry+beehives) AND IF (sheep+goats+pigs+(horses and donkeys)+poultry+beehives>0) THEN *Total cattle=of these cows kept for milking*. Furthermore, for all the questions related to the ownership of the animals, it was necessary to check the correctness of the answers given to question (11); it must have a value of *1* if the household owns at least one animal, otherwise it must have the value of *2* (no animals owned);

Agriculture machines: in these questions the general idea was that a valid value in the *Number* must correspond to a value in the horse-power *HP* (for each of the different machines that were considered). Obviously the contrary is true (a value found in the *HP* field must correspond to a value in the *Number* of agricultural machines). For what concerns the answers to the questions 13, it should be *1* in case the household owns one of the machines described, *2* if he doesn't own any machine.

2. Data processed

The census data that refer to the questionnaires used are *315,513*. Of these, *18,535* were not considered according to what derived from the controls made (cases where the household had no permanent resident). The data that were finally considered refer to *297,078* households.

3. General considerations on the errors

According to the rules defined above, it was possible to define for each record a score of correctness; it could assume a value between 0 (no errors) to 10 (errors for all the groups that were considered).

Were considered in particular:
Error in *Tenure Status;*
Error in *Type of ownership of the housing unit;*
Error in *Does the household have a telephone*;
Error in *Does the household have a computer;*
Error in *Does the household have access to the internet from the housing unit*;
Error in *Does the household have any land in ownership or in use in Kosovo*;
Error in *Does the household own any livestock/animals*;
Error in *Does the household own any agriculture machines*:
Error in interpreting the *Total cattle*;

General error in *land surfaces or in agriculture machine*.
The table below shows the distribution of the total amount of such errors for the census data (frequencies of records according to their number of errors).

Errors	Count	Percentage
0	188719	63.525
1	90997	30.631
2	13510	4.548
3	1341	0.451
4	189	0.064
5	2215	0.746
6	79	0.027
7	23	0.008
8	5	0.002
TOTAL	***297078***	***100***

This table shows that the majority of the records were corrected. The most frequent situation was records having only one error (according to the rules that were defined). In the next paragraph the single sources of errors are analyzed.

4. Errors in general questions

The general questions were defined by considering the following fields: *Tenure status, Type of ownership of the housing unit, Does the household have a telephone, Does the household have computer, Does the household have access to the internet from the housing unit*.

The table below shows the crossed frequencies between *Tenure Status* and *Type of ownership of the housing unit*.

		Type of ownership of the housing unit							
		(-1) Missing	(1) Private ownership	(2) Public ownership (local or central government)	(3) Mixed ownership	(4) Co-operative ownership	(5) Other type of ownership	NOT PRESENT	***Total***
Tenure Status	(-1) Missing	2548	104	10	4	5	3	0	***2674***
	(1) A member of the household is the owner of all or part of the housing unit	26	26182	243	132	241	47	253864	***280735***
	(2) A member of the household is a tenant of all or part of the housing unit	1000	5953	613	305	95	296	0	***8262***
	(3) The household occupies all or part of the housing unit under some other form of tenure	303	2378	509	384	519	298	0	***4391***
	(4) The household is living in collective dwelling	1	48	40	9	13	2	903	***1016***
	Total	**3878**	**34665**	**1415**	**834**	**873**	**646**	**254767**	***297078***

Values in this table show the following error rates for these two questions:
0.9% for the *Tenure Status* (it cannot be missing);
9.51% for the *Type of ownership of the housing unit*; in particular:

- 9.07% because it cannot have a value if the *Tenure status* is *1* or *4*;
- 0.44% because it cannot be missing when the *Tenure status* is *2* or *3*.

Next table indicates values found in the questions on the telephone, computer and internet (absolute values and percentages on total records).

Question	Values	Frequencies	Percentage (on total cases)
Does The Household Have A Telephone	(-1) Missing	2877	0.97
	(1) Fixed telephone only	10186	3.43
	(2) Mobile phone only	227807	76.68
	(3) Both fixed and mobile	51869	17.46
	(4) Neither fixed nor mobile telephone	4339	1.46
Does The Household Have A Computer	(-1) Missing	3691	1.24
	(1) Yes	183427	61.74
	(2) No	109960	37.01
Does The Household Have Access To The Internet from the housing unit	(-1) Missing	4537	1.53
	(1) Yes	166805	56.15
	(2) No	125736	42.32

The following error rates were found in these questions:
Does the household have a telephone: 0.97%;
Does the household have a computer: 1.46%;
Does the household have access to the internet from the housing unit: 1.53%.

5. Errors in agriculture questions

The errors in agriculture section can be deducted from the general question Q6: *Does the household have any land in ownership or in use in Kosovo* by checking its coherence with data on the values given in questions 7 to 9.

For what concerns the latter, they are analyzed in the paragraph related to the numerical questions, but for the general question it is useful to consider the frequencies in the following table where rows present the true values, and columns the values derived just by verifying if there is a numerical value in any of the questions from (7) to (9).

		Does Your Household Have Any Land In Ownership or in use in Kosovo (derived from check)		
		(1) Yes	**(2) No**	***Total***
Does Your Household Have Any Land In Ownership or in use in Kosovo	(-1) Missing	3281	954	***4235***
	(1) Yes	195099	167	***195266***
	(2) No	577	97000	***97577***
	Total	***198957***	***98121***	***297078***

For that question, the error rate is 1.7% (100*(577+167+4235)/297078).

6. Errors in animals owned

For what concerns the errors in the animal owned, according to the rule previously defined, there are 1923 records for which it is possible to consider a misunderstanding of the meaning of *Total cattle*; the error rate, in this case, is 0.6%.

The correctness of the answers given to the question (11) is derived from the following data on crossed frequencies for the original answers to the question *Does the household own any livestock animals* and the same derived by verifying if there is a numerical value in any of the fields related to the questions (12).

		Does Your Household Have Any Land In Ownership or in use in Kosovo (derived from check)		
		(1) Yes	**(2) No**	***Total***
Does Your Household Have Any Land In Ownership or in use in Kosovo	(-1) Missing	3281	954	***4235***
	(1) Yes	195099	167	***195266***
	(2) No	577	97000	***97577***
	Total	***198957***	***98121***	***297078***

In this case the error rate is: 1.07%.

7. Errors in agriculture machines

Also In this case, as for the land owned, errors can be detected by comparing answers given to the question (13) with the values in the question (14), which show an error rate of 0.9%.

		Does your household own any livestock animals (derived from check)		
		(1) Yes	**(2) No**	***Total***
Does your household own any livestock animals	(-1) Missing	1009	1704	***2713***
	(1) Yes	88719	338	***89057***
	(2) No	155	205153	***205308***
	Total	***89883***	***207195***	***297078***

8. Errors in numerical variables

For what concerns the numerical variables (i.e. all those related to the dimension of the land owned and to the number of agriculture machines), all the coherency rules introduced have been the following:

Rule	Explanation
D7B-D7B1-D7B2=0	Of which land owned and given to others=for money or in-kind+given for free
-D7A+D7B+D7C=0	Total area owned by the household=of which area owned and used by the household+of which area owned and given to others
D8D-D8D1-D8D2=0	Total area rented or used from others=for money or in-kind+used for free
-D7C-D8D+D9=0	What is the use of the household's total land as of census day=Of which area owned and used by the household+total area rented or used from others in Kosovo

D9-D9A-D9B-D9C-D9D-D9E-D9F-D9G-D9H-D9I=0	What is the use of the household's total land as of census day=Arable land+greenhouses+Meadows+Orchards+Vineyards+Fallow+Pasture+Forests+Other
D7A-D7B>=0	What is the total area of land in ownership by the household in Kosovo>=what is the total area rented and/or used from others in Kosovo
D8D-D8D1>=0	Total area rented and/or used from others in Kosovo>=for money or in-kind
D8D-8D2>=0	Total area rented and/or used from others in Kosovo>=Used for free
-D14A1+D14A2>=0	HP for tractor single axletree>=Number of tractor single axletree
D14A1-1.0E-6*D14A2>=0	Number of tractor single axletree>=0.000001* HP for tractor single axletree
-D14B1+D14B2>=0	HP for tractor two axletrees>=Number of tractor two axletrees
D14B1-1.0E-6*D14B2>=0	Number of tractor two axletrees>=0.000001* HP for tractor two axletrees
-D14C1+D14C2>=0	HP for cereals combines >=Number of cereals combines
D14C1-1.0E-6*D14C2>=0	Number of cereals combines>=0.000001* HP for cereals combines
-D14D1+D14D2>=0	HP for truck>=Number of truck
D14D1-1.0E-6*D14D2>=0	Number of truck>=0.000001* HP for truck
D14A1>=0	All the variables must have a value greater or equal to 0
D14A2>=0	
D14B1>=0	
D14B2>=0	
D14C1>=0	
D14C2>=0	
D14D1>=0	
D14D2>=0	
D7A>=0	
D7B>=0	
D7B1>=0	
D7B2>=0	
D7C>=0	
D8D>=0	
D8D1>=0	
D8D2>=0	
D9>=0	
D9A>=0	
D9B>=0	
D9C>=0	
D9D>=0	
D9E>=0	
D9F>=0	
D9G>=0	
D9H>=0	
D9I>=0	

The application of these rules to census data gave the results summarized in the table hereafter, which provides data on number of satisfied or unsatisfied edits.

Edit indicator	Times the edit is correctly satisfied (absolute values)	Times the edit is correctly satisfied (percentage on total records)	Times the edit is not satisfied without others (absolute values)	Times the edit is not satisfied without others (percentage on total records)	Times the edit is not satisfied in association with others edit (absolute values)	Times the edit is not satisfied in association with others edit (percentage on total records)
D7B-D7B1-D7B2=0	294238	99.04	264	0.09	2576	0.87
-D7A+D7B+D7C=0	238824	80.39	3305	1.11	54949	18.5
D8D-D8D1-D8D2=0	294964	99.29	342	0.12	1772	0.6
-D7C-D8D+D9=0	242191	81.52	2646	0.89	52241	17.58
D9-D9A-D9B-D9C-D9D-D9E-D9F-D9G-D9H-D9I=0	261880	88.15	14315	4.82	20883	7.03
D7A-D7B>=0	297009	99.98	0	0	69	0.02
D8D-D8D1>=0	296873	99.93	0	0	205	0.07
D8D-D8D2>=0	296708	99.88	0	0	370	0.12
-D14A1+D14A2>=0	293461	98.78	1780	0.6	1837	0.62
D14A1-1.0E-6*D14A2>=0	296946	99.96	78	0.03	54	0.02
-D14B1+D14B2>=0	295639	99.52	650	0.22	789	0.27
D14B1-1.0E-6*D14B2>=0	297030	99.98	17	0.01	31	0.01
-D14C1+D14C2>=0	296877	99.93	57	0.02	144	0.05
D14C1-1.0E-6*D14C2>=0	297077	100	1	0	0	0
-D14D1+D14D2>=0	296611	99.84	186	0.06	281	0.09
D14D1-1.0E-6*D14D2>=0	297075	100	2	0	1	0
D14A1>=0	297078	100	0	0	0	0
D14A2>=0	297078	100	0	0	0	0
D14B1>=0	297078	100	0	0	0	0
D14B2>=0	297078	100	0	0	0	0
D14C1>=0	297078	100	0	0	0	0
D14C2>=0	297078	100	0	0	0	0
D14D1>=0	297078	100	0	0	0	0
D14D2>=0	297078	100	0	0	0	0
D7A>=0	297078	100	0	0	0	0
D7B>=0	297078	100	0	0	0	0
D7B1>=0	297078	100	0	0	0	0
D7B2>=0	297078	100	0	0	0	0
D7C>=0	297078	100	0	0	0	0
D8D>=0	297078	100	0	0	0	0

D8D1>=0	297078	100	0	0	0	0
D8D2>=0	297078	100	0	0	0	0
D9>=0	297078	100	0	0	0	0
D9A>=0	297078	100	0	0	0	0
D9B>=0	297078	100	0	0	0	0
D9C>=0	297078	100	0	0	0	0
D9D>=0	297078	100	0	0	0	0
D9E>=0	297078	100	0	0	0	0
D9F>=0	297078	100	0	0	0	0
D9G>=0	297078	100	0	0	0	0
D9H>=0	297078	100	0	0	0	0
D9I>=0	297078	100	0	0	0	0

The most frequent cause of incoherencies derives from the rule: *Total area owned by the household = area owned and used by the household + area owned and given to others*; 20% of the records presented an error in that rule.

The second cause of errors comes from the requested equation: *The use of the household's total land as of census day = area owned and used by the household + total area rented or used from others in Kosovo*; in this case errors were found in 19% of the records;
The third type of errors derived from inconsistencies in the fields related to the details of the use of the land (all the questions in section (9)). In general it is possible to conclude that most of the errors derive from the not respect of the balancing variables (i.e. those variables that are used to verify the totals).

The table below presents the frequencies of single edit not satisfied or edits not satisfied in association with others, and shows to what extent an error in one rule could be associated to an error in another rule.

Number of single or joint edits not satisfied	Number of corresponding records (absolute values)	Number of corresponding records (percentage on total records)
1	23643	7.96
2	43528	14.65
3	14206	4.78
4	1252	0.42
5	232	0.08
6	53	0.02
7	6	0

The last interesting analysis was conducted in order to verify the fields that were more frequently cause of errors; in this case the results in the table below refer to the end of the Editing and

Imputation steps, in which the algorithm was able to identify how many times each variable was considered erroneous and, so, required to be imputed.

Variable	Index
What is the total area of land owned and really usable by the household? (ha*100+are)	68.08
What is the use of households total land as of census day (ha*100+are)	44.65
Arable land (annual crops) (ha*100+are)	36.24
Other (house yard, ponds,...) (ha*100+are)	30.80
Meadows (clover, mixed grass, etc.) (ha*100+are)	30.48
What is the total area owned by household? (ha*100+are)	30.32
Forests (ha*100+are)	29.71
Fallow (more than 2 years not cultivated field) (ha*100+are)	26.70
Pasture (ha*100+are)	25.75
Orchards (ha*100+are)	25.25
Vineyards (ha*100+are)	25.05
Greenhouses (ha*100+are)	24.89
What is the total area rented and used by others in Kosovo? (ha*100+are)	14.54
What is the total area rented and used by others in Kosovo (used for free)? (ha*100+are)	10.55
What is the total area rented and used by others in Kosovo (for money)? (ha*100+are)	10.01
What is the total area of land owned and given to others? (ha*100+are)	6.52
Does the household own tractor single axle (hp)	6.43
What is the total area of land owned and given to others for free? (ha*100+are)	3.89
What is the total area of land owned and given to others for money? (ha*100+are)	3.30
Does the household own tractor two axletree (hp)	2.45
Does the household own truck (hp)	0.73
Does the household own cereals combines (hp)	0.29
Does the household own tractor single axle (number)	0.25
Does the household own tractor two axletree (number)	0.08
Does the household own truck (number)	0.01
Does the household own cereals combines (number)	0.00

This table shows the value of the indexes that were evaluated, for each field and after the editing steps, in order to measure the number of times a variable was identified as erroneous (greater values of the index identifies a variable often erroneous).

In detail the index is calculated by considering the percentage of times, on all the erroneous records, the variable was changed.

It is interesting to observe that, during the editing and imputation steps, the number of deterministic solutions that were identified was 76,189, on a total of records that did not satisfy the rules equal

to 82,920. It must be noticed that a deterministic solution identifies an error that happened not for random reasons but for systematic reasons (in this case it is possible to consider a misunderstanding of some questions in the questionnaires).

The part of the census data on household characteristics appears to be of a good quality; the bigger source of errors comes from the interpretation of the balancing questions and not from missing or incoherent answers in the qualitative fields.

In particular the question that was in many cases not understood was the necessity of writing the dimension of the area owned and used by the household (question 7.c), as a simple subtraction of two fields. Furthermore it should be mentioned that in 10 percent of the cases the *skips* implicit in the values *1* or *4* for the variable *Tenure status* were not respected.

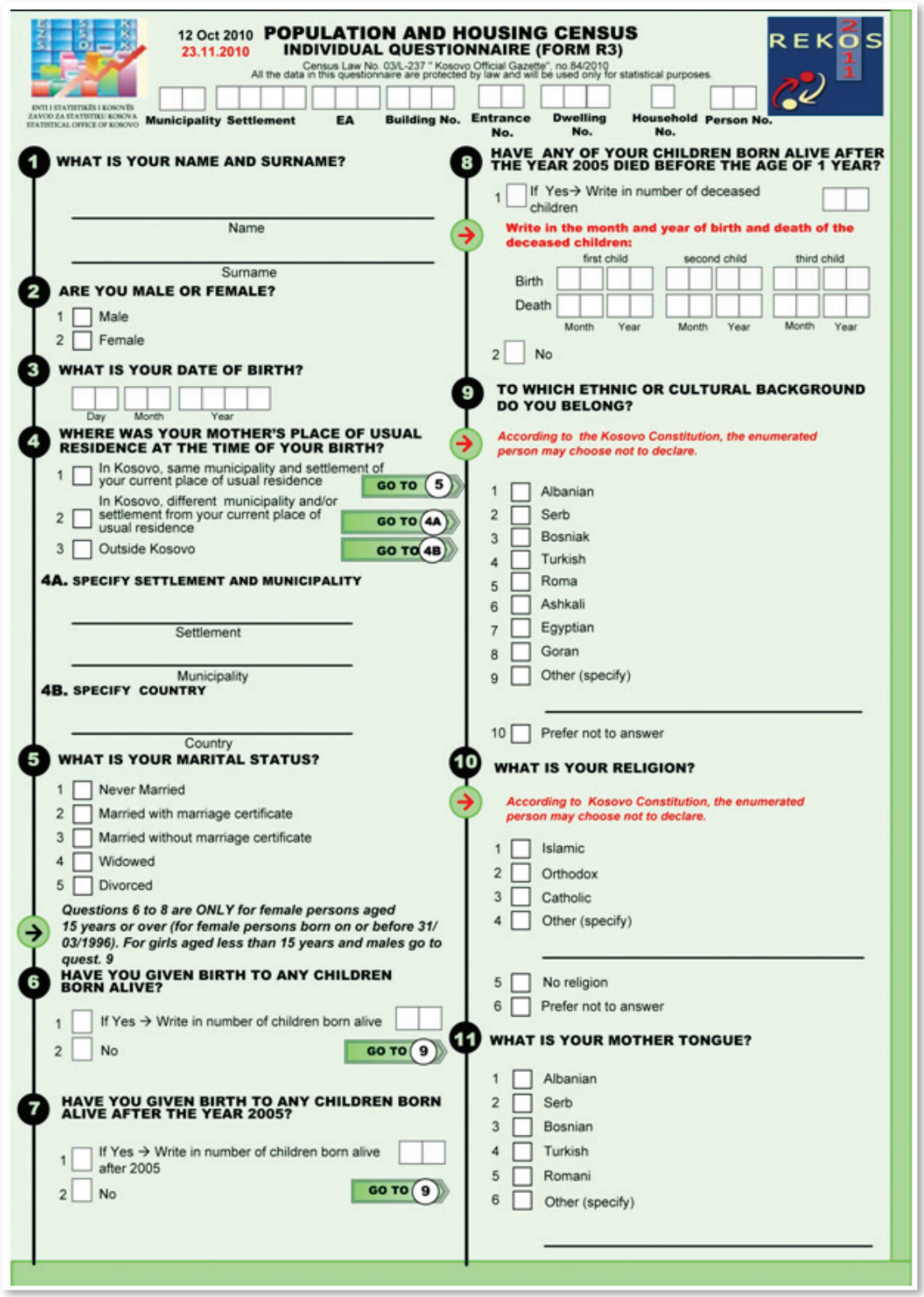
Annex 2 presents the results of editing and imputation of R2 questionnaires, by confronting the initial (or raw) data with the ones obtained by the cleaning process.

3.4.3 Data on individuals

The Form R3 – population and housing census – individual questionnaire collected all personal data about each usual resident of Kosovo. Most important demographics are collected for all enumerated people: age, sex, marital status, place of residence of the mother at the moment of respondent's birth, ethnicity, religion, etc, Many questions, however, had a filter based on age (education and labour history, which are relevant only after a threshold age) and, for some, also on sex (fertility facts, which concern only women having completed a certain age).

Usually, data on each household member were collected by the head of household (see Form R2, List 1). Cultural data for children were attributed by the head of household (for example, the mother tongue of children who are under the age of being able to speak was the one they were suppose to learn at home).

The individual questionnaires are presented hereafter.



12 Oct 2010
23.11.2010

POPULATION AND HOUSING CENSUS
INDIVIDUAL QUESTIONNAIRE (FORM R3)

Census Law No. 03/L-237 " Kosovo Official Gazette", no.84/2010
All the data in this questionnaire are protected by law and will be used only for statistical purposes.

REKOS 2011

ENTI I STATISTIKËS I KOSOVËS
ZAVOD ZA STATISTIKU KOSOVA
STATISTICAL OFFICE OF KOSOVO

Municipality | Settlement | EA | Building No. | Entrance No. | Dwelling No. | Household No. | Person No.

1 WHAT IS YOUR NAME AND SURNAME?

Name

Surname

2 ARE YOU MALE OR FEMALE?

1 Male
2 Female

3 WHAT IS YOUR DATE OF BIRTH?

Day | Month | Year

4 WHERE WAS YOUR MOTHER'S PLACE OF USUAL RESIDENCE AT THE TIME OF YOUR BIRTH?

1 In Kosovo, same municipality and settlement of your current place of usual residence → GO TO 5
2 In Kosovo, different municipality and/or settlement from your current place of usual residence → GO TO 4A
3 Outside Kosovo → GO TO 4B

4A. SPECIFY SETTLEMENT AND MUNICIPALITY

Settlement

Municipality

4B. SPECIFY COUNTRY

Country

5 WHAT IS YOUR MARITAL STATUS?

1 Never Married
2 Married with marriage certificate
3 Married without marriage certificate
4 Widowed
5 Divorced

Questions 6 to 8 are ONLY for female persons aged 15 years or over (for female persons born on or before 31/03/1996). For girls aged less than 15 years and males go to quest. 9

6 HAVE YOU GIVEN BIRTH TO ANY CHILDREN BORN ALIVE?

1 If Yes → Write in number of children born alive
2 No → GO TO 9

7 HAVE YOU GIVEN BIRTH TO ANY CHILDREN BORN ALIVE AFTER THE YEAR 2005?

1 If Yes → Write in number of children born alive after 2005
2 No → GO TO 9

8 HAVE ANY OF YOUR CHILDREN BORN ALIVE AFTER THE YEAR 2005 DIED BEFORE THE AGE OF 1 YEAR?

1 If Yes→ Write in number of deceased children

Write in the month and year of birth and death of the deceased children:

	first child		second child		third child	
Birth						
Death						
	Month	Year	Month	Year	Month	Year

2 No

9 TO WHICH ETHNIC OR CULTURAL BACKGROUND DO YOU BELONG?

According to the Kosovo Constitution, the enumerated person may choose not to declare.

1 Albanian
2 Serb
3 Bosniak
4 Turkish
5 Roma
6 Ashkali
7 Egyptian
8 Goran
9 Other (specify)

10 Prefer not to answer

10 WHAT IS YOUR RELIGION?

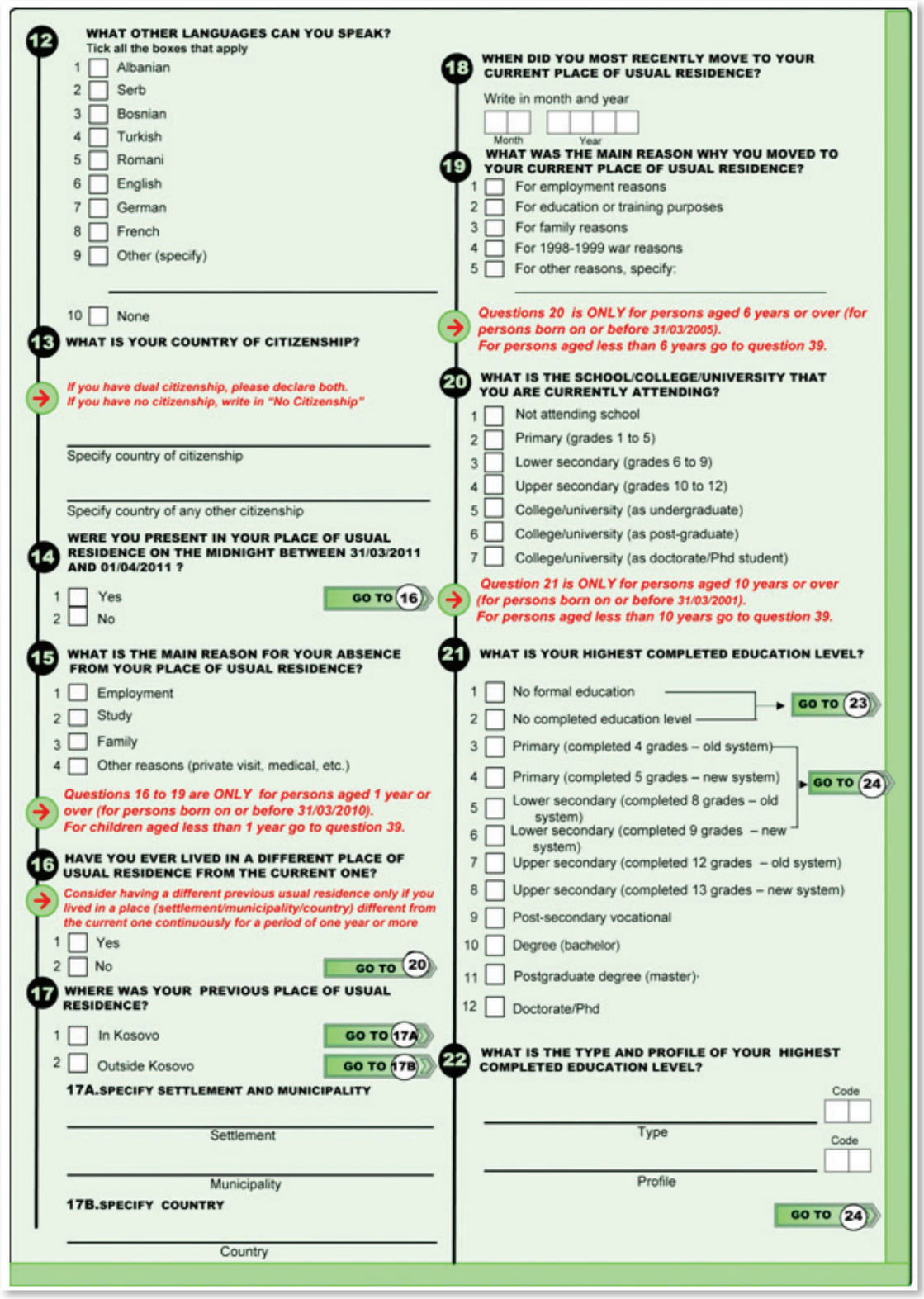
According to Kosovo Constitution, the enumerated person may choose not to declare.

1 Islamic
2 Orthodox
3 Catholic
4 Other (specify)

5 No religion
6 Prefer not to answer

11 WHAT IS YOUR MOTHER TONGUE?

1 Albanian
2 Serb
3 Bosnian
4 Turkish
5 Romani
6 Other (specify)



12 WHAT OTHER LANGUAGES CAN YOU SPEAK?
Tick all the boxes that apply

1 ☐ Albanian
2 ☐ Serb
3 ☐ Bosnian
4 ☐ Turkish
5 ☐ Romani
6 ☐ English
7 ☐ German
8 ☐ French
9 ☐ Other (specify)

10 ☐ None

13 WHAT IS YOUR COUNTRY OF CITIZENSHIP?

If you have dual citizenship, please declare both.
If you have no citizenship, write in "No Citizenship"

Specify country of citizenship

Specify country of any other citizenship

14 WERE YOU PRESENT IN YOUR PLACE OF USUAL RESIDENCE ON THE MIDNIGHT BETWEEN 31/03/2011 AND 01/04/2011 ?

1 ☐ Yes → GO TO 16
2 ☐ No

15 WHAT IS THE MAIN REASON FOR YOUR ABSENCE FROM YOUR PLACE OF USUAL RESIDENCE?

1 ☐ Employment
2 ☐ Study
3 ☐ Family
4 ☐ Other reasons (private visit, medical, etc.)

Questions 16 to 19 are ONLY for persons aged 1 year or over (for persons born on or before 31/03/2010).
For children aged less than 1 year go to question 39.

16 HAVE YOU EVER LIVED IN A DIFFERENT PLACE OF USUAL RESIDENCE FROM THE CURRENT ONE?

Consider having a different previous usual residence only if you lived in a place (settlement/municipality/country) different from the current one continuously for a period of one year or more

1 ☐ Yes
2 ☐ No → GO TO 20

17 WHERE WAS YOUR PREVIOUS PLACE OF USUAL RESIDENCE?

1 ☐ In Kosovo → GO TO 17A
2 ☐ Outside Kosovo → GO TO 17B

17A. SPECIFY SETTLEMENT AND MUNICIPALITY

Settlement

Municipality

17B. SPECIFY COUNTRY

Country

18 WHEN DID YOU MOST RECENTLY MOVE TO YOUR CURRENT PLACE OF USUAL RESIDENCE?

Write in month and year

Month Year

19 WHAT WAS THE MAIN REASON WHY YOU MOVED TO YOUR CURRENT PLACE OF USUAL RESIDENCE?

1 ☐ For employment reasons
2 ☐ For education or training purposes
3 ☐ For family reasons
4 ☐ For 1998-1999 war reasons
5 ☐ For other reasons, specify:

Questions 20 is ONLY for persons aged 6 years or over (for persons born on or before 31/03/2005).
For persons aged less than 6 years go to question 39.

20 WHAT IS THE SCHOOL/COLLEGE/UNIVERSITY THAT YOU ARE CURRENTLY ATTENDING?

1 ☐ Not attending school
2 ☐ Primary (grades 1 to 5)
3 ☐ Lower secondary (grades 6 to 9)
4 ☐ Upper secondary (grades 10 to 12)
5 ☐ College/university (as undergraduate)
6 ☐ College/university (as post-graduate)
7 ☐ College/university (as doctorate/Phd student)

Question 21 is ONLY for persons aged 10 years or over (for persons born on or before 31/03/2001).
For persons aged less than 10 years go to question 39.

21 WHAT IS YOUR HIGHEST COMPLETED EDUCATION LEVEL?

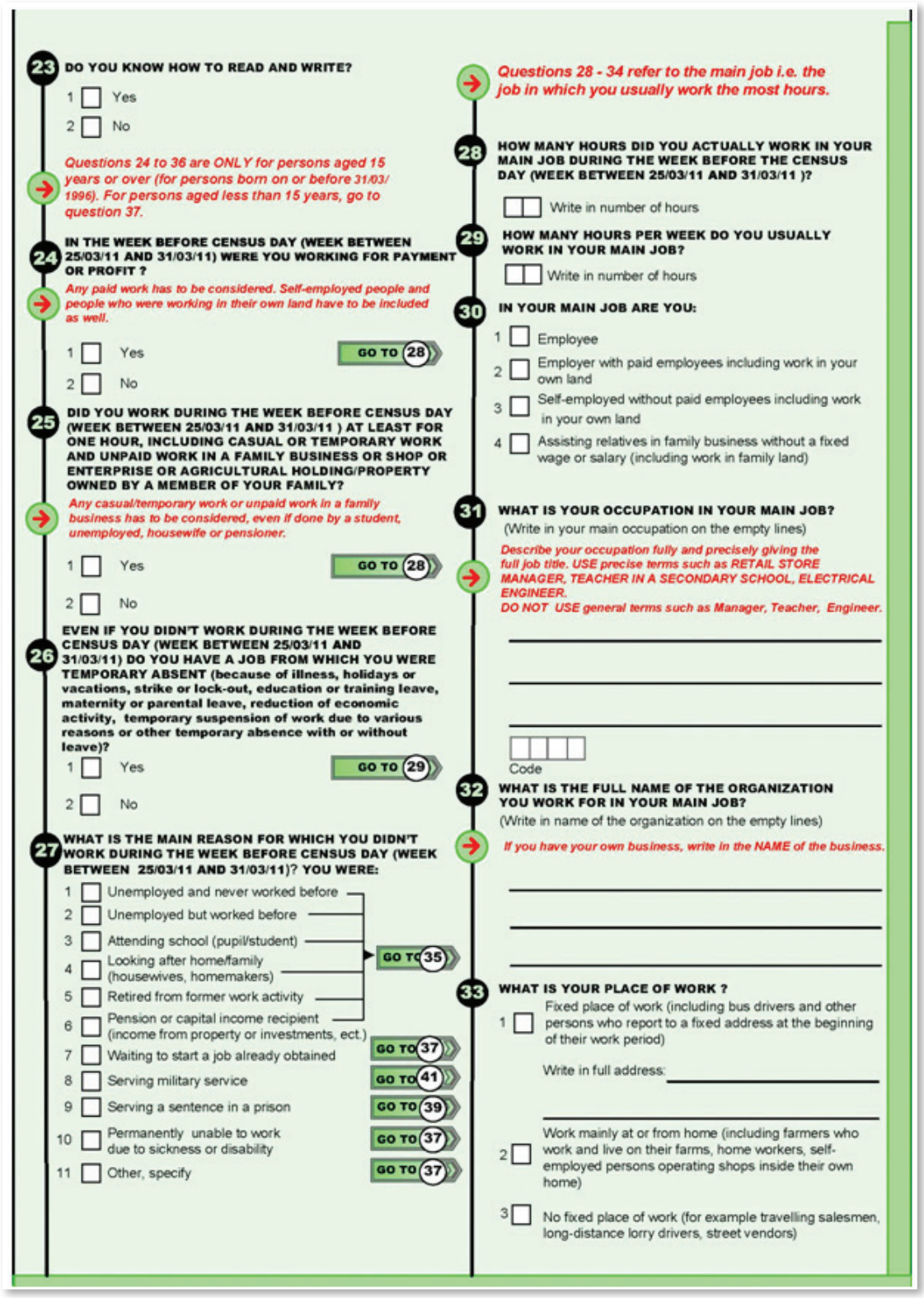
1 ☐ No formal education → GO TO 23
2 ☐ No completed education level → GO TO 23
3 ☐ Primary (completed 4 grades – old system)
4 ☐ Primary (completed 5 grades – new system) → GO TO 24
5 ☐ Lower secondary (completed 8 grades – old system)
6 ☐ Lower secondary (completed 9 grades – new system)
7 ☐ Upper secondary (completed 12 grades – old system)
8 ☐ Upper secondary (completed 13 grades – new system)
9 ☐ Post-secondary vocational
10 ☐ Degree (bachelor)
11 ☐ Postgraduate degree (master)
12 ☐ Doctorate/Phd

22 WHAT IS THE TYPE AND PROFILE OF YOUR HIGHEST COMPLETED EDUCATION LEVEL?

Type Code

Profile Code

→ GO TO 24



23 DO YOU KNOW HOW TO READ AND WRITE?

1 ☐ Yes
2 ☐ No

Questions 24 to 36 are ONLY for persons aged 15 years or over (for persons born on or before 31/03/1996). For persons aged less than 15 years, go to question 37.

24 IN THE WEEK BEFORE CENSUS DAY (WEEK BETWEEN 25/03/11 AND 31/03/11) WERE YOU WORKING FOR PAYMENT OR PROFIT ?

Any paid work has to be considered. Self-employed people and people who were working in their own land have to be included as well.

1 ☐ Yes — GO TO 28
2 ☐ No

25 DID YOU WORK DURING THE WEEK BEFORE CENSUS DAY (WEEK BETWEEN 25/03/11 AND 31/03/11) AT LEAST FOR ONE HOUR, INCLUDING CASUAL OR TEMPORARY WORK AND UNPAID WORK IN A FAMILY BUSINESS OR SHOP OR ENTERPRISE OR AGRICULTURAL HOLDING/PROPERTY OWNED BY A MEMBER OF YOUR FAMILY?

Any casual/temporary work or unpaid work in a family business has to be considered, even if done by a student, unemployed, housewife or pensioner.

1 ☐ Yes — GO TO 28
2 ☐ No

26 EVEN IF YOU DIDN'T WORK DURING THE WEEK BEFORE CENSUS DAY (WEEK BETWEEN 25/03/11 AND 31/03/11) DO YOU HAVE A JOB FROM WHICH YOU WERE TEMPORARY ABSENT (because of illness, holidays or vacations, strike or lock-out, education or training leave, maternity or parental leave, reduction of economic activity, temporary suspension of work due to various reasons or other temporary absence with or without leave)?

1 ☐ Yes — GO TO 29
2 ☐ No

27 WHAT IS THE MAIN REASON FOR WHICH YOU DIDN'T WORK DURING THE WEEK BEFORE CENSUS DAY (WEEK BETWEEN 25/03/11 AND 31/03/11)? YOU WERE:

1 ☐ Unemployed and never worked before
2 ☐ Unemployed but worked before
3 ☐ Attending school (pupil/student)
4 ☐ Looking after home/family (housewives, homemakers)
5 ☐ Retired from former work activity
6 ☐ Pension or capital income recipient (income from property or investments, ect.)
— GO TO 35
7 ☐ Waiting to start a job already obtained — GO TO 37
8 ☐ Serving military service — GO TO 41
9 ☐ Serving a sentence in a prison — GO TO 39
10 ☐ Permanently unable to work due to sickness or disability — GO TO 37
11 ☐ Other, specify — GO TO 37

Questions 28 - 34 refer to the main job i.e. the job in which you usually work the most hours.

28 HOW MANY HOURS DID YOU ACTUALLY WORK IN YOUR MAIN JOB DURING THE WEEK BEFORE THE CENSUS DAY (WEEK BETWEEN 25/03/11 AND 31/03/11)?

☐☐ Write in number of hours

29 HOW MANY HOURS PER WEEK DO YOU USUALLY WORK IN YOUR MAIN JOB?

☐☐ Write in number of hours

30 IN YOUR MAIN JOB ARE YOU:

1 ☐ Employee
2 ☐ Employer with paid employees including work in your own land
3 ☐ Self-employed without paid employees including work in your own land
4 ☐ Assisting relatives in family business without a fixed wage or salary (including work in family land)

31 WHAT IS YOUR OCCUPATION IN YOUR MAIN JOB?

(Write in your main occupation on the empty lines)

Describe your occupation fully and precisely giving the full job title. USE precise terms such as RETAIL STORE MANAGER, TEACHER IN A SECONDARY SCHOOL, ELECTRICAL ENGINEER.
DO NOT USE general terms such as Manager, Teacher, Engineer.

_____
_____
_____

☐☐☐☐ Code

32 WHAT IS THE FULL NAME OF THE ORGANIZATION YOU WORK FOR IN YOUR MAIN JOB?

(Write in name of the organization on the empty lines)

If you have your own business, write in the NAME of the business.

_____
_____
_____

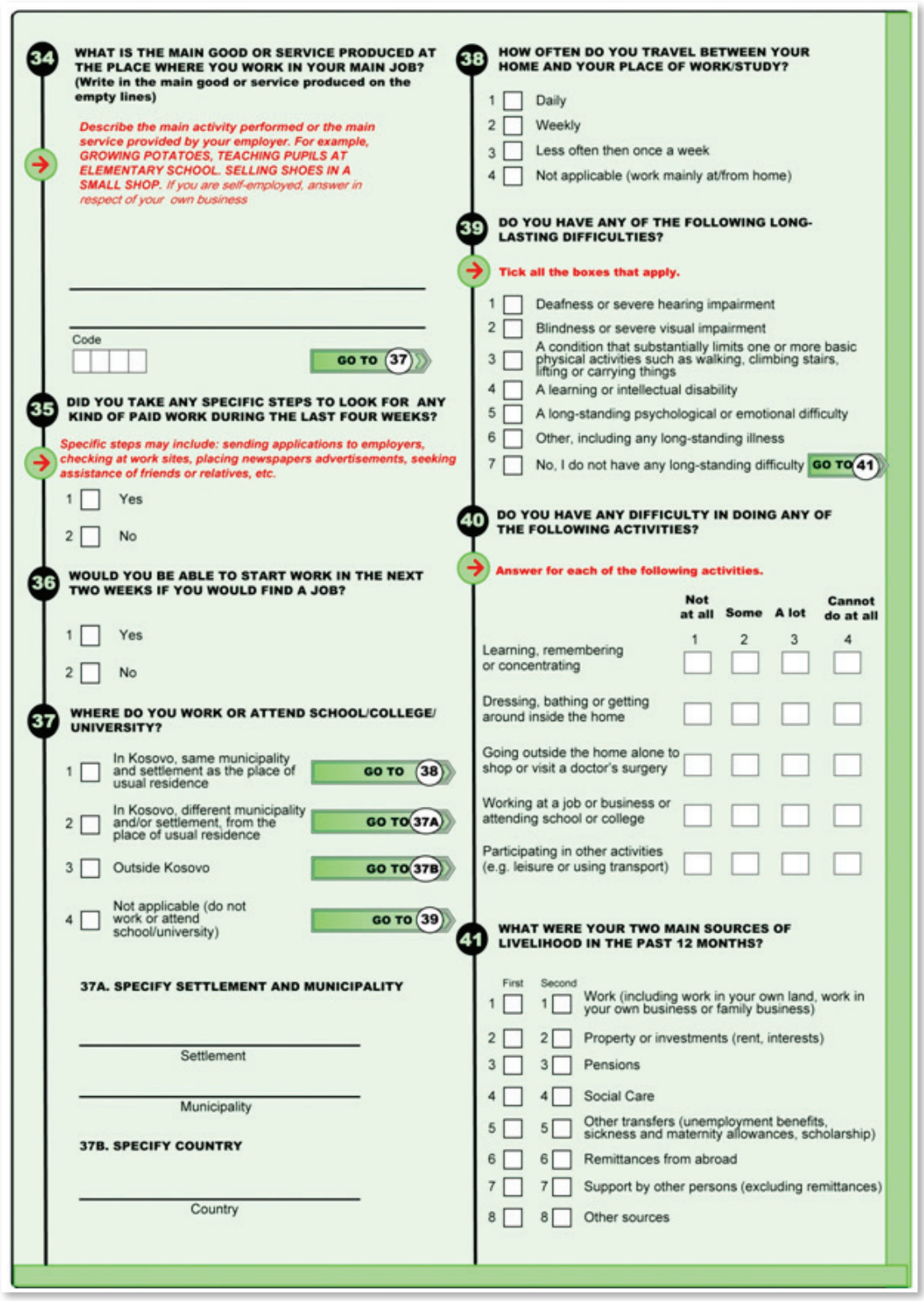
33 WHAT IS YOUR PLACE OF WORK ?

1 ☐ Fixed place of work (including bus drivers and other persons who report to a fixed address at the beginning of their work period)

Write in full address: _____
_____

2 ☐ Work mainly at or from home (including farmers who work and live on their farms, home workers, self-employed persons operating shops inside their own home)

3 ☐ No fixed place of work (for example travelling salesmen, long-distance lorry drivers, street vendors)



34 WHAT IS THE MAIN GOOD OR SERVICE PRODUCED AT THE PLACE WHERE YOU WORK IN YOUR MAIN JOB? (Write in the main good or service produced on the empty lines)

Describe the main activity performed or the main service provided by your employer. For example, GROWING POTATOES, TEACHING PUPILS AT ELEMENTARY SCHOOL. SELLING SHOES IN A SMALL SHOP. If you are self-employed, answer in respect of your own business

Code

GO TO 37

35 DID YOU TAKE ANY SPECIFIC STEPS TO LOOK FOR ANY KIND OF PAID WORK DURING THE LAST FOUR WEEKS?

Specific steps may include: sending applications to employers, checking at work sites, placing newspapers advertisements, seeking assistance of friends or relatives, etc.

1 Yes
2 No

36 WOULD YOU BE ABLE TO START WORK IN THE NEXT TWO WEEKS IF YOU WOULD FIND A JOB?

1 Yes
2 No

37 WHERE DO YOU WORK OR ATTEND SCHOOL/COLLEGE/ UNIVERSITY?

1 In Kosovo, same municipality and settlement as the place of usual residence — GO TO 38
2 In Kosovo, different municipality and/or settlement, from the place of usual residence — GO TO 37A
3 Outside Kosovo — GO TO 37B
4 Not applicable (do not work or attend school/university) — GO TO 39

37A. SPECIFY SETTLEMENT AND MUNICIPALITY

Settlement

Municipality

37B. SPECIFY COUNTRY

Country

38 HOW OFTEN DO YOU TRAVEL BETWEEN YOUR HOME AND YOUR PLACE OF WORK/STUDY?

1 Daily
2 Weekly
3 Less often then once a week
4 Not applicable (work mainly at/from home)

39 DO YOU HAVE ANY OF THE FOLLOWING LONG-LASTING DIFFICULTIES?

Tick all the boxes that apply.

1 Deafness or severe hearing impairment
2 Blindness or severe visual impairment
3 A condition that substantially limits one or more basic physical activities such as walking, climbing stairs, lifting or carrying things
4 A learning or intellectual disability
5 A long-standing psychological or emotional difficulty
6 Other, including any long-standing illness
7 No, I do not have any long-standing difficulty — GO TO 41

40 DO YOU HAVE ANY DIFFICULTY IN DOING ANY OF THE FOLLOWING ACTIVITIES?

Answer for each of the following activities.

	Not at all (1)	Some (2)	A lot (3)	Cannot do at all (4)
Learning, remembering or concentrating				
Dressing, bathing or getting around inside the home				
Going outside the home alone to shop or visit a doctor's surgery				
Working at a job or business or attending school or college				
Participating in other activities (e.g. leisure or using transport)				

41 WHAT WERE YOUR TWO MAIN SOURCES OF LIVELIHOOD IN THE PAST 12 MONTHS?

	First	Second	
1	☐	1 ☐	Work (including work in your own land, work in your own business or family business)
2	☐	2 ☐	Property or investments (rent, interests)
3	☐	3 ☐	Pensions
4	☐	4 ☐	Social Care
5	☐	5 ☐	Other transfers (unemployment benefits, sickness and maternity allowances, scholarship)
6	☐	6 ☐	Remittances from abroad
7	☐	7 ☐	Support by other persons (excluding remittances)
8	☐	8 ☐	Other sources

The individual questionnaire comprises 41 questions that can be classified as follows, according to the filters used:

Questions asked to all respondents:

1. Basic demographics: Q1 to Q5

In this part, name and surname have been used only to check eventual duplicates or missing persons from the List 1 in R2 form. Nominal data have been encrypted at the data entry phase.

2. Personal characteristics: Q9 to Q15

These questions relate to the ethnic and religious background, mother tongue, knowledge of other languages, citizenship, and the presence at usual residence at census day.

3. Data on disability and source of livelihood: Q39 to Q41

Some inconsistencies were found in these questions for very young children. Indeed, the interpretation of the questions was not always easy in case of babies. For instance, many questionnaires relating to new-born indicated "work" as main source of livelihood (instead of "family support"), most probably referring to the fact that the father was working. Similarly, disability questions for new-born or very young children were incongruous such as Q40 "Do you have any difficulty in doing any of the following activities?... Learning, remembering or concentrating; ... Dressing, bathing or getting around inside the home; ..etc."

Questions asked only to women aged 15 years and over

4. Fertility and infant mortality questions: Q6 to Q8

Unfortunately, fertility data collected in the census are not sufficient to calculate total fertility rates. These data on fertility have been of poor quality as regards the question 8, where not only children dead before 1 year of age were reported. Some inconsistencies were also found in the number of children ever born for some young females, and in some cases in the number of children born after 2005.

Question asked only to persons aged 6 years and over

5. Education level currently attended: Q20

Schooling starting at the age of 6 in Kosovo, this question was asked only to persons having completed this age. It has been observed that this school attendance question was sometimes confused with completed education level, which was the next question.

Questions asked only to persons aged 10 years and over

6. Highest completed education level and literacy: Q21 to Q23

The answers took into account the changes occurred from the year 2000 in the education system of Kosovo (old system/new system). The type and profile (Q22) helped in coding data according to the ISCED classification.

7. Commuting characteristics: Q37 to Q38

The origin-destination movements for study or working reasons and their frequency were collected with these questions.

Questions asked only to persons aged 1 year and over

8. Migration history: Q16 to Q19

Questions refer to the last change of residence only and collect origin of the last movement. Reason for the migration was also collected as well as the date of the last migration. Some problems were found in these data as some persons reported the last migration of the head of household as their own migration.

Questions asked only to persons aged 15 years and over

9. Characteristics of employed persons: Q24 to Q26; Q28 to Q34

Employment – and in general activity – questions have been often misunderstood. Occupations and activity sector, coded manually according to ISCO and NACE classifications had therefore been disseminated only at 2 digits level. Editing and imputations of all these categories of questions has been also quite complex, due to the inter-linkage of all variables between them.

10. Characteristics of unemployed and inactive persons: Q27; Q35 to Q36

Despite the fact that the census questionnaires underwent several field and pilot tests some questions were still incompatible with Kosovo reality (serving military service does not exist in Kosovo) or wrongly understood (place of work inconsistent with commuting history; unemployed housewives coded as workers in service sector; etc.).

Preliminary treatment

The goal of this phase was to prepare all variables necessary for data cleaning. The structure of output record was described in a specific EXCEL file.

Steps performed by the software application:

Read R3 – Individual raw data;
Eliminate duplicated records;
Prepare a unique code for R3 data;

Merge with list of usually resident persons (R2 – list1) adding flag of presence in R2 and R2 variables SEX (HH1_04), DATE OF BIRTH (HH1_05), AGE (HH1_AGE), RELATIONSHIP WITH THE HEAD OF HOUSEHOLD (HH1_06_B), ORDINAL NUMBER OF FAMILY (HH1_07) and FAMILY STATUS (HH1_08_B);
Fill R3 variables Municipality, settlement, EA, building, entrance, dwelling, household number and individual number (ID codes) and sex, age, date of birth and, when possible, marital status;
Add new variables as error flags, unique code, and other flags;
Sort by ID codes: Municipality, settlement, EA, building, entrance, dwelling, household number and individual number;
Write R3 file with new variables.
Write a quantitative report of performed operation.

Systematic errors processing

Main reasons for deterministic imputation of errors were:
Values of variables asked also in R2 form (sex, age) are considered more reliable than values of R3 form;
Sensible variables as ethnicity or religion are not imputed;
Routing errors: deterministic imputation of routing errors if values are coherent with the rule (as recommended by the IMO) or skip rule is based on a reliable variable (i.e. males responding questions about fertility);
A filter question missing and detailed questions answered.

No answers were accepted for the following questions:

Question 4 – Where was your mother's place of usual residence at the time of your birth?
Question 6 – Have you given birth to any children born alive?
Question 6B - Number of Born Alive Children
Question 7 – Have you given birth to any children born alive after the year 2005?
Question 7B - Number of children born alive after 2005
Question 8 – Have you given birth to any children born alive after the year 2005 died before the age of 1 year?
Question 8B - Number of Born Alive after the year 2005 Children died before the age of 1 year
Question 9 – To which ethnic or cultural background do you belong?
Question 10 – What is your religion?
Question 11 – What is your mother tongue?
Question 12 – What other languages do you speak (First- Ninth)?
Question 13 – What is your country of citizenship?
Question 17 – Where was your previous place of usual residence? No answer is accepted for settlement, municipality and country codes.

Question 35 – Did you look for a paid work in last four week?
Question 36 – Would you be able to start work in the next two weeks?
Question 39 – Do you have any of the following long-lasting difficulties?
Question 40 – Do you have any difficulty in doing any of the following activities?

Questions 8 (All dates), 31, 32, 33 (Address), 34 are not involved in data cleaning.

In the following table there is the list of deterministic imputations performed for individual questionnaire.
The error conditions are described in the CONDITION column. The corresponding imputations are under ACTION.

Deterministic imputations

ERROR CODE	QUESTION	CONDITION	ACTION
		Computing of AGE for people born on 1st April	Subtract 1 from variable AGE
E300		*Empty record* - Records with all variables of R3 Form missing (probably corresponding to missing R3 Form in presence of individual in R2 – List1 Form)	Fill R3 sex, date of birth and age and, if possible, marital status with the corresponding R2 values and follow normal check process.
E301	*2*	*Q2 - Sex*: Missing sex Different sex values between R2 -list1 and R3	Fill R3 sex with the corresponding R2 value, if it is a valid one
E302	*3*	*Q3 – Date of birth:* Missing or different date of birth from R2 list1 and R3 If R3 Date of birth/age is missing or invalid or different from R2 valid date of birth/age	R3 date of birth (IN_03) assumes R2 date of birth (HH1_05) valid values.
E303	*3*	*Q3 – Date of birth: Age of head of household <= 15 years (first person in the household):* In this case the Age and year of birth fields are prepared for Scia imputation.	R2, R3 Year of birth and AGE are changed in -1 (no response value)
E304	*3*	*Q3 - Born after Census Date (1/4/2011)* Erroneous year of birth (prepared for Scia imputation)	R2, R3 Year of birth = -1

E309	*4*	*Q4 - mothers place of usual residence (total changes)* Q4A answer 2 or 3 is accepted even if codes are not specified (no answer) If Q4A is missing (-1 or space) and municipality and settlement codes are filled If Q4A is equal to 1 (in Kosovo, same municipality and settlement) or 2 (in Kosovo, different municipality or settlement) and municipality and settlement codes are filled and municipality and settlement code of birth are equal to municipality and settlement code of residence If Q4A is equal to 1 ((In Kosovo, same municipality and settlement) and municipality and settlement codes are filled and municipality and settlement code of birth are different to municipality and settlement code of residence.	Q4A = 2 (in Kosovo, different municipality or settlement) Q4A code = 1 (same municipality and settlement) and municipality/settlement codes = spaces Qq4A code = 2 (in Kosovo, different municipality or settlement)
E310	*6-8*	*Q6-8- Fertility Questions – Routing rule:* Fertility questions are ONLY for female persons aged 15 years or over IF AGE < 15 years or Sex = 2 (Male) and at least one question about fertility is answered	Q6-8 = spaces
	6	*Q6- Have You Given Birth To Any Children Born alive?* Number of children > 0 and Q6 missing or no Number of children = 0 and Q6 missing Q7 = Yes AND Q7B (number of children) > 0 AND Q6 = NO or missing Q7 Number of children<= 10 and greater than Q6 Number of children Number of children > 25	No answer permitted: code 0 Q6 = YES (1) Q6 = NO (2) Q6=Yes and Q6B=Q7B Q6B=Q7B Number of children = -1
	7	*Q7- Have You Given Birth To Any Children Born alive after 2005?* Number of children > 0 and Q7 missing or no Number of children = 0 and Q7 missing Q8 = Yes AND Q8B (number of children) > 0 AND Q7 = NO or missing Q7 Number of children<= 10 and greater than Q7 Number of children Number of children > 10 Age > 51	No answer permitted: code 0 Q7 = YES (1) Q7 = NO (2) Q6=Yes and Q6B=Q7B Q6B=Q7B Number of children = -1 Q7 = NO (2)

	8	*Q8- Have You Given Birth To Any Children Born alive after 2005 Died before the age of 1 year?* Number of children > 0 and Q8 missing or no Number of children = 0 and Q8 missing Number of children > 10 Age > 51 DATE OF BIRTH AND DEATH ARE NOT CHECKED	No answer permitted: code 0 Q8 = YES (1) Q8 = NO (2) Number of children = -1 Q8 = NO (2)
E306	*9*	*Q9 - Ethnic or cultural background*: If the ethnic or cultural background is missing (space or -1)	No answer permitted: code 0
E307	*10*	*Q10 - Religion*: If the religion is missing	No answer permitted: code 0
E308	*11*	*Q11 - Mother tongue:* If mother tongue is missing (-1 or space)	No answer permitted: code 0.
E312	*12*	*Q12 – What other languages can you speak?* Other language must be different from mother tongue: If Q12A (values 1-5) and Q12B is equal to space (no other languages) If Q12 is equal to Q11 (values 1-5) and Q12b is different from space	No answer permitted: code 0 Q12A = 10 (None) Shift left Q12B-Q12I: Q12A=Q12B, Q12B=Q12C,...
E313	*13*	*Q13 - country of citizenship* If Q13A is equal Q13B	No answer permitted: code 0 Q13B = space
E305	*14-15*	*Q14 - Present in usual residence:* If Q14 is missing or YES (1) and Q15 (reason for absence) has a valid answer	Q14 = NO (2)

E311	*16*	*Q16 - Q19: Previous place of usual residence (Total changes)* If Q16 - Q19 are all equal -1 (not answered) then Q16 = 2 (never lived in a different place) IF Q16 != 1 and Q17 Municipality and settlement are different from NULL or country is different from NULL IF Q16 = 1 (other place of residence) and Q17 municipality and settlement are equal to current municipality and settlement codes IF Q16 = 2 (same place of residence) and Q17 municipality and settlement are different from current municipality and settlement codes IF Q16 = 2 (same place of residence) and Q17 Country is not null No answer is valid for municipality, settlement and country codes	Q16 = 2 (same place of residence Q16 = 1 (other place of residence) Q16 = 2 a(same place of residence) and Q17 is skipped (spaces) Q16 = 1 (other place of residence) Q16 = 1 (other place of residence)
EDUCATION QUESTIONS (Q20-23)			
Deterministic imputation of routing errors if values are coherent with the rule			
E315	*20-23*	*Q20-23 - Education Questions – Routing rule: Q20 is for persons aged 6 years and over, Q21-23 for persons aged 10 years and over (Total)* Of which: values are coherent: attending primary school,...	Q20-23 = spaces
E316	*21-23*	*Q21-23 - Education Questions – Routing rule: Q21-23 for persons aged 10 years and over* Of which: values are coherent: attending primary school,...	Q21-23 = spaces
E318	*23*	*Q23 – Literacy – Routing rule: Q23 is for people without completed education level* Q23 = 1 (able to read and write) and Q21 > 2 (Educated)	Q23 = space
E317	*21*	*Q21 -* AGE = 10 and questions 20 = 2 (attending primary school) and Q21 = 1 (no formal education) or missing (-1)	Q21 = 2 (no completed education level)
E319	*20*	*Q20, Q21 – if Q21 (Highest completed education level greater than grade of school currently attended and AGE > 25 years)* IF (Age > 25 and Q20 = 2 (primary) and Q21 >= 3 (primary) or Q20 = 3 (lower secondary) and Q21 >= 5 (lower secondary)or Q20 = 4 (upper secondary) and Q21 >= 7 (upper secondary) or Q20 = 5 (college) and Q21 >= 9	Q20 = 1 (not attending school)

E329	*20*	*Q20, Q21 – Incoherence between answers of Q21(completed education level) and grade of attending school* - Attending primary school (Q20=2) AND no formal education (Q21 =1) - Attending primary school (Q20=2) AND Q21 = 3 (Primary –old system) and age between 9 and 11 - Attending lower secondary school (Q20=3) AND Q21 = 3 (Primary –old system) and age less than 17 - Attending upper secondary school (Q20=4) AND Q21 = 5 (Lower secondary – old system) and age less than 20 - Attending college (Q20=5) AND Q21 = 7 (Upper secondary – old system) and age less than 24	- Q21 = 2 (no completed education level) - Q21 = 2 (no completed education level) - Q21=4 (Primary school-new system) - Q21 = 6 (Lower secondary – new system) - Q21 = 8 (Upper secondary – new system)
E330	*21*	*Q21 and AGE – Change from old to new system* IF Q21 =3 AND AGE < 17 IF Q21 =5 AND AGE < 20 IF Q21 =7 AND AGE < 24	 Q21=4 Q21=6 Q21=8
		WORK QUESTIONS (Q24-36)	
E320	*24-36*	*Q24-36 – Questions about work – Routing rule:* work questions (24-36) are for persons aged 15 years and over IF AGE < 15 years and at least one question about work is answered	 Q24-36 = spaces
E323	*24-26*	*Q24 -26 Questions about work: there is a missing instead of a NO –* If first question is not answered and next question is answered Q25 -26 If Q24 is not answered and Q25 is answered THEN Q24 = NO (2) Q24 -26 If Q25 is not answered and Q26 is answered THEN Q25 = NO (2)	Preceding question = NO (2)
E321	*33*	*Q33 Place of work:* If code is missing (-1) and address is written.	Q33 code = 1 (fixed place of work)
E322	*37*	*Q37-Q38 – Routing rule:* Those questions are reserved for persons aged 10 years and over If age < 10 years and Q37 or Q38 are answered	Q37 = space Q38 = space
E324	*37*	*Q37 - Place of work/study*: Q37 not answered but municipality and settlement are declared (Total) of which Municipality and/or settlement different from those of residence of which Municipality and/or settlement equal to those of residence	 Q37 = 2 Q37 = 1

E325	*39-40*	Q39-Q40 - Long lasting difficulty - If no difficulty (Q39=7) and Q40 always equal to 1 (not at all) If no difficulty (Q39=7) and Q40 always equal to -1 (no answer) No duplicated answer	No answer permitted: code 0 Q40 is skipped (= space) Q40 is skipped (= space
E327	*41*	*Q41 – Main sources of livelihood - Shift (Total)* IF Q41A (first source of livelihood) is missing (-1) and Q41B (second source of livelihood) IF Q41A (first source of livelihood equal to Q41B (second source of livelihood)	Q41A = Q41B and Q41B = space Q41B = space

Random errors processing

To impute random errors in categorical variables, SCIA software was used for probabilistic imputations.
The edits of SCIA are incompatibilities rules that are activated when the record is an erroneous one.

There is an inconsistency when:

DESCRIPTION	EDIT
AGE OF HEAD OF HOUSEHOLD MUST BE GREATER OR EQUAL TO 12 YEARS	
Age of head of household must be greater or equal to 12 years	AGE (0-11) IND(1)
Age of Married head of household must be greater or equal to 15 years	AGE (12-15) IND(1) HH108B(1-5)
RELATIONS BETWEEN MARITAL STATUS (Q5) AND FAMILY STATUS (HH1_08_B)	
Marital status different from married with certificate and family status equal to married with certificate	IN05 <2) HH108B (1)
Marital status different from married with certificate and relationship to the head of household = husband/wife with certificate and ordinal number of family equal to 1	IN05<2) HH106B(2) HH107(1)
QUESTION 5 – MARITAL STATUS	
Age <= 15 years and marital status (Q5) different from single	IN05 <1) AGE(0-15)
Q14 – PRESENCE IN USUAL PLACE OF RESIDENCE Q15 – REASON FOR ABSEBCE FROM USUAL PLACE OF RESIDENCE	
IF Q14 = YES (PRESENT AT USUAL RESIDENCE) Q15 MUST BE MISSING	
Question 14 = 1 (yes) and question 15 (reason for absence) different from space	IN14 <2) IN15 <)
Question 14 = 2 (no) and question 15 (reason for absence) missing	IN14 (2) IN15 ()
SKIP RULE: QUESTIONS FROM 16 -38 ARE ONLY FOR PERSONS AGED 1 YEAR OR OVER	
Age less than 1 and one or more questions 16-38 answered	AGE (0) SKIP2 <)
Age greater than 0 and missing in Q16	AGE (1-120) IN16( )

Q16 – HAVE YOU EVER LIVED IN A DIFFERENT PLACE OF USUAL RESIDENCE Q17 – WHERE WAS YOUR PREVIOUS PLACE OF USUAL RESIDENCE Q18 – DATE OF MOVE TO CURRENT PLACE OF RESIDENCE Q19 – MAIN REASON FOR MOVING TO CURRENT PLACE OF RESIDENCE	
Question 16 (ever lived in different residence) = 1 (yes) and question 17 (previous place of residence) equal to from space	IN16 (1) IN17 ( )
Question 16 (ever lived in different residence) = 1 (yes) and question 18B (year of move) equal to space	IN16 (1) IN18B ( )
Question 16 (ever lived in different residence) = 1 (yes) and question 19A (main reason) equal to space	IN16 (1) IN19A ( )
Question 16 (ever lived in different residence) = 1 (yes) and question 17 (previous place of residence) different from space	IN16 (2) IN17 <)
Question 16 (ever lived in different residence) = 1 (yes) and question 18B (year of move) different from space	IN16 (2) IN18B <)
Question 16 (ever lived in different residence) = 2 (no) and question 19A (main reason) different from space	IN16 (2) IN19A <)
Question 17 (previous place of residence) = 1 (in Kosovo) and other country code different from space	IN17 (1) IN17C<)
Question 17 (previous place of residence) = 2 (outside Kosovo) and other country code equal to space	IN17 (2) IN17C( )
Question 17 (previous place of residence) = 1 (in Kosovo) and municipality and settlement codes equal to spaces	IN17 (1) IN17A( ) IN17B( )
Question 17 (previous place of residence) = 2 (outside Kosovo) and municipality code different from space	IN17 (2) IN17A<)
Question 17 (previous place of residence) = 2 (outside Kosovo) and settlement code different from space	IN17 (2) IN17B<)
Q20-23 QUESTIONS ABOUT EDUCATION	
Q20 –SCHOOL CURRENTLY ATTENDED	
ROUTING RULE: QUESTIONS 20 IS ONLY FOR PERSONS AGED 6 YEARS OR OVER	
Age < 6 years and question 20 (attending school) different from space	AGE (0-5) IN20 <)
Age >= 6 years and question 20 (attending school) not answered (equal to space)	AGE (6-120) IN20 ( )
ROUTING RULE: QUESTIONS 21-23 ARE ONLY FOR PERSONS AGED 10 YEARS AND OVER	
Question 21 (Highest completed education level) answered (different from spaces) and age under 10 years	AGE (0-9) IN21 <)
Question 21 (Highest completed education level) missing and age greater or equal 10 years	AGE (10-120) IN21 ( )
Question 22B (Highest education level profile) is only for people aged 10 years and over	AGE (0-9) IN22B <)

Question 23 (Literacy) is only for people aged 10 years and over	AGE (0-9) IN23 <)
Q21 - HIGHEST COMPLETED EDUCATION LEVEL Q23 – DO YOU KNOW HOW TO READ AND WRITE	
IF YOU HAVE NO FORMAL EDUCATION YOU MUST ASWER YOU CAN READ AND WRITE	
Question 21 = 1,2 (no formal education or no completed education level) and question 23 (literacy) missing and age is greater or equal to 10 years	IN21 (1,2) IN23 ( ) AGE (10-120)
Question 21 >2 (some formal education) and question 23 (literacy) equal to 2 (no literacy) and age is greater or equal to 10 years	IN21 <1,2) IN23 (2) AGE (10-120)
EDUCATION RULES (QUESTIONS 20 AND 21)	
Q20 – SCHOOL *CURRENTLY* ATTENDING Q21 – HIGHEST *COMPLETED* EDUCATION LEVEL	
Question 20 different from 1 (not attending school) and question 21 equal to 1 (no formal education) and age greater or equal to 10 years.	IN20<1) IN21(1) AGE (10-120)
Question 20 different from 2 (Primary school) or 1 (not attending school) and question 21 equal to 2 (no completed education level) and age greater or equal to 10 years.	IN20<1,2) IN21 (2) AGE (10-120)
Question 20 = 2 (Primary school) and question 21 different from 2 (no completed education level) and age is greater or equal to 10 years.	IN20(2) IN21<2) AGE (10-120)
Question 20 = 3 (lower secondary) and question 21 different from 3 or 4 (primary school) and age >= 10	IN20(3) IN21<3-4) AGE (10-120)
Question 20 = 4 (upper secondary) and question 21 different from 5 or 6 (lower secondary) and age >= 10	IN20(4) IN21<5-6) AGE (10-120)
Question 20 = 5 (college as under graduate) and question 21 different from 7 or 8 (upper secondary) and age >= 10	IN20(5) IN21<7-9) AGE (10-120)
Question 20 = 6 (college as post graduate) or 7 (college as doctorate) and question 21 different from 10 (degree (bachelor)or 11 (post-graduate degree) or 12 (doctorate) and age >= 10	IN20(6-7) IN21<10-12) AGE (10-120)
Q20 - SCHOOL *CURRENTLY* ATTENDING - the answers will be based on the group ages	
IN20 = 1 or 2 (Primary grades) and age 6-9 years	AGE(6-9) IN20<1,2)
Age from 10 to 12 years and school attendance different from primary(2) or lower secondary(3) or not attending (1)	AGE(10-12) IN20<1,2,3)
IN20 = 1 or 3 (Lower secondary) and age 13 years	AGE(13) IN20<1,3)
IN20 = 1 or 3 (Lower secondary) or 4 (Upper secondary) and age 14-16 years	AGE(14-16) IN20<1,3,4)
IN20 = 1 or 4 (Upper secondary) and age 17 years	AGE(17) IN20<1,4)

IN20 = 1 or 4 (Upper secondary) or 5 (undergraduate) and age 18-20 years	AGE(18-20) IN20<1,4,5)
IN20 = 1 or 5 (undergraduate) or 6 (post-graduate) or 7 (doctorate) and age 22-120 years	AGE(21-120) IN20<1,5,6,7)
Q21 – HIGHEST COMPLETED EDUCATION LEVEL - the answers will be based on the group ages	
Age from 10 to 12 years and completed education level greater than primary school	AGE(10-12) IN21<1,2,4)
Age from 13 to 16 years and completed education level greater than lower secondary	AGE(13-16) IN21<1,2,4,6)
Age from 17 to 19 years and completed education level greater than upper secondary	AGE(17-19) IN21<1,2,4,6,8)
Age = 20 years and completed education level greater than upper secondary (also old system)	AGE(20) IN21<1-6,8)
Age from 21 to 23 years and completed education level greater than degree (also old system)	AGE(21-23) IN21<1-6,8,9,10)
Incompatibility between attending school different from primary without a completed education level and age greater or equal to 10 years	IN20(3-7) IN21(1-2)
For ages greater than 9 years If the Q21 has for an answer the modality 3 or 4 then Q20 can not have the modality 1 or 3	IN21(3-4) IN20<1,3) AGE (10-120)
For ages greater than 9 years If the Q21 has an answer for modality 5 or 6 then Q20 must have the modality 1 or 4	IN21(5-6) IN20<1,4) AGE (10-120)
For ages greater than 9 years If the Q21 has for an answer the modality 7,8,9,10,11,12 then the Q20 must be modality 1 or 5 or 6 or 7	IN21(7-12) IN20<1,5,6,7) AGE (10-120)
If the Q21 has for an answer the modality 7,8,9,10,11,12 then Q22 must have the education level declared	IN21(7-12) IN22B( )
If the Q21 has an answer different from modalities 7,8,9,10,11,12 then Q22 must be missing	IN21<7-12) IN22B<)
Q24 – WORK FOR PAY OR PROFIT DURING THE WEEK BEFORE THE CENSUS Q25 – UNPAID WORK FOR AT LEAST ONE HOUR DURING THE WEEK BEFORE THE CENSUS	
SKIP RULE: QUESTIONS FROM 24 TO 37 ARE ONLY FOR PERSONS AGED 15 YEARS OR OVER	
Age < 15 years and question 24 - 37 (work) different from space	AGE (0-14) SKIP3 <)
Age >= 15 years and question 24 (work for pay or profit during the week before census) not answered (equal to space)	AGE (15-120) IN24 ( )
SKIP RULE: IF QUESTION 24 = YES GO TO Q28, IF Q24 = NO GO TO Q25	
Question 24 = 1 (work for pay) and question 25 (unpaid work...) different from space	IN24(1) IN25 <)
Question 24 = 1 (work for pay) and question 26 (temporary absence from work) different from space	IN24(1) IN26 <)

Question 24 = 1 (work for pay) and question 27 (reason for not work) different from space	IN24(1) IN27 <)
Question 24 = 2 (no work for pay) and question 25 (unpaid work) missing	IN24(2) IN25 ( )
IF q25 =1 GO TO Q28, IF Q25 = 2 GO TO Q26	
Question 25 (unpaid work) = 1 and question 26 (temporary absence from work) different from spaces	IN25(1) IN26 <)
Question 25 (unpaid work) = 1 and question 27 (reason for not work) different from spaces	IN25(1) IN27 <)
Question 25 (no unpaid work) = 1 and question 26 (temporary absence from work) missing	IN25(2) IN26 ( )
IF q26 =1 GO TO Q28, IF Q26 = 2 GO TO Q27	
question 26 (temporary absence from work) = 1 and question 27 (reason for not work) different from spaces	IN26(1) IN27 <)
question 26 (temporary absence from work) = 1 and question 28 (hours actually worked during the week before the census) different from spaces	IN26(1) IN28 <)
question 26 (no temporary absence from work) = 2 and question 27 (reason for not work) missing	IN26(2) IN27 ( )
QUESTION 28-29 If IN28 is answered, also should have an answer to question 24 or 25	
Question 24 (work for pay) equal to yes and question 28 (number of hours worked the week before census day) without answer	IN24 (1) IN28( )
Question 24 (work for pay) equal to yes and question 29 (number of hours usually worked) without answer	IN24 (1) IN29( )
Question 25 (unpaid work) equal to yes and question 28 (number of hours worked the week before census day) without answer	IN25 (1) IN28( )
Question 25 (unpaid work) equal to yes and question 29 (number of hours usually worked) without answer	IN25 (1) IN29( )
Question 26 (temporary absent from work) equal to yes and question 29 (number of hours usually worked) without answer	IN26 (1) IN29( )
Question 24 (work for pay) different from yes and Question 25 (unpaid work) different from yes and question 28 (number of hours worked the week before census day) not missing	IN24 <1) IN25 <1) IN28<)
ROUTING RULES FOR Q27	
Question 27 need answer only if all questions 24-26 are equal to NO	IN24 <2) IN25 <2) IN26 <2) IN27 <)
Question 27 missing and all questions 24-26 equal to NO	IN24 (2) IN25 (2) IN26 (2) IN27 ( )
QUESTION 27 =1-6 (UNEMPLOYED, STUDENT, PENSION,...) GO TO Q.35	

Question 27 (reason for not work) < 7 and question 28 – 33A different from space	IN27 (1-6) SKIP4 <)
Question 27 (reason for not work) from 1 to 6 and question 35 missing	IN27 (1-6) IN35( )
Question 27 (reason for not work) from 1 to 6 and question 36 missing	IN27 (1-6) IN36( )
Question 27 missing and question 35 different from space	IN27 ( ) IN35<)
Question 27 missing and question 35 different from space	IN27 ( ) IN36 <)
QUESTION 27 =7 (WAITING TO START A JOB) OR 10 (PERMANENTLY UNABLE TO WORK) OR 11 (OTHER) GO TO Q. 37 (PLACE OF WORK OR STUDY)	
question 27 (reason for not work) = 7 or 10 or 11 and question 28 – 36 different from space	IN27 (7,10,11) SKIP5 <)
QUESTION 27 =8 (MILITARY SERVICE) GO TO Q. 41 (MAIN SOURCES OF LIVELIHOOD)	
question 27 (reason for not work) = 8 and question 28 – 40 different from space	IN27 (8) SKIP6 <)
QUESTION 27 =9 (PRISON) GO TO Q. 39	
question 27 (reason for not work) = 9 and question 28 – 38 different from space	IN27 (9) SKIP7 <)
WORK RULES (QUESTIONS 33, 37, 38, 27)	
Q27 – REASON FOR WHICH YOU DIDN'T WORK THE WEEK BEFORE THE CENSUS Q33 – PLACE OF WORK Q37 – PLACE OF WORK/SCHOOL (KOSOVO/ OUTSIDE KOSOVO) Q38 – HOW OFTEN DO YOU TRAVEL BETWEEN YOUR HOME AND WORK/STUDY PLACE	
If Question 28 or 29 is different from zero, Question 30,31,32,33,34 should have answer	
Question 28 (number of hours worked the week before census day) with an answer and questions 30-34 without answer	IN28(1-72) IN30( ) IN31( ) IN33A( ) IN34( )
Question 29 (number of hours usually worked) with an answer and questions 30-34 without answer	IN29(1-72) IN30( ) IN31( ) IN33A( ) IN34( )
Question 33 = 2 (work at home) and question 37 different from 1 (work not in Kosovo)	IN33A(2) IN37A<0,1)
Question 33 = 2 (work at home) and question 38 different from 4 (not applicable)	IN33A(2) IN38<0,4)
Question 33 not equal to 2 (work at home) and question 38 equal to 4 (not applicable)	IN33A<2) IN38(4)
Question 27 = 10 (permanently unable to work) and question 39 equal to 7 (no difficulties)	IN27(10) IN39A(7)
ROUTING RULES: Question 37-38 are for people aged 10 years and over	
Question 37 is for people aged 10 years and over	AGE (0-9) IN37A <)
Question 38 is for people aged 10 years and over	AGE (0-9) IN38 <)

Question 37 must be skipped if question 27 is equal to 8 (military service) or 9 (prison)	AGE (10-120) IN37A( ) IN27<8,9)
Question 38 must be skipped if question 27 is equal to 8 (military service) or 9 (prison)	AGE (10-120) IN38 ( ) IN27<8,9) IN37A (1,3)
Question 37 is mandatory for people of 10 years and over currently attending school (Q20)	IN20(2-7) IN37A<4) AGE(10-120)
Question 27 = 3 (attending school) and question 20 equal to 1 (not attending school)	IN20(1) IN27(3)
Q37 – PLACE OF WORK OR SCHOOL	
QUESTION 37 = 4 (NOT APPLICABLE) GO TO Q39	
question 37 (place of work/school) = 4 and question 38 different from space	IN37A(4) IN38<)
Settlement an municipality specified and Q37 different from 2 (in Kosovo other municipality/settlement)	IN37A (1,3,4) IN37C<)
Settlement an municipality specified and Q37 different from 2 (in Kosovo other municipality/settlement)	IN37A (1,3,4) IN37B<)
Country specified and Q37 different from 2 (outside Kosovo)	IN37A (1,2,4) IN37D<)
Settlement an municipality missing and Q37 equal to 2 (in Kosovo other municipality/settlement)	IN37A (2) IN37C( )
Settlement an municipality missing and Q37 equal to 2 (in Kosovo other municipality/settlement)	IN37A (2) IN37B( )
Country missing and Q37 equal to 3 (outside Kosovo)	IN37A (3) IN37D( )
Q39 – LONG LASTING DIFFICULTIES Q41 – MAIN SOURCES OF LIVELIHOOD	
QUESTION 39 = 7 (NO DIFFICULTIES) GO TO Q41	
question 39 = 7 (no difficulties) and question 40A (learning,...) different from space	IN39A(7) IN40A<)
question 39 = 7 (no difficulties) and question 40B (dressing,...) different from space	IN39A(7) IN40B<)
question 39 = 7 (no difficulties) and question 40C (going outside alone) different from space	IN39A(7) IN40C<)
question 39 = 7 (no difficulties) and question 40D (work or attending school) different from space	IN39A(7) IN40D<)
question 39 = 7 (no difficulties) and question 40E (other activities) different from space	IN39A(7) IN40E<)
	IN39A(7) IN39B<)
	IN39A(7) IN39C<)
QUESTION 39A, Q39B, Q39C must be different	
	IN39A(1) IN39B(1)
	IN39A(2) IN39B(2)

	IN39A(3) IN39B(3)
	IN39A(4) IN39B(4)
	IN39A(5) IN39B(5)
	IN39A(6) IN39B(6)
	IN39A(1) IN39C(1)
	IN39A(2) IN39C(2)
	IN39A(3) IN39C(3)
	IN39A(4) IN39C(4)
	IN39A(5) IN39C(5)
	IN39A(6) IN39C(6)
	IN39B(1) IN39C(1)
	IN39B(2) IN39C(2)
	IN39B(3) IN39C(3)
	IN39B(4) IN39C(4)
	IN39B(5) IN39C(5)
	IN39B(6) IN39C(6)
Q41 - SOURCE OF LIVELIHOOD AND AGE	
Question 41 equal to 1 (first source of livelihood equal to work) and age <= 14	AGE(0-14) IN41A(1)
Question 41 equal to 1 (second source of livelihood equal to work) and age <= 14	AGE(0-14) IN41B(1)
Q41 - FIRST SOURCE OF LIVELIHOOD HAS TO BE DIFFERENT FROM SECOND SOURCE OF LIVELIHOOD	
	IN41A(1) IN41B(1)
	IN41A(2) IN41B(2)
	IN41A(3) IN41B(3)
	IN41A(4) IN41B(4)
	IN41A(5) IN41B(5)
	IN41A(6) IN41B(6)
	IN41A(7) IN41B(7)
	IN41A(8) IN41B(8)
Q24, Q25,26 AND AGE > 80 - IT IS IMPOSSIBLE TO WORK AFTER 80 YEARS	
	IN24(1) AGE(80-120)
	IN25(1) AGE(80-120)
	IN26(1) AGE(80-120)
	IN27(1,2,3,7,8) AGE(80-120)
	IN35(1) AGE(80-120)
	IN36(1) AGE(80-120)

Final treatment

The goal of this step was to complete the data cleaning for some variables and prepare the final file.

Question 3 - Date of birth
Day of birth: random choice of a number between 1 and 28
Month of birth: random choice of a number between 1 and 12
Year of birth: Compute as 2011-AGE. If month of birth > 4 then year of birth = year of birth – 1.

Questions 6, 7, 8 - Fertility question
Respect of skip rule for males.

Question 18 – When did you recently move to your current place of residence.
Month: random choice of a number between 1 and 12.

Annex 3 presents the results of editing and imputation of R3 questionnaires, by confronting the initial (or raw) data with the ones obtained by the cleaning process.

CHAPTER 4

Assessment of data quality based on demographic techniques

The purpose of this chapter is to assess the quality of 2011 census of Kosovo using demographic methods and comparisons with other relevant sources; a separate report on census coverage and content quality based on the post-enumeration survey is also available.

The present analysis includes evaluation of data on age-sex distribution, education characteristics, marital status, children ever born, economic characteristics and housing characteristics. Depending on the amount of demographic information available for Kosovo, demographic methods for this study are selected to assess the quality of data. It should be noted that, some demographic methods particularly used for cohort analysis to measure coverage errors are not included in this report because of the absence of periodical censuses for 30 years in Kosovo. Therefore this report mainly focuses on evaluation of content errors rather than coverage errors. It is expected that this report would provide further insight into the types of errors in the data in addition to the results of the Post Enumeration Survey.

In the following, evaluation of the quality of census data are carried out by age groups, sex and rural/urban areas for all census topics included in the analysis to detect eventual inconsistencies of data for these groups of people or type of areas. Using age, sex and location breakdown in evaluation of data may provide better insight in creating a typology of reasons for errors found in the data.

4.1 Data Sources Available for Census Evaluation

The Population and Housing Census was conducted in April 2011, with reference moment of 31 March 2011 at midnight. For the purpose of the evaluation of the 2011 census, the following data sources will be used in this study considering the data requirement of the methods and coherence between the reference date of additional data sources and the census:

4.1.1 Population and Housing Censuses

Population and housing censuses were not conducted periodically in Kosovo. The previous census was carried out 30 years ago with different definition of population and considerable changes in the border of administrative geographical units. Therefore, evaluation of the quality of age-sex distribution of population enumerated in 2011 census will be carried out using the demographic methods which require only single (current) census results.

4.1.2 Birth and death registers

The Kosovo Agency of Statistics (ASK) collects data on births and deaths statistics and publishes them annually. Data on births include all births registered in a given year and concern new-born babies delivered by all Kosovo nationals whether they live in Kosovo or in other countries. Fortunately, additional information for the year of birth (not of registration of this birth) and a distinction of registration of births occurred in Kosovo and in other countries are available in the registers. Regarding death statistics, it seems that data quality is not as good as birth statistics. Particularly deaths occurred at home may not be registered properly in the system. Comparison of infant deaths derived from vital registration system and estimated from Demographic and Health Survey implies that death registers might be reliable with reasonable coverage rate. However "Death Statistics" publications do not provide child mortality rate, therefore infant and child mortality estimation from the "Demographic and Health Survey" is used in this report to estimate population size of children. It should be noted that all administrative sources do not cover data for the Serbian community; therefore special attention is requested in order to ensure that the coverage of both census and administrative registers would not affect the results of analysis undertaken for measuring the quality of census data.

4.1.3 School enrolment

ASK collects data on education statistics including data on pupils and schools and publishes annually. The data for this publication is collected by sending the standard questionnaire to all types of school including both private and public schools. In this report, school enrolment data is used to analyze the quality of school attendance by the level of education.

4.1.4 Households Surveys

Three households surveys- Demographic and Health Surveys, Labour Force Survey and Household Budget Survey- can be used to evaluate the quality of census data. Regarding population characteristics, Demographic and Health Survey and Labour Force Survey conducted in 2009 are main sources to evaluate the quality of data on population characteristics. Additionally, Household Budget Survey conducted in 2010 is used to evaluate the quality of data on housing characteristics. In fact, analysis of HBS showed that the types of questions and categories of these questions on housing characteristics used in the census and the survey are not harmonized properly, but some data like number of rooms and the facilities in housing units (such as availability of bath, flush toilet, central heating system) are comparable with the census data. Therefore, a limited number of characteristics of housing units is covered in the report.

4.2 Methodological Framework

There are fairly large numbers of methods which can be applied in census evaluation. These methods can be selected according to purposes of census evaluation and available data sources in the

country. In general, these methods differs from very simple methodology which requires the results of household surveys and/or administrative registers for analyzing coherence between the census and other sources to somehow more sophisticated methodologies using demographic techniques which require a series of census results.
According to the availability of data sources in Kosovo, the methods used in this study are categorized into three groups;

4.2.1 Demographic methods

Demographic methods will be extensively used to assess the quality of age-sex distribution of population using population pyramid, age ratio, sex ratio and Whipple index. Also, child population will be estimated based on birth registers and child mortality rate to evaluate the coverage of enumerated child population.

4.2.2 Comparison of the census with administrative registers

One of the more conceptually straightforward approaches for assessing the plausibility of a census data is the comparison of the census data with a corresponding data from independent data source. This study will use birth registers and school enrolment registers as independent data sources.

4.2.3 Comparison of the census with existing household surveys

Household surveys conducted within a few years of the census will be used to assess the plausibility of the enumerated distribution of population by selected characteristics. The rationale for making such comparisons is that because of the greater degree of operational control that can normally be imposed over a survey activity in comparison with a census, survey data are often less seriously affected by non-sampling errors. However, survey data are subject to sampling errors, which in many cases limits the strength of the conclusions that can be derived from comparisons of the survey and census data. Nevertheless, serious discrepancies between census and survey distributions for selected characteristics are indicative of error in one or both sets of data.

The main limitation of using demographic techniques to evaluate the census of Kosovo is the absence of consecutive censuses. If two or more consecutive censuses were available, a wider range of demographic methods would be applied and defensible estimates of the magnitude of census coverage and content error could be derived based on these techniques. The availability of results from one or more demographic surveys conducted during inter-censual period would further increase the usefulness of demographic analysis. Because of this limitation, it is not possible to measure the coverage error of the enumerated population. Yet this coverage estimates are available through the PES analysis.

4.3 Evaluation of 2011 Census Data

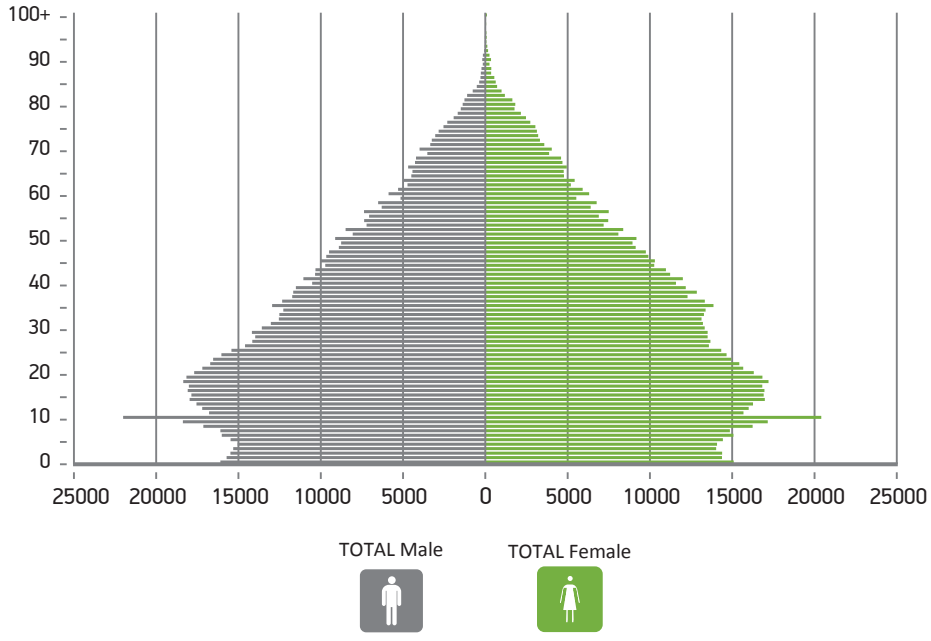
4.3.1 Age-sex distributions

Evaluation of the "reasonableness" of the distribution of the population enumerated in the census by age and sex can provide considerable insight into the quality of the census enumeration. The reason for this is that the age and sex distribution of a population, which is determined by the levels of fertility, mortality and migration flows to which the population has been exposed, follows a fairly predictable pattern in the absence of error in the data. Accordingly, significant discrepancies in a census age-sex distribution, which cannot be accounted by extraordinary events; other demographic events such as age or sex selective migration; or structural changes in fertility/mortality, are usually indicative of census errors.

Population pyramids

Figure 1 displays the size of the population enumerated in each age (or cohort) by sex and urban/rural. The population distribution for ages 20 and over illustrates the typical shape of a population pyramid. The narrower base of the pyramid for ages 20 and younger reflects the rapid decline in fertility. Both urban and rural population pyramids have very similar shape to the pyramid of the total population.

Figure 1. Population pyramid by single age and sex



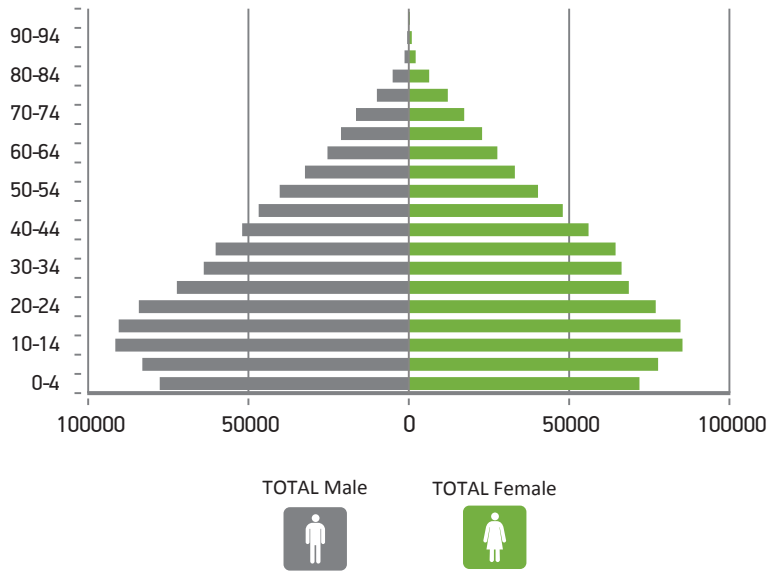
The population pyramid indicates irregular pattern at age 10 for both sexes and localities. The size of population at age 10 is 30 percent higher than the population at age 11 and 20 percent higher than the population at age 9 for males and females. Looking at the shape of the pyramid, there is strong indication for a baby-boom in the year 2000 immediately after the 1999 conflict in the country. However, there is no other source to support this indication considering that birth statistics are not available before the years 2002 and DHS-2003 does not provide total fertility rate for single years including the year 2000.

According to population distribution by single age, there is no clear sign for age misreporting or heaping for both sexes living in rural and urban areas.

Figure 2 presents population distribution by five-year age group and sex for total population (see the annex 4 for the population pyramids for urban and rural areas). In general, the pyramid reflects the typical shape of a population pyramid with gradually decreasing population size with increasing age because of mortality. This general shape of the pyramid is distorted by young generations. Narrowing shape at the bottom of the pyramid clearly shows the rapid decline in fertility particularly during the last 10 years.

It can be observed from the pyramid that there is no significant indication for age or sex selective international migration or structural changes in age distribution because of migration or extraordinary events. This assessment might also be considered as good sign for the possibility of enumerating people living abroad as residents in the census. Otherwise, some cohorts could have irregular sizes of population particularly in working age groups.

Figure 2. Population pyramid by five-year age group and sex



Whipple Index

The Whipple index measures the magnitude of a particular type of error in census age data with the extent of heaping on ages ending in the digits of 0 and 5. The index ranges from 100 to 500 where 100 representing no preference for digits "0" or "5" and 500 indicating that only digits "0" and "5" were reported in the census.

Table 1 shows the results of the index by sex and localities. It is clear from the figures that there is no age preference or misreporting in the census data. The results of Whipple Index imply a perfect age reporting in the census for both sexes. It should be highlighted that as many countries experienced similar success, the quality of age data is increased by collecting this information with date of birth (year, month and day) instead of asking the completed years of age.

Table 1: Whipple Index by sex and location

Location	Male	Female
Total	101.53	101.17
Urban	101.66	101.13
Rural	101.45	101.20

Age and Sex Ratio

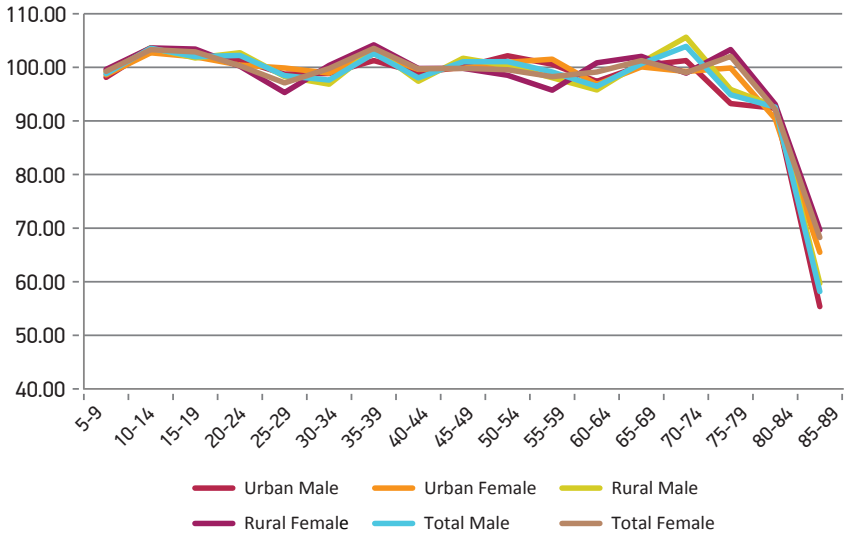
The assessment of the "reasonableness" of census age-sex distribution may be approached in a more quantitative fashion through the examination of age and sex ratios in the enumerated census population. The diagnostic value of these ratios lies in their predictable patterns in human populations. Accordingly, significant departures from ratio values in the absence of measurement error which cannot be explained by extraordinary or distorting factors indicate census error (US, 1985. P. 133).

Age Ratios

Age ratios provide a measure of the "smoothness" of the age distribution of the population over a restricted age range. In the absence of sharp swings in fertility or mortality, significant level of migration, or other distorting factors, the enumerated size of a particular cohort should be approximately equal to the average size of the immediately preceding and subsequent cohorts. In other words, the ratio of the census count for a particular cohort to the average of the counts for the adjacent cohorts should be approximately equal to 100. Significant departures from this expected ratio indicate either the presence of error in the census enumeration or of other distorting factors (US, 1985. P. 133).

Figure 3 presents age ratios by sex and localities. The ratios range between 104 and 95 for all ages except two very old age groups because of very high level of mortality, age ratios decreases with increasing of age. The pattern of age ratios in general reflects smooth age distribution of population for both sexes and localities. However, there are some departures from this pattern particularly for females living in rural areas. The figure shows that females in rural areas at ages 25-29 and 55-59 (age ratios are 95.3% and 95.7% respectively) have lower population size compared to the adjacent cohorts. This distortion cannot be explained by lower fertility level for these cohorts since only female population living in rural areas are affected. This finding might indicate that the females in age groups 25-29 and 55-59 were exposed to migration more than males (particularly from rural to urban and/or international migration) compared to the adjacent cohorts. Further analysis of migration data collected in the census might provide better information to understand if there is different pattern of migration between males and females. Except these slight distortions, the values of age ratios indicate highly consistent population size by age groups by rural/urban areas and sex.

Figure 3. Age ratio by sex and localities

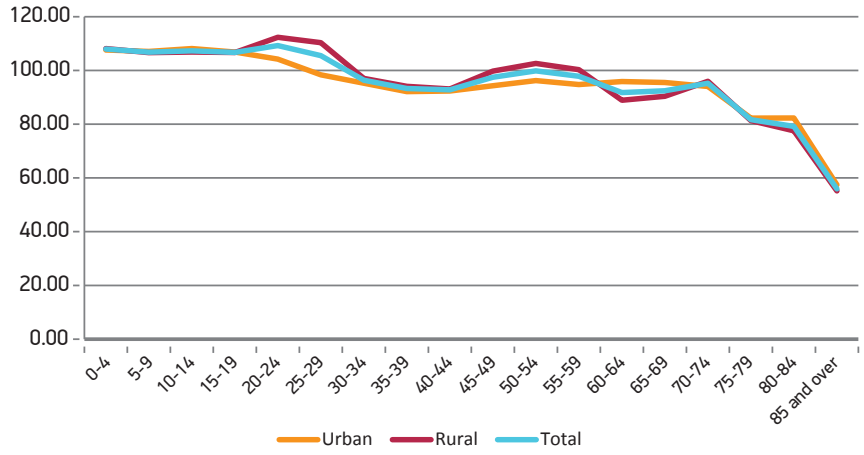


Sex Ratios

Sex ratios are defined as the ratio of the enumerated male population to the enumerated female population in a particular age group. The sex ratio at birth for most populations lies between 102 and 107 and due to higher mortality rates for males in most populations, the ratio declines gradually among older cohorts of the populations. Accordingly, significant variations from this pattern are implications of either census errors or of the presence of distorting factors such as sex-selective migration or mortality (US, 1985. P. 133).

Sex ratios by age group and localities are given in Figure 4. Although the general pattern of sex ratios reflects the expected pattern of declining sex ratio with increasing age, the patterns of rural and urban areas are somehow different from each other indicating the possibility of sex-based selective migration. The distortions in the sex ratios do not compensate between rural and urban areas, therefore this might be indication of sex and age selective immigration in the country. It might be concluded that the high sex ratios for 20-24 and 25-29 age groups (112 and 110 respectively) in rural areas and relatively fast decline in sex ratios for ages from 30 to 44 indicates the possibility that more female migration occurred in young ages and more male migration occurred in older ages. For further and more precise investigation, migration data should be analyzed to evaluate how the pattern of migration differs between sexes and locations.

Figure 4. Sex Ratio by age group



Estimation of Child Population- Age group (0-4)

A more commonly available source of information for census evaluation purposes is a vital registration system. A considerable number of census evaluations have been undertaken in which the registered number of births in the months/years immediately preceding the census – adjusted for infant and child mortality - have been used as a basis for assessing census coverage of the population for the youngest cohort.

Number of births registered in vital registration system and child mortality rate estimated from DHS-2009 are used as independent sources to estimate child population (age group 0-4). ASK publishes birth statistics for Kosovo nationals living in the country and abroad. For this estimation, number of births for mothers living in the country (given in Table 2) is separated from total births. The number of births occurred in specific year is calculated using information on number of registered

births by the year of birth. Annual registered births include the births occurred in the same year of registration and occurred one year before the registration. Table 2 provides the total number of births by the years from 2002 to 2010 taking into consideration their year of registration. However, the births provided in the birth statistics for the year 2010 are incomplete since the births registered one year later will be included in the statistics in the following year 2011. In order to adjust the number of births for 2010, it is assumed that five percent of the births would be registered in the following year considering current experiences in Kosovo. Using this ratio, total births are estimated as 28,106 for the year 2000. With this method, the total number of births for the period 2006 and 2010 are calculated by adding number of births occurred in each year.

Table 2. Population estimates for age group (0-4) based on birth registers

Year of birth	Year of registration		Proportion of births registered in the year following the year of the birth (%)	Total births
	Number of births registered in the same year as birth occurred	Number of births registered in the year following the year of the birth		
2002	31079	3516	11.3	34595
2003	28101	4855	17.3	32956
2004	27680	4285	15.5	31965
2005	29387	1640	5.6	31027
2006	28744	1526	5.3	30270
2007	27405	1229	4.5	28634
2008	27577	1325	4.8	28902
2009	27118	834	3.1	27952
2010(1)	26768	1338.4	5	28106
2006-2010 (2)				143864
			Number of children deaths	
			(Child mortality rate: 11.2‰	
			from DHS-2009)	*1611*
			Estimated population	
			at age group (0-4)	*142253*
			Enumerated population	*149714*
			at age group (0-4)	

Notes: (1) Number of births for the year 2010 is estimated based on assumption of 5 percent of births registered one year after the birth. (2) Total births for five-year period is used to estimate the youngest population (0-4 age group).

Using total number of births for 2006-2010 period (estimated as 143,864) and child mortality rate from DHS-2009 (11.2‰) the child population is estimated as 142,253. According to this estimation, it seems that there is a 5 percent over-count for child population in the census. However, the reliability of this estimation depends on the quality of birth registers and of the child mortality rate. There is strong expectation for under-coverage of births instead of over-coverage because the births which are registered two years or more after the births occurred are not included in birth statistics. There is probably under-coverage of births in vital statistics, which certainly will create a bias in estimating child population.
The difference between the number of births from administrative data (recorded in the civil registry) and the population census 2011 is a consequence of not recorded births of the Serbian community (recorded in the civil register) whereas they participated in the population census.

Also the data on infants from the vital statistics (2006-2010) and the population census 2011 are almost fully in line. According to the vital statistics (civil register) during the period 2006-2010 the annual number of infant mortality deaths was 335, whereas the data from the census2011 shows that the annual number of infant mortality deaths for the same period (2006-2010) was 329. This small difference is a consequence that Serbian community is not registering their death cases in the Kosovo's civil register (i.e. infant death cases) while they have participated in the population census.

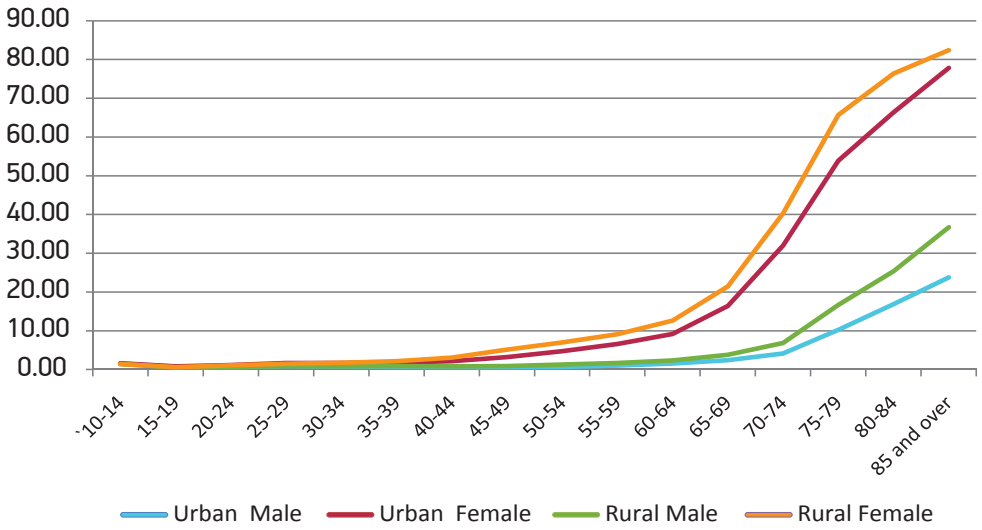
4.3.2 Education Characteristics

The possible sources of data for assessing the quality of census results for education characteristics are the results of household surveys -regarding illiteracy, school attendance and the completed level of education- and administrative registers on school enrolment, both data published by ASK. The DHS-2009 publication provides very detailed data on education characteristics of population; therefore the DHS is used to assess the quality of census data in this report. This analysis can eventually be extended including the results of LFS-2010.

Illiteracy

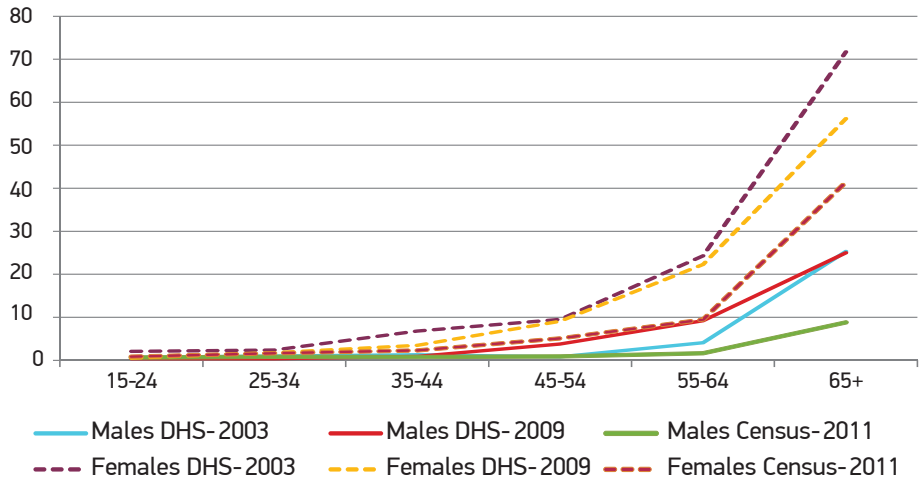
Proportion of illiterate population by sex, age group and localities from the census results is given in Figure 5. The figure indicate very low illiteracy for females until age 50s and for males until age 70s. This proportion is dramatically increased for females with increasing age particularly after age 60s. In general the pattern of illiterate population is reasonable with higher rate among females and for population living in rural areas. In this regard, data on illiteracy seems very good quality because its pattern is highly compliance with general expectation.

Figure 5. Proportion of illiterate population by sex, age group and localities



Census data on illiteracy is compared with the results of DHS-2003 and 2009. Proportion of illiterate population by age group and sex is given in Figure 6. It can be observed from the results of DHS, illiterate population is decreasing over time in all ages. Same trend in proportion of illiterate population is experienced in the census for males and females. The figure indicates that census results are consistent with the results of DHSs in all ages.

Figure 6. Proportion of illiterate population and age groups and sex from the census and DHS-2003-2009



School Attendance

There are two data sources to analyze the results of the census on school attendance. Education statistics yearly published by ASK are collected in collaboration with schools as aggregated data by level of education and sex. The table 3 provides the number of students from the census and school enrolment by sex and level of education. The basic assumption for this comparison is that school enrolment data as independent source is considered as "expected" counts to assess the quality of census data.

The table 3 shows that the number of students for primary and secondary school enumerated in the census is around 2.8 percent superior to the number of students registered in the schools for both sexes. Differences in number of students between the census and the school enrolment for upper secondary school present similar results with 4.5 percent for males and 3.4 percent for females. However, the differences between two data sources are less than five percent for two levels of education indicating good coherence between the census and administrative data.

There might be several reasons for this difference, for example the reference date for the two sources may create this difference and/or the quality of data derived from school enrolment might not be good quality since it is collected as aggregated data from the schools rather than individual administrative registers.

Table 3. Population currently attending school , Census-2011 and school enrolment-2010-2011

Level of Education		Census-2011	School enrolment 2010/2011	Difference (%) (Census-Enrolment)/ Enrolment
Primary and lower secondary	Male	161110	156578	2.89
	Female	149757	145675	2.80
	Total	310867	302253	2.85
Upper secondary	Male	61348	58715	4.48
	Female	51473	49788	3.38
	Total	112821	108503	3.98

As second source, the results DHS-2009 are used to evaluate the quality of census data on school attendance. Table 4 provides proportion of population who are currently attending school by selected age group from the census and DHS-2009. Differences between two sources are very low for both sexes and rural/urban ranging from -1.5 to 4.5 by age groups. It can be observed that among localities, the difference in rural population (particularly for male population) is higher compared to urban population. This might be indication of the higher risk of response errors among people living in rural areas compared to people living in urban areas because of big household size in which two-three generations live together and information for most of household members are usually collected from proxy respondent.

Table 4. Proportion of population currently attending school by sex and age group

	Urban			Rural			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
CENSUS-2011									
`10-14	96.8	96.6	96.7	97.0	96.6	96.8	96.9	96.6	96.8
15-19	85.4	85.1	85.3	80.2	70.9	75.7	82.1	76.0	79.1
20-24	37.7	39.9	38.8	25.4	23.0	24.3	29.8	29.5	29.7
25-29	16.7	13.7	15.2	9.7	5.2	7.6	12.3	8.7	10.5
30-34	8.6	6.4	7.5	3.8	1.5	2.7	5.6	3.5	4.5
DHS-2009									
`10-14	97.3	98.0	97.7	96.5	97.5	97.0	96.8	97.7	97.2
15-19	84.5	82.4	83.4	75.9	70.7	73.3	79.0	74.6	76.8
20-24	39.0	37.0	38.1	22.7	18.7	20.8	28.8	25.4	27.2
25-29	16.3	11.9	14.1	8.4	4.5	6.5	11.1	7.3	9.2
30-34	8.2	4.5	6.3	3.7	1.3	2.4	5.4	2.5	3.9
Difference									
Census-DHS									
`10-14	-0.5	-1.4	-1.0	0.5	-0.9	-0.2	0.1	-1.1	-0.4
15-19	0.9	2.7	1.9	4.3	0.2	2.4	3.1	1.4	2.3
20-24	-1.3	2.9	0.7	2.7	4.3	3.5	1.0	4.1	2.5
25-29	0.4	1.8	1.1	1.3	0.7	1.1	1.2	1.4	1.3
30-34	0.4	1.9	1.2	0.1	0.2	0.3	0.2	1.0	0.6

Tables 3 and 4 indicate that census data on school attendance on the basis of detail cross table by age group, sex and localities is highly consistent with data from school enrolment and DHS-2009.

Completed Level of Education

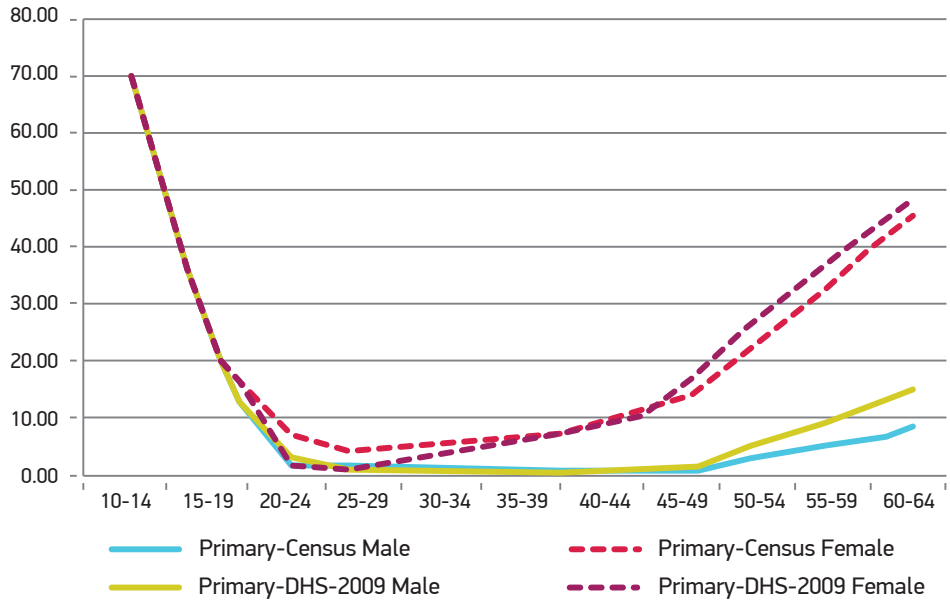
Data on educational attainment in the census was collected for the level of old and new systems of education. On the other hand, DHS-2009 uses 4 categories of educational attainment. In order to compare two sources, the categories used in the census are re-grouped to provide harmonised data with the DHS-2009 as follows:

No schooling : No schooling +No completed level of education
Primary: Primary (4 years)+Primary (5 years)
Lower secondary: Lower secondary (8 grades)+Lower secondary (9 grades)
Higher secondary: Upper secondary (12 grades) + upper secondary (13 grades)
University : Post-secondary vocational + bachelor + postgraduate degree + doctorate

Based on the proposed categories, population by the completed level of education for the census is derived from the original table. In the current report, data quality on educational attainment is analysed using two levels of education-primary and university- as these levels are available in the DHS-2009 report.

The figure 7 presents proportion of population with primary education from the census and the survey. The figure illustrates a clear distinction between males and females indicating that more female population stay with primary education in their life time compared to the male population. The census data is very compatible with the results of the DHS-2009.

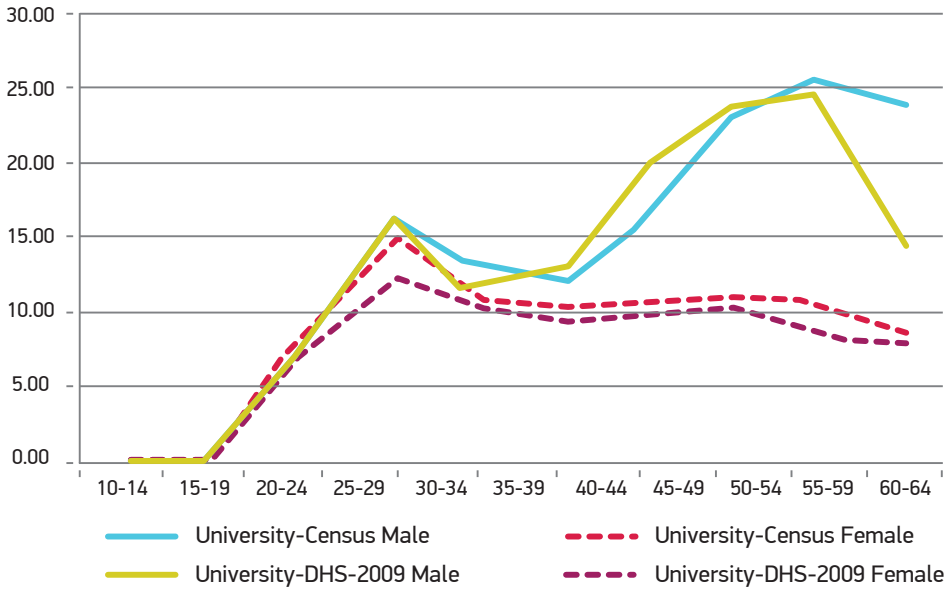
Figure 7. Proportion of population with primary education by age group and sex



Proportion of population with university education is given in the Figure 8. The figure shows unexpected pattern for older males (age 45 and older) who were more likely to have a university education compared to younger cohorts of males. It is clearly seen from the figure this is real situation in the country since both the census and the survey have same pattern for male population. Female population having university education represents somehow smooth pattern with decreasing proportion of university degree for older cohort. The youngest cohort for university degree (age group 25-29) has relatively more population with university education for both sexes that might be indication of recent increase in population having university education.

This analysis clearly indicates the high quality of census data on educational attainment with very coherent results with the results of DHS.

Figure 8. Proportion of population with university education by age group and sex



4.3.3 Marital Status

The distribution of male residents who are single and who are currently married for the ages 15 to 74 from the census and DHS-2009 are presented in the Figures 9-a and 9-b. These distributions are given by urban and rural areas. The figures clearly show that the proportions of single males from the census and the survey are highly consistent with DHS results for urban and rural.

It is interesting that the pattern of single male population and currently married population are almost identical for rural and urban areas. This indicates that male population living in urban and rural areas have similar mean age at first marriage. The distribution of currently married population shows that almost all males are married by age 40 and stay married until very old ages.

Figure 9-a. Proportion of single male population by age group, Census-2011 and DHS-2009

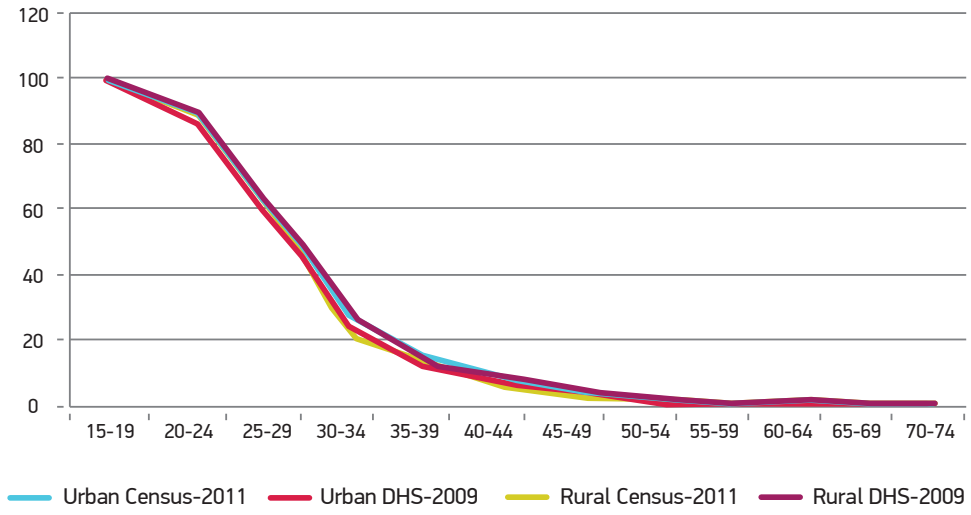
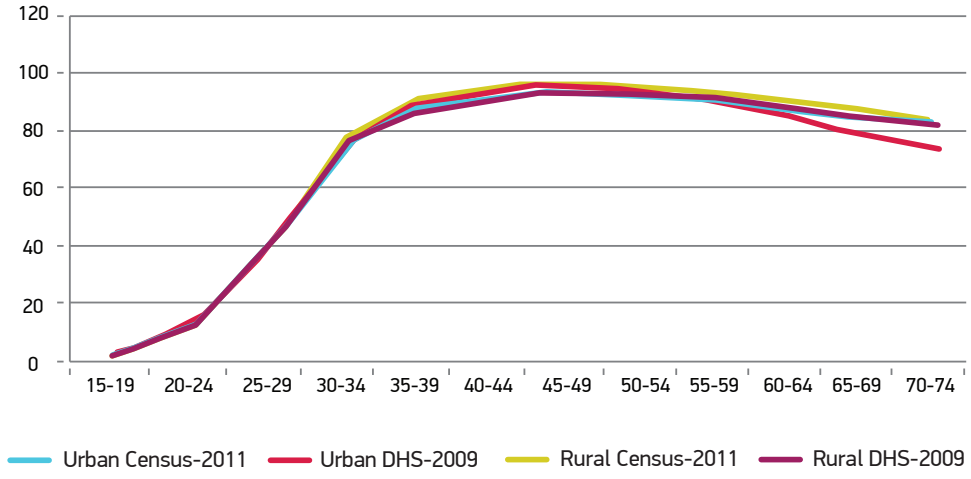


Figure 9-b. Proportion of currently married male population by age group, Census-2011 and DHS-2009



The distribution of female population aged 15 to 74 who are single and currently married is presented in Figures 10-a and 10-b. The proportion of single and currently married females from the census is consistent with the distribution of females from the survey but it is observed that the coherence between two sources are lower compared to the coherence between two sources observed for male population.

The proportion of single female population shows that mean age at first marriage for females living in rural areas is lower than the one of females living in urban areas.

Figure 10-a. Proportion of single female population by age group, Census-2011 and DHS-2009

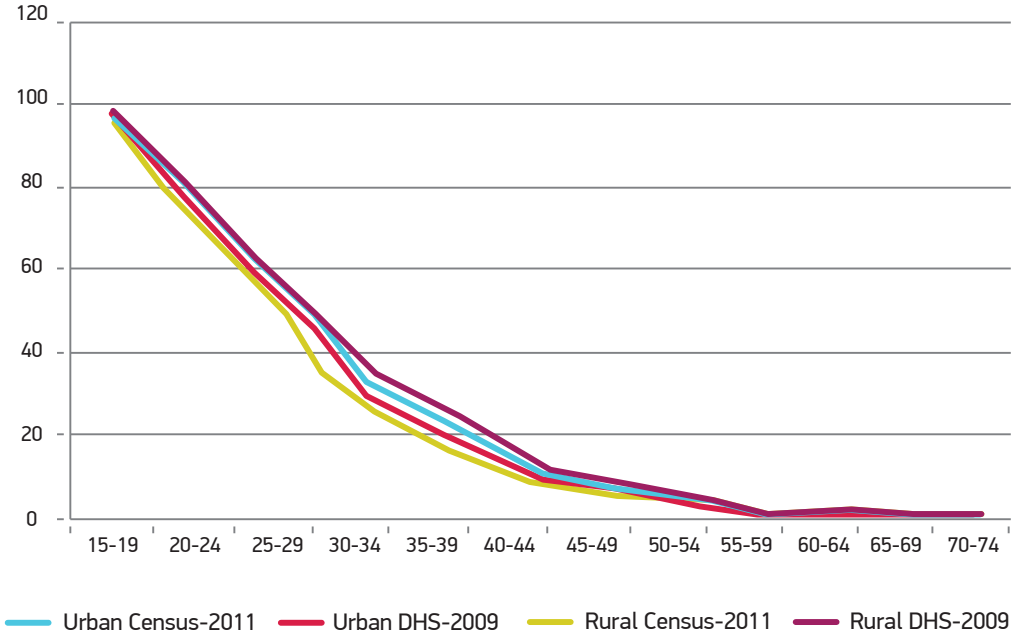
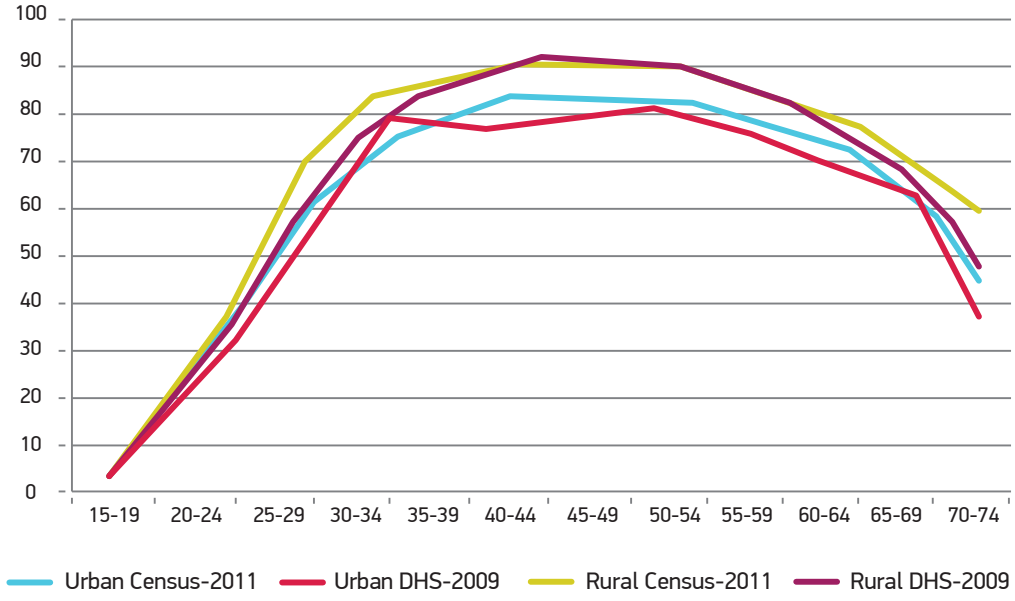


Figure 10-b. Proportion of currently married female population by age group, Census-2011 and DHS-2009



The proportion of married women has a peak at age 40 with around 90 percent of women in this age being married, and then the proportion is decreasing with increasing age. This pattern seems reasonable compared to the experience of many countries. The census and survey have a similar pattern for married females. However, there are slight differences between census and survey particularly in the middle ages. The results of the survey for urban areas illustrate some fluctuations by age groups in the proportion of married females after age 30. This indicates the effect of sampling errors in the results of that survey. Therefore, it can be concluded that the pattern of currently married females observed in the census is consistent with the results of the survey and it can be accepted as more reliable compared the results of the survey.

4.3.4 Children Ever Born

The census questionnaire includes data on number of children ever born during the life time of women. In the census, data required for measuring current level of fertility is not collected as such therefore the quality of census data for fertility is only undertaken for data on children ever born.

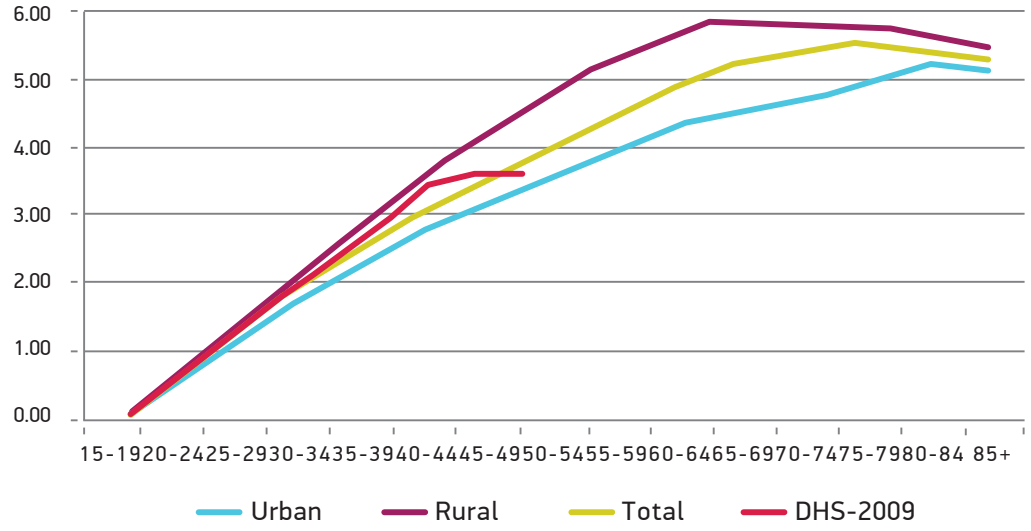
In order to measure the quality of census data on children ever born, the mean (or average) of children ever born (number of children ever born per woman) by age group and urban/rural are calculated from the census and is compared with the results of DHS-2009. The DHS results provide this information only for reproductive ages (ages 15 to 49), while the census provides this data for all ages. Therefore the comparison of the census data with the DHS-2009 is given for women in reproductive ages.

Figure 11 presents average children ever born (MCEB) by age group and rural/urban distribution. As countries with decreasing fertility level, the MCEB per woman is gradually increasing with increasing ages in rural and urban areas. The differences in MCEB between urban and rural areas are getting larger with increasing of age until very old ages.

The figure shows that the levels of MCEB calculated from the census are almost identical to the levels estimated from DHS-2009 for ages from 15 to 45. The MCEB for age group 45-49 is 3.48 and 3.31 children per woman for the census and survey respectively. Although there is slight difference between two sources, DHS-2009 reflects no changes in MCEB in this age group while the census reflects increasing level of MCEB.

The publication of DHS-2009 provides this indicator only for total population; therefore it is not possible to compare the levels of MCEB by rural and urban areas.

Figure 11. Mean children ever born by age group of women and locations



4.3.5 Economic Characteristics

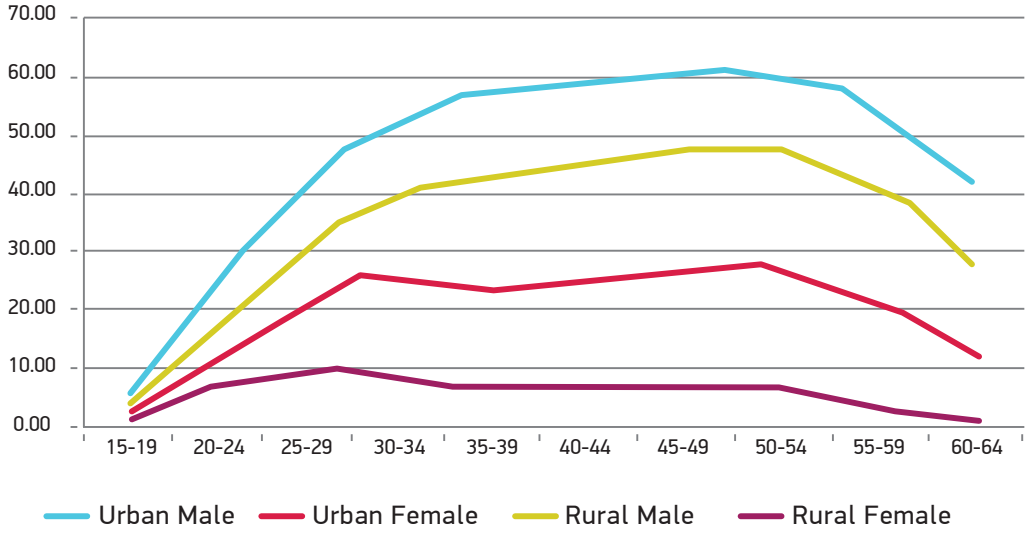
Labour Force Survey (LFS) conducted in 2009 is used to assess the quality of census data on economic characteristics. There is no available data from the social security system, therefore the LFS is the only source to evaluate the census results.

Definitions of economic activity, employment and unemployment used in the census are consistent with the definitions used in the survey according to the information given in the LFS report. In the current report, the quality of census data on economic activity will be assessed using the indicators on employed and unemployed population.

Employed population

Figure 12 presents the proportion of employed population within the working age population compiled from the census by age groups and localities. The distribution of employed population by age groups indicates that there is big difference in the level of employment by urban/rural and sexes. As it is expected the male population has higher proportion of employment in both rural and urban areas compared to female population living in these areas. Moreover, it might be remarkable to note that population living in urban areas have higher proportion of employment in all ages. The LFS-2009 results are not provided by the locations, therefore it is not possible to conclude whether this is real situation in the country or whether there is reporting error in the census because of problems of understanding the definition of economic activity particularly for agricultural activities and mostly for women. This conclusion needs further investigation to clarify the level of employment in rural and urban areas.

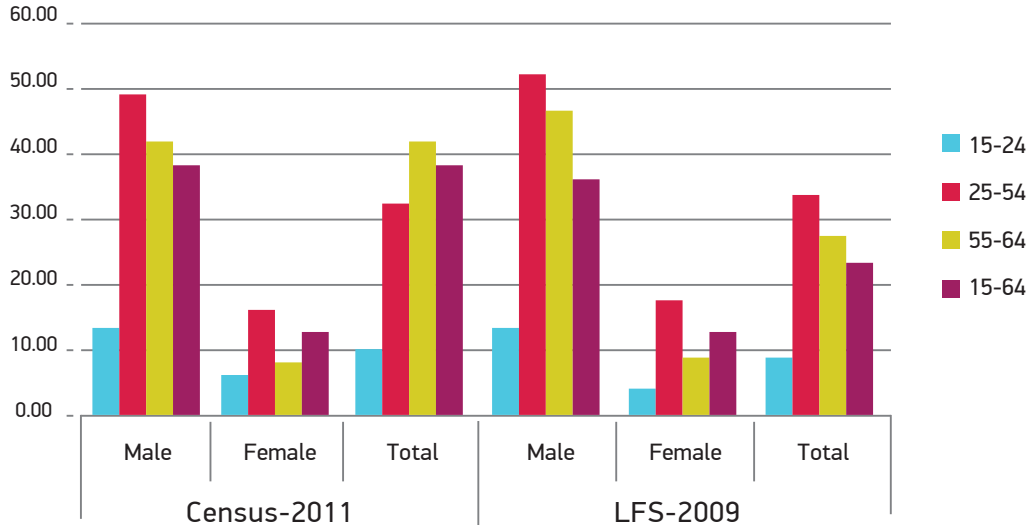
Figure 12. Proportion of employed population in total population by age group



Proportion of employed population by the selected age groups is provided in the Figure 13 for the census and the survey. LFS report provides the proportion of employed and unemployed population only for broad age groups therefore comparison of the census with the survey is limited with these age groups.

The figure indicates that the level of employment for the selected age groups from the census is highly consistent with the results of the survey for both sexes. The proportions of employment population in the survey are slightly higher than the census particularly for males in age group 55-64.

Figure 13. Proportion of employed population in total population by selected age group



Unemployment

Unemployment rate is the percentage of unemployed population in economic active population (labour force). The figure 14 presents unemployment rates by five-year age groups and urban/rural from the census. The figure illustrates interesting pattern of unemployment rates as it is decreasing with increasing age. Schooling ratio for teenagers (age group 15-19) is around 80 percent for both sexes (see Table 4) and unemployment rate for this cohort is over 80 percent reflecting the fact that most of young people are looking a job while they are students. Therefore very high unemployment rate for this cohort may be implausible result. However, it seems this is real situation in Kosovo, since the LFS-2009 also indicates similar results for young cohort.

Further investigation on the relationship between unemployment and education is suggested to understand why young people have very high unemployment rate compared to people in working age group.

The figure 15 provides unemployment rates for selected age groups from the census and LFS-2009. Both sources indicate very high unemployment rates for all age groups and both sexes. It is clearly seen from the figures that the results of the census are consistent with the results of the LFS-2009 with the exception of females in the age group 55-64. Unemployment rate for this age group is 34.7 percent from LFS-2009 and 13.2 percent from the census. The rates for males are 23.9 from LFS-2009 and 16.5 percent from the census. Except the age group 55-64, the difference in unemployment ratios between two sources is very low (less than 3). Based on these results it could be concluded that the census results might be more realistic considering the risk of sampling errors in LFS-2009 having relatively small number of cases for the age group 55-64 in the sample. In general the census results seem reliable and consistent with the results of LFS-2009.

Figure 14. Unemployment rate by age group and urban/rural

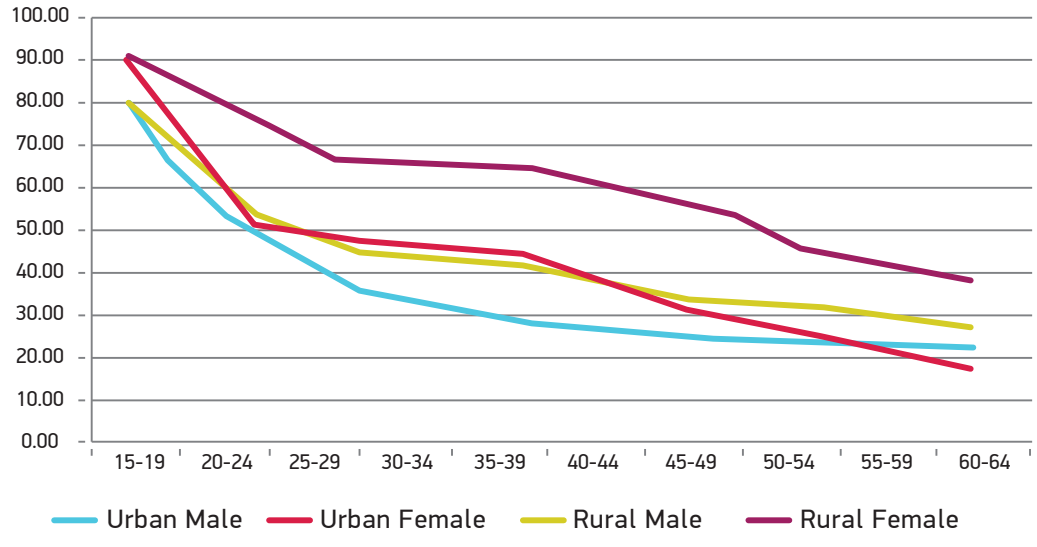
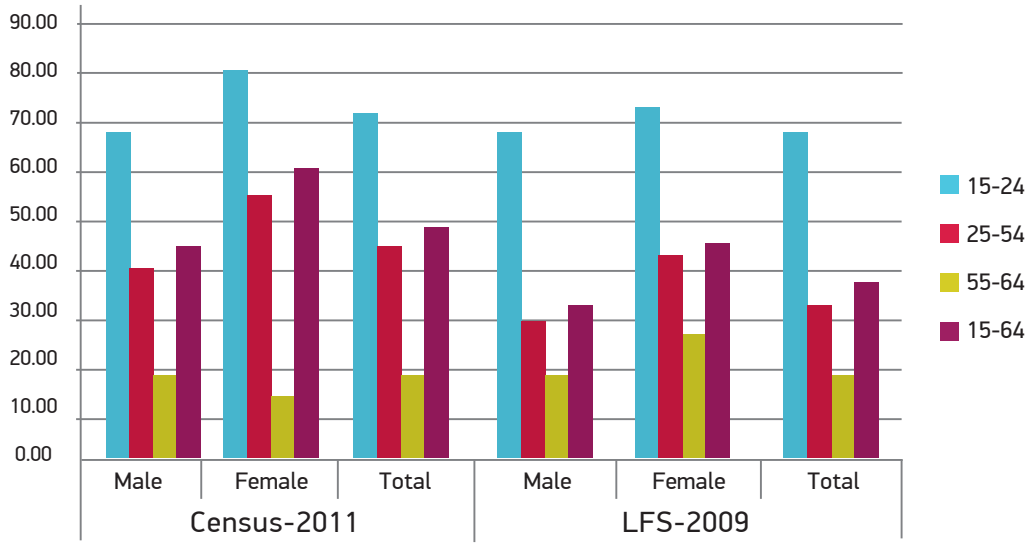


Figure 15. Unemployment rate by selected age group, Census-2011 and LFS-2009



4.3.6 Housing Characteristics

The quality of data on housing characteristics collected in the census can be compared to two household surveys-Demographic, Social and Reproductive Health Survey in Kosovo conducted in 2009 and Households Budget Survey conducted in 2010. This section covers the topics included in the census and in one of the two household surveys. There are significant numbers of housing topics included in the census but both household surveys include a limited number of housing topics.

Number of rooms

Data on number of rooms is available in the HBS-2010. According to the definition used in the census, the kitchens are included in the number of rooms. For this analysis, number of rooms including kitchen is derived from the HBS-2010 database to compare it with the results of the census.

Distribution of housing units by number of rooms from the census and HBS-2010 is presented in Figure 16. Percentage of housing units by the number of rooms reached a peak at 3 rooms in urban areas and 4 rooms in rural areas in both sources. As the figure shows, the pattern of distribution of housing units by number of rooms from the census and survey are very similar with slight difference in urban areas. The figures clearly show that the results of the census are consistent with the results of HBS.

Figure 16. Percentage of housing units by number of rooms

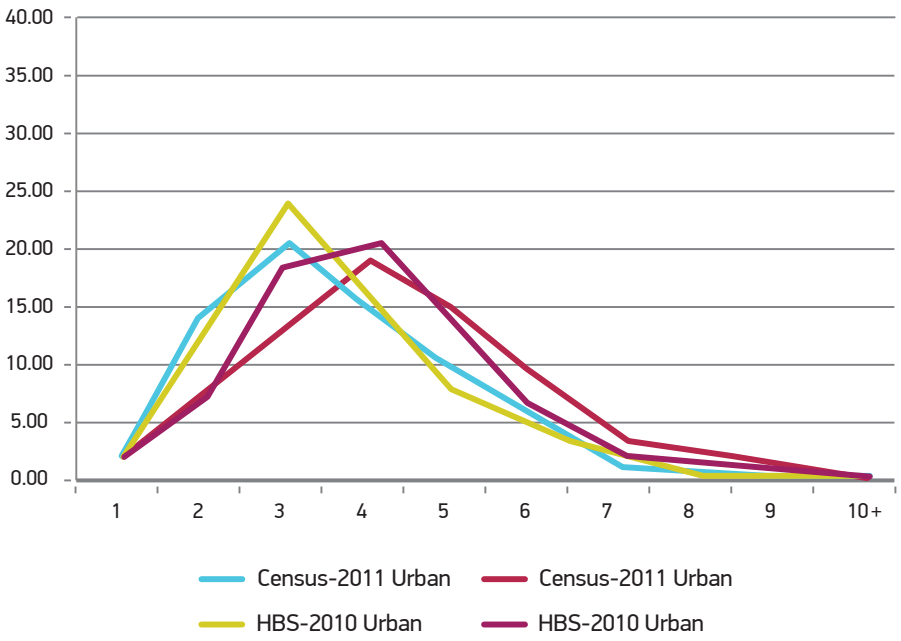


Table 5 presents the proportions of housing units by selected characteristics of housing units from the census and both DHS and HBS. Most of the characteristics included in the table have similar proportions for three sources with the exception of availability of flush toilet facilities in urban areas. This proportion is 10 percent higher in the census compared to the result of HBS. Based on the limited number of characteristics of housing units included in the study, the results of the census might be considered reliable and accurate.

Table 5. Proportion of housing units by selected facilities

Housing Characteristics	Census-2011		HBS-2010		DHS-2009	
	Urban	Rural	Urban	Rural	Urban	Rural
Piped water in the housing unit	99.00	89.10	98.90	93.33	...	...
Flush toilet in the housing unit	96.08	80.56	83.14	80.64	...	...
Bathing facility in the housing unit	95.55	86.77	96.71	92.51	93.90	82.90
Central heating system	14.63	3.27	10.35	2.65	16.20	3.00

CONCLUSION

Among all procedures put in place to ensure a quality population census in Kosovo, the major failure regards the non-coverage of three northern municipalities. Communication and outreach measures implemented with the objective of fostering the participation of minorities did also not completely achieve their goals in the enumerated part of Kosovo territory. Yet, all municipalities lead by Kosovo-Serb Mayors created a municipal census commission in the enumerated area, and a large part of its residents participated in the census. Compared to the pilot census experience, other minorities largely participated in the census as well.

The census organization worked well and achieved a satisfactory level of performance from the enumeration staff. On its turn, the very well secured data-capture and document management systems allowed to keep a full control on the census paper and electronic version of the data.

Editing and imputation processes experienced delays, but the initial error rates found in the data and the corrections requested had all a low rate. The final census database can be considered highly consistent.

Over all analysis of census data suggests very good quality of data collected in the census. The results of the census is highly coherence with the reasonable pattern of the characteristics of population and consistent with other sources of data available in Kosovo.

Main findings of the analysis of census data and the conclusions are given below for each census topic.

Age-sex distribution

The assessment of the reasonableness of the age and sex distribution of the population enumerated in the census is typically the first step taken in evaluating a census by means of demographic methods. These methods are applied to provide an indication of the general quality of the data. The results of demographic methods suggest the followings:

- Age-sex distributions of population are highly reliable and consistent with the expected pattern;
- There is a clear sign of baby-boom for the cohort born in the year 2000;
- There is no age misreporting or heaping observed in the distribution of age; the results of
- Whipple Index indicates perfect age reporting;
- The pattern of age ratios in general reflects smooth age distribution of population with relatively low ratios among female population in age group 25-29 and 55-59 living in rural area;
- The result of child population (age group 0-4) estimation indicates the possibility of over-count of this cohort for about 5 percent. However, this estimation might not be reliable because of high risk of under-coverage of births with the benchmark source (vital statistics data) since the births which are registered 2 or more years after the births are not included in birth statistics.

Main limitation of this assessment is the lack of data for using sophisticated demographic methods providing some measurements on coverage errors in the census. If periodical censuses and demographic indicators during the inter-censual period were available, the demographic methods used for cohort analysis such as cohort survival ratio and cohort component method could be included in this report. Therefore only demographic methods which do not require the results of consecutive censuses are covered in this assessment.

Education characteristics

Analysis of data on education characteristics is undertaken using three types of data regarding illiteracy, school attendance and the completed level of education and this data is compared with the result of school enrolment 2010-2011 and DHS-2009 . The analysis of data suggests very high quality of census data as the distribution of illiterate population and population with education attainment by age cohorts and sex are coherence with the results of the survey. Moreover, the size of students enumerated in the census is consistent with administrative data on school attendance by sex and age cohorts of the students.

Marital status

The quality of census data on marital status is analysed by comparing the census data with the results of DHS-2009. This analysis indicates that the distribution of single and currently married population by age group, sex and urban/rural is highly consistent with the results of DHS and the expected pattern.

Children ever born and infant mortality

The quality of census data on children ever born is evaluated using the mean children ever born calculated for the census and the DHS-2009. The levels of MCEB by age groups of women calculated from the census and survey are almost identical for all ages indicating that the census data on children ever born is reliable and consistent with the survey results. Data on annual infant mortality in the period 2005-2010 from the census are consistent with data from vital statistics.

Economic characteristics

Labour Force Survey (LFS) conducted in 2009 was used to assess the quality of census data on economic characteristics. The analysis compared data on employment and unemployment among the active population by age groups, sex and location (rural/urban). General patterns follow very close characteristics in the census and in the survey.

The distribution of employed population by age groups indicates a big difference in the level of employment by locations and sexes. As it is expected the male population has higher proportion of employment in both rural and urban areas compared to female population living in these areas. Moreover, according to the census the population living in urban areas have remarkably higher

proportion of employment in all ages. The LFS-2009 results cannot be broken down by rural/ urban criteria, therefore it is not possible to conclude whether this is real situation in the country or whether there is reporting error in the census because of problems of understanding the definition of economic activity particularly for agricultural activities and mostly for women. This conclusion needs further investigation to clarify the level of employment in rural and urban areas.

Regarding unemployment, data from both sources show similar patterns, with the particularity for the youngest cohort (aged 15-19) being affected by very high unemployment rates. This implausible data might be partially explained by students looking for a job; yet, it is highly consistent in both data sources and can reflect a country reality. Some unemployment differences were detected in the age group 55-64, with lower rates in the census. Their analysis lead to conclude that the census results might be more realistic considering the risk of sampling errors in LFS-2009, which has relatively small number of cases for this age group in the sample.

Housing characteristics

For the reason of compatibility of definitions, only some characteristics of housing units (number of rooms, availability of central heating, piped water, bathing and flush toilet facilities in the housing unit) from the census could be compared with data from the Demographic Health Survey and with the Household Budget Survey. Results from the census are highly compatible with the ones of both surveys, and the analysis indicates that the slight difference can be attributed to sampling errors in the survey in a larger extent than to errors of census data collection.

Bibliography

A systematic approach to automatic edit and imputation. Fellegi, I.P., and Holt, D., Journal of the American Statistical Association, 71, 17-35 (1976)

Algorithm for finding a general formula for the non-negative solutions of a system of linear equations. Chernikova, N.V., USSR Computational Mathematics and Mathematical Physics, 4, 151-158 (1964)

Algorithm for finding a general formula for the non-negative solutions of a system of linear inequalities. Chernikova, N.V., USSR Computational Mathematics and Mathematical Physics, 5, 228-233 (1965)

Evaluating Censuses of Population and Housing Census. US Census Bureau, Statistical Training Document, ISP-TR-5, 1985; http://www.census.gov/apsd/techdoc/rr85-24.pdf

Large scale computation of elementary flux modes with bit pattern trees. Terzer, M. and Stelling, J., Bioinformatics, July 28 (2008)

Methods and Materials of Demography. 2nd edition, Siegel and Swanson, 2004

Post Enumeration Survey Operational Guidelines. UN Department of Economic and Social Affairs, Statistics Division; http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manual_PESen.pdf

Principles and Recommendations for Population and Housing Censuses, Revision 2. Sales No. E.07.XVII.8, United Nations, New York, 2008; http://unstats.un.org/unsd/publication/SeriesM/Seriesm_67rev2e.pdf

Processing of erroneous and unsafe data. De Waal, T., Ph.D. Thesis, Erasmus University Rotterdam (2003)

Two algorithms for Error Localization Problems. Scarnò M and Caramanna L., 45th Scientific Meeting of the Italian Statistical Society (2010); http://adamsoft.caspur.it/English/Guide.html

ANNEX 1

Resulting data from editing and imputations of R1 questionnaire on dwellings

QUESTIONS	MODALITIES	RAW DATA	FINAL DATA
1.Type of Housing	Missing	257	0
	Conventional dwelling	412284	412519
	Other housing unit (former shop, tent, etc.) with one or more usual residents	400	365
	No shelter (homeless)	40	0
	Skipped(DataEntry Mistakes)	397	0
Total Dwellings		413378	412884
2.Occupancy Status Of Dwelling	Missing	2373	0
	Occupied by one or more usual residents	290870	293078
	Occupied by residents not to be enumerated (diplomatic personnel, etc.)	1083	1083
	Reserved for seasonal/occasional stay or secondary use (holidays,seasonal works etc.)	18570	18550
	Vacant (available for sale,rent,demolition,etc.)	99765	99808
	Skipped	717	365
Total Dwellings		413378	412884
3A. Rooms in the housing unit (including rooms for business)	Missing	1105	0
	Valid	292476	293443
	Skip Questions	119797	119441
Total Dwellings		413378	412884
3B.Total surface area in the housing unit (including rooms for business)	Missing	3167	0
	Valid	290414	293443
	Skip Questions	119797	119441
Total Dwellings		413378	412884
4A. Rooms Used Mainly For Business Or Professional Purposes	Missing	1801	0
	Valid	291780	293443
	Skip Questions	119797	119441
Total Dwellings		413378	412884

ANNEX 1 - Resulting data from editing and imputations of R1 questionnaire on dwellings

4B.Total surface used Mainly For Business Or Professional Purposes	Missing	1851	0
	Valid	7457	293431
	Skip Questions	404070	119453
Total Dwellings		413378	412884
5.Water Supply System In The Housing Unit	Missing	406	0
	Piped water from a public service	194728	204365
	Piped water from other sources	78436	82609
	Piped water within the building but outside the housing unit	9863	899
	Piped water outside the building	7213	3413
	No piped water available	2935	2157
	Skipped	119797	119441
Total Dwellings		413378	412884
6.Type Of Sewage Disposal System	Missing	518	0
	Wastewater empties into a piped system connected to a public sewage disposal plant	168356	168453
	Wastewater empties into a piped system connected to a private sewage disposal plant	41872	41914
	All other arrangements (e.g. open ditch,a pit, a river, etc.)	70194	70311
	No sewage disposal system	12641	12765
	Skipped	119797	119441
Total Dwellings		413378	412884
7.Flush Toilet In The Housing Unit	Missing	986	0
	Flush toilet in the housing unit	255431	257428
	Flush toilet outside the housing unit but in the building	7190	6327
	Flush toilet outside the building	15893	15147
	No flush toilet	14081	14541
	Skipped	119797	119441
Total Dwellings		413378	412884
8.Bathing Facilities In The Housing Unit	Missing	2431	0
	Fixed bath or shower	264326	265181
	No fixed bath or shower	26824	28262
	Skipped	119797	119441
Total Dwellings		413378	412884

9.Kitchen/Cooking Facilities In The Housing Unit	Missing	1183	0
	Kitchen	205061	205653
	Kitchenette	57755	57996
	Cooking facilities provided in another type of room	15975	16064
	No kitchen/kitchenette/other cooking facilities	13607	13730
	Skipped	119797	119441
Total Dwellings		413378	412884
10.Main Type Of Heating	Missing	592	0
	Central heating from a public heating centre	9631	9698
	Central heating from an installation in the building or in the housing unit	14093	14231
	Stove	252506	252837
	Fireplace	4233	4187
	Portable electric heater	11456	11402
	Other	860	882
	No heating at all	210	206
	Skipped	119797	119441
Total Dwellings		413378	412884
11.Main Type Of Energy Used For Heating	Missing	1348	0
	Wood	260228	259031
	Coal	7317	6932
	Oil	6768	10812
	Gaseous fuel	981	817
	Electricity	16403	15356
	Other type of energy	337	289
	Skipped	119996	119647
Total Dwellings		413378	412884
12A. Type Of Building	Missing	1435	0
	Residential	291445	316966
	Non-residential	423	387
	Skipped	120075	95531
Total Dwellings		413378	412884

12B.Type Of Building If residential, answer to the following question	Missing	9276	0
	Detached house	234275	245686
	Semi-detached	18933	20879
	Row or terraced house (with at least 3 attached or connected dwellings)	4778	6458
	Apartment building or block of flats	25292	43555
	Other	376	388
	Skipped	120448	95918
Total Dwellings		413378	412884
13.Year Of Construction Of The Building	Missing	5336	0
	Valid	287966	317353
	Skipped	120076	95531
Total Dwellings		413378	412884
14.Number Of Floors In The Building	Missing	3078	0
	Valid	290225	317353
	Skipped	120075	95531
Total Dwellings		413378	412884
15A.Materials Of Which Outer Walls Of The Building Are Constructed	Missing	3016	0
	Concrete and reinforced concrete	11421	15142
	Concrete blocks	11530	13345
	Bricks and/or stones	58677	64215
	Mixed materials	16334	17901
	Red and White Brick	184951	198812
	Other (specify)	7373	7938
	Skipped	120076	95531
Total Dwellings		413378	412884

ANNEX 2

Resulting data from editing and imputations of R2 questionnaire on households

QUESTIONS	MODALITITES	RAW DATA	FINAL DATA
1.Tenure Status	Missing	2674	0
	A member of the household is the owner of all or part of the housing unit	280735	284023
	A member of the household is a tenant of all or part of the housing unit	8262	8367
	The household occupies all or part of the housing unit under some other form of tenure	4391	4432
	The household is living in collective dwelling	1016	268
	Skipped	18435	0
	Total	315513	297090
2.Type Of Ownership Of The Housing Unit	Missing	3878	0
	Private ownership	34665	9430
	Public ownership (local or central government)	1415	1260
	Mixed ownership	834	783
	Co-operative ownership	873	670
	Other type of ownership	646	656
	Skipped	273202	284291
	Total	315513	297090
3.Does The Household Have A Telephone	Missing	2877	0
	Fixed telephone only	10186	10298
	Mobile phone only	227807	229889
	Both fixed and mobile	51869	52518
	Neither fixed nor mobile telephone	4339	4385
	Skipped	18435	0
	Total	315513	297090
4.Does The Household Have A Computer?	Missing	3691	0
	Yes	183427	185849
	No	109960	111241
	Skipped	18435	0
	Total	315513	297090

5.Does The Household Have Access To The Internet From The Housing Unit?	Missing	4537	0
	Yes	166805	169451
	No	125736	127639
	Skipped	18435	0
	Total	315513	297090
6.Does Your Household Have Any Land In Ownership Or In Use In Kosovo?	Missing	4235	0
	Yes	195266	195124
	No	97577	101966
	Skipped	18435	0
	Total	315513	297090
7.What is the total area of land in ownership by thehousehold	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
8.What is the total area rented and used by others	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.What is the use of households total land	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Arable land (annual crops)	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Greenhouses	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Meadows (clover, mixed grass, lucerne, etc.)	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Orchards	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090

9.Vineyards	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Fallow (more than 2 years not cultivated field)	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Pasture	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Forests	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
9.Other (house yard, ponds,...)	Missing	0	0
	valid	200081	195124
	Skipped	115432	101966
	Total	315513	297090
10.Wheat	Missing	0	0
	valid	200081	200093
	Skipped	115432	96997
	Total	315513	297090
10.Maize	Missing	0	0
	valid	200081	200093
	Skipped	115432	96997
	Total	315513	297090
10.Patato	Missing	0	0
	valid	200081	200093
	Skipped	115432	96997
	Total	315513	297090
10.Pepers	Missing	0	0
	valid	200081	200093
	Skipped	115432	96997
	Total	315513	297090

11.Does your household own any livestock animals	Missing	2713	0
	Yes	89057	89894
	No	205308	207196
	Skipped	18435	0
	Total	315513	297090
12.Total cattle(young and adult total)	Missing	0	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Cows kept for milking	Missing	0	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Sheep(young and adult total)	Missing	0	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Goats(young and adult total)	Missing	0	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Pigs(young and adult total)	Missing	0	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Horses and donkeys(young and adult total)	0	88380	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Poultry(young and adult total)	0	35696	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090
12.How many Beehives	0	87193	0
	valid	91941	89894
	Skipped	223572	207196
	Total	315513	297090

ANNEX 2 – Resulting data from editing and imputations of R2 questionnaire on households

13.Does your household own any agriculture machinery	Missing	2068	0
	Yes	51947	51663
	No	243063	245427
	Skipped	18435	0
	Total	315513	297090
14.Does The Household Own Tractor single axletree Number	Missing	545	0
	valid	53485	51663
	Skipped	261483	245427
	Total	315513	297090
14.Does The Household Own Tractor single axletree HPower	Missing	4029	0
	valid	34899	51663
	Skipped	276585	245427
	Total	315513	297090
14.Does The Household Own Tractor two axletree Number	missing	610	0
	valid	53420	51663
	Skipped	261483	245427
	Total	315513	297090
14.Does The Household Own Tractor two axletree HPower	missing	2001	0
	valid	12375	51663
	Skipped	301137	245427
	Total	315513	297090
14.Does The Household Own Cereals combines Number	missing	745	0
	valid	53285	51663
	Skipped	261483	245427
	Total	315513	297090
14.Does The Household Own Cereals combines HPower	missing	945	0
	valid	474	51663
	Skipped	314094	245427
	Total	315513	297090
14.Does The Household Own Truck Number	Missing	718	0
	valid	53312	51663
	Skipped	261483	245427
	Total	315513	297090
14.Does The Household Own Truck combines HPower	Missing	1182	0
	valid	1512	51663
	Skipped	312819	245427
	Total	315513	297090

ANNEX 3

Resulting data from editing and imputations of R3 questionnaire on individuals

QUESTIONS	MODALITIES	RAW DATA	FINAL DATA
2.Are you male or female ?	Male	874422	875900
	Female	863232	863925
	Skip Questions	694	0
3.Day of birth	Missing	11502	2
	Valid	1726152	1739823
	Skip Questions	694	0
3.Month of birth	Missing	11502	2
	Valid	1726152	1739823
	Skip Questions	694	0
3.Year of birth	Valid	454	0
	Missing	1737200	1739825
	Skip Questions	694	0
4.Where was your mothers place of usual residence at the time of your birth	Missing	5700	0
	In Kosovo	1221034	1232368
	Kosovo Different Municipality	456440	452741
	Outside Kosovo	54480	54716
	Skip Questions	694	0
4a.Municipality Code	Missing	8007	3176
	Valid	461284	449565
	Skip Questions	1269057	1287084
4a.Settlement Code	Missing	10209	7404
	Valid	459082	447362
	Skipped	1269057	1285059
4b.Country Code	Missing	233	233
	Valid	54227	54483
	Skip Questions	1683888	1685109
5.Marital status	Missing	18135	0
	Never Married	876214	882696
	Married with Marriage Certificate	687420	715026
	Married without marriage Certificate	83896	76007
	Widowed	67089	60865
	Divorced	4900	5231
	Skip Questions	694	0

6.Have You Given Birth To Any Children Born Alive?	Missing	24691	18945
	Yes	403909	408980
	No	203056	200781
	Skip Questions	1106692	1111119
6.Number of Born Alive Children	Missing	22054	1008
	Valid	408578	407698
	Skip Questions	1307716	1331119
7.Have You Given Birth To Any Children Born Alive After The Year 2005?	Missing	31716	23121
	Yes	112611	113205
	No	286327	293532
	Skip Questions	1307694	1309967
7.Number of children born alive after 2005	Missing	34226	479
	Valid	113970	112669
	Skip Questions	1590152	1626677
8.Children Born Alive After The Year 2005 Died Before The Age Of 1 Year?	Missing	37758	30818
	Yes	1397	1452
	No	109041	115468
	Skipped	1590152	1592087
8.Number of Died Children	Missing	37701	50
	Valid	1459	1396
	Skipped	1699188	1738379
8.First Birth Month	Missing	192	191
	Valid	1327	1323
	Skip Questions	1736829	1738311
8.First Birth Year	Missing	185	184
	Valid	1334	1330
	Skip Questions	1736829	1738311
8.Second Birth Month	Missing	118	117
	Valid	189	189
	Skip Questions	1738041	1739519
8.Second Birth Year	Missing	109	108
	Valid	198	198
	Skip Questions	1738041	1739519
8.Third Birth Month	Missing	80	79
	Valid	20	2
	Skip Questions	1738248	1739727

8.Third Birth Year	Missing	78	77
	Valid	22	21
	Skip Questions	1738248	1739727
8.First Death Month	Missing	440	437
	Valid	1079	1077
	Skip Questions	1736829	1738311
8.First Death Year	Missing	417	414
	Valid	1102	1100
	Skip Questions	1736829	1738311
8.Second Death Month	Missing	178	177
	Valid	129	129
	Skip Questions	1738041	1739519
8.Second Death Year	Missing	165	164
	Valid	142	142
	Skip Questions	1738041	1739519
8.Third Death Month	Missing	91	89
	Valid	9	9
	Skip Questions	1738248	1739727
8.Third Death Year	Missing	86	84
	Valid	14	14
	Skip Questions	1738248	1739727
9.To Which Ethnic Or Cultural Background Do You Belong?(First)	Missing	1143	1840
	Albanian	1615563	1616869
	Serb	25461	25532
	Turk	18727	18738
	Bosnia	27517	27533
	Roma	8801	8824
	Ashkali	15420	15436
	Egyptian	11519	11524
	Goran	10262	10265
	Other(specify)	2329	2352
	Prefer not to answer	909	912
	Skipped	697	0

ANNEX 3 - Resulting data from editing and imputations of R3 questionnaire on individuals

9.To Which Ethnic Or Cultural Background Do You Belong?(Second)	Missing	4	0
	Albanian	6	8
	Serb	2	2
	Turk	3	3
	Bosnia	2	3
	Roma	0	0
	Ashkali	0	0
	Egyptian	2	2
	Goran	0	0
	Other(specify)	6	6
	Prefer not to answer	0	0
	Skipped	1738323	1739801
10.What is your religion ?	Missing	1802	2495
	Islam	1662228	1663412
	Orthodox	25755	25837
	Catholic	38253	38438
	Other Specify	1188	1188
	No religion	1240	1242
	Prefer not to answer	7189	7213
	Skipped	693	0
11.What is your mother tongue (First)	Missing	1559	2252
	Albanian	1643527	1644865
	Serb	27908	27983
	Turk	19558	19568
	Bosnian	28974	28989
	Roma	5847	5860
	Other (specify)	10282	10308
	Skipped	693	0
11.What is your mother tongue (Second)	Missing	0	0
	Albanian	0	0
	Serb	0	0
	Bosnian	3	3
	Turk	0	0
	Roma	0	0
	Other (specify)	3	3
	Skipped	1738342	1739819

12.What other languages do you speak(First)	Missing	35682	36382
	Albanian	112795	32593
	Serb	236027	246633
	Turkish	22074	23183
	Bosnian	15912	15022
	Roma	663	587
	English	216408	226101
	German	51759	54223
	French	6828	7205
	Other (Specify)	14647	15510
	None	1024859	1082386
	Skipped	694	0
12.What other languages do you speak(Second)	Missing	0	0
	Albanian	392	36
	Serb	24903	12274
	Turkish	29465	29383
	Bosnian	25677	25172
	Roma	1027	957
	English	67542	61061
	German	51845	51918
	French	9959	10107
	Other (Specify)	18078	18184
	Skipped	1509460	1530733
12.What other languages do you speak(Third)	Missing	0	0
	Albanian	105	7
	Serb	27	24
	Turkish	3340	1178
	Bosnian	10001	9039
	Roma	919	702
	English	18415	16374
	German	15037	13390
	French	5831	5655
	Other (Specify)	11616	11332
	Skip Questions	1673057	1682124

12.What other languages do you speak(Fourth)	Missing	0	0
	Albanian	21	0
	Serb	3	2
	Turkish	2	2
	Bosnian	1061	248
	Roma	310	267
	English	4043	3309
	German	3335	2851
	French	2038	1845
	Other (Specify)	4886	4609
	Skipped	1722649	1726692
12.What other languages do you speak(Fifth)	Missing	0	0
	Albanian	2	0
	Serb	0	0
	Turkish	0	0
	Bosnian	1	0
	Roma	35	8
	English	396	146
	German	743	595
	French	564	489
	Other (Specify)	1466	1350
	Skipped	1735141	1737237
12.What other languages do you speak(Sixth)	Missing	0	0
	Albanian	0	0
	Serb	0	0
	Turkish	0	0
	Bosnian	0	0
	Roma	0	0
	English	18	5
	German	91	62
	French	138	112
	Other (Specify)	353	292
	Skipped	1737748	1739354

12.What other languages do you speak(Seventh)	Missing	0	0
	Albanian	1	1
	Serb	0	0
	Turkish	0	0
	Bosnian	0	0
	Roma	0	0
	English	0	0
	German	6	0
	French	37	33
	Other (Specify)	71	60
	Skipped	1738233	1739731
12.What other languages do you speak(Eight)	Missing	0	0
	Albanian	0	0
	Serb	0	0
	Turkish	0	0
	Bosnian	0	0
	Roma	0	0
	English	0	0
	German	0	0
	French	5	0
	Other (Specify)	17	16
	Skipped	1738326	1739809
12.What other languages do you speak(Nineth)	Missing	0	0
	Other (Specify)	4	4
	Skipped	1738344	1739821
13.Single citizenship code	Missing	15160	15395
	Valid	1722494	1724430
	Skipped	694	0
13.In case of double citizenship code	Missing	237	0
	Valid	40212	38161
	Skipped	1697899	1701664
14.Were you present in your placeof usual residence on the midnight between 31/03/2011 and 01/04/2011	Missing	5916	0
	Yes	1692340	1698407
	No	39398	41418
	Skipped	694	0

ANNEX 3 - Resulting data from editing and imputations of R3 questionnaire on individuals

15.What is the reasons of your absence	Missing	8316	0
	Employment	11244	11898
	Study	7796	8291
	Family	13414	14343
	Other reasons (private visit, medical, etc.)	6346	6886
	Skipped	1691232	1698407
16.Have you ever lived in a different place	Missing	8273	0
	Yes	521993	517992
	No	1180094	1190546
	Skipped	27988	31287
17.Where was your previous residence	Missing	15637	0
	In Kosovo	388859	389184
	Outside Kosovo	130522	128808
	Skipped	1203330	1221833
17a.Municipality No	Missing	15109	8800
	Valid	389402	380384
	Skipped	1333837	1350641
17a.Settlement No	Missing	16877	10558
	Valid	387634	378626
	Skipped	1333837	1350641
17b.Country No	Missing	541	505
	Valid	129966	128303
	Skipped	1607841	1611017
18.When did you move to current place of usual (month)	Missing	72729	0
	Valid	462289	517992
	Skip Questions	1203330	1221833
18.When did you move to current place of usual (year)	Missing	27027	0
	Valid	507991	517992
	Skip Questions	1203330	1221833
19.Why did you move to current place of usual	Missing	29572	0
	For employment reasons	43679	44448
	For education or training purposes	14543	14403
	For family reasons	353236	361831
	For 1998-1999 war reasons	35190	35529
	For other reasons	58798	61781
	Skipped	1203330	1221833

20.School/college/university	Missing	26699	0
	Not attending school	1017925	1054437
	Primary (grades 1 to 5)	180313	177617
	Lower secondary (grades 6 to 9)	138017	133292
	Upper secondary (grades 10 to 12)	119282	112848
	College/University(as undergraduate)	75306	69600
	College/University(as post-graduate)	11060	11149
	College/University(as doctorate/Phd)	1339	1235
	Skipped	168407	179647
21.What is your highest completed education level	Missing	25651	0
	No formal education	73986	60727
	No completed education level	84098	76462
	Primary (completed 4 grades - old system)	113280	100788
	Primary (completed 5 grades - new system)	125412	144764
	Lower secondary (completed 8 grades - old system)	353226	339965
	Lower secondary (completed 9 grades - new system)	145971	161652
	Upper secondary (completed 12 grades - old system)	330673	304245
	Upper secondary (completed 13 grades - new system)	94432	123660
	Post secondary Vocational	31425	31568
	Degree (bachelor)	74776	75213
	Postgraduate degree(master)	8906	8795
	Doctorate/Phd	1323	1336
	Skipped	275189	310650
22.Type Code	Missing	0	0
	Valid	543428	543435
	Skipped	1194920	1196390
22.Profile Code	Missing	40998	40096
	Valid	533518	534980
	Skipped	1163832	1164749
23.Do you know how to read and write	Missing	6053	0
	Yes	111959	82188
	No	57904	55001
	Skipped	1562432	1602636

24.In the week before the census day, did you work	Missing	10706	0
	Yes	267793	257353
	No	1017869	994895
	Skipped	441980	487577
25.Did you work during the previous week	Missing	13255	0
	Yes	28088	18077
	No	1005388	976818
	Skipped	691617	744930
26.Work during the week before the census day	Missing	21033	0
	Yes	6212	5024
	No	996864	971794
	Skipped	714239	763007
27.Reason for not working	Missing	14381	0
	Unemployed and never worked before	231295	230559
	Unemployed but worked before	136275	138557
	Attending school (pupil/student)	222995	173213
	Looking after home/family (housewives,homemakers)	309945	313368
	Retired from former work activity	54141	56147
	Pension or capital income recipient (income from property or investments ect.)	28190	29399
	Waiting to start a job already obtained?	2051	3700
	Serving military service	91	0
	Serving a sentence in a prison	387	1687
	Permanently unable to work due to sickness or disability	12851	17172
	Other, specify	6915	7992
	Skipped	718831	768031
28.How many hours working during the week before the census day	Missing	8173	0
	Valid	270603	275430
	Skip Questions	1459572	1464395
29.How many hours per week do you usually work in your main job	Missing	6751	0
	Valid	276617	280454
	Skip Questions	1454980	1459371

ANNEX 3 - Resulting data from editing and imputations of R3 questionnaire on individuals

30.In your main job are you	Missing	7366	490
	Employee	223789	227002
	Employer with paid employees including work in your own land	13313	13507
	Self-employed, without paid employees (including work in your own land)	33523	33998
	Assisting relatives in family business without a fixed wage or salary, including work in family land)	5377	5457
	Skipped	1454980	1459371
31.Isco Code	Missing	8586	0
	Valid	274782	280454
	Skipped	1454980	1459371
33.What is your place of work	Missing	15119	3636
	Fixed place of work	237246	249596
	Work mainly at or from home	14520	9748
	No fixed place of work	16483	17474
	Skipped	1454980	1459371
34.Main type of activity of employer	Missing	0	0
	Valid	283368	280454
	Skipped	1454980	1459371
35.Search for work in last four week	Missing	12129	0
	Yes	254870	251248
	No	751557	689936
	Skipped	719792	798641
36.Able to start to work	Missing	13163	0
	Yes	459711	428276
	No	545682	512908
	Skipped	719792	798641
37.Where do you work or attend school or university	Missing	26496	44276
	In Kosovo same Municipality and settlement	484266	461909
	In Kosovo Different Municipality and or settlement	152240	150692
	Outside of Kosovo	15464	15190
	Not Applicable/Not Work or Attend School	784223	755421
	Skipped	275659	312337

38.How often you travel between your home and work	Missing	35262	35405
	Daily	589321	584993
	Weekly	22373	23632
	Less often then once a week	16204	18285
	Not Applicable(work mainly at/from home)	18092	9752
	Skipped	1057096	1067758
39.Do you have any of the following long-lasting difficulties(First)	Missing	3850	4652
	Deafness or severe hearing impairment	9040	9800
	Blindness or severe visual impairment	6915	7048
	A condition that substantially limits one or more basic physical activities such as walking, climbing stairs, lifting or	27298	28185
	A learning or intellectual disability	3014	3234
	A long-standing psychological or emotional difficulty	5430	5789
	Other, including any long-standing illness	23136	23674
	No, I do not have any long-standing difficulty	1658885	1657443
	Skipped	780	0
39.Do you have any of the following long-lasting difficulties(Second)	Missing	12	0
	Deafness or severe hearing impairment	165	33
	Blindness or severe visual impairment	2318	2321
	A condition that substantially limits one or more basic physical activities such as walking, climbing stairs, lifting or	2524	2427
	A learning or intellectual disability	1719	1729
	A long-standing psychological or emotional difficulty	1417	1467
	Other, including any long-standing illness	3467	3471
	No, I do not have any long-standing difficulty	22	0
	Skipped	1726704	1728377

39.Do you have any of the following long-lasting difficulties(Third)	Missing	45	0
	Deafness or severe hearing impairment	546	102
	Blindness or severe visual impairment	72	66
	A condition that substantially limits one or more basic physical activities such as walking, climbing stairs, lifting or	1064	955
	A learning or intellectual disability	695	644
	A long-standing psychological or emotional difficulty	878	877
	Other, including any long-standing illness	1540	1466
	No, I do not have any long-standing difficulty	7	0
	Skipped	1733501	1735715
40.Learning, remembering or concentrating	Missing	12540	17888
	Not at all	29936	1683011
	Some	19068	19260
	A lot	8559	8617
	Cannot do at all	10741	10930
	Skipped	1657504	119
40.Dressing, bathing or getting around inside the home	Missing	16464	18620
	Not at all	37472	33866
	Some	16366	16503
	A lot	6314	6373
	Cannot do at all	6908	7020
	Skipped	1654824	1657443
40.Going outside the home alone to shop or visit a doctor's	Missing	14958	17097
	Not at all	28049	24397
	Some	16697	16792
	A lot	9927	9991
	Cannot do at all	13893	14105
	Skipped	1654824	1657443
40.Working at a job or business or attending school or college	Missing	15763	17858
	Not at all	13296	9566
	Some	9361	9441
	A lot	9334	9397
	Cannot do at all	35770	36120
	Skipped	1654824	1657443

40.Participating in other activities	Missing	16411	18518
	Not at all	15391	11711
	Some	11307	11381
	A lot	8318	8364
	Cannot do at all	32097	32408
	Skipped	1654824	1657443
41.First livelihood	Missing	20915	0
	Work (including work in your own land,work in your own or family business)	387721	390281
	Property or other investments (rent, shop, etc.)	6264	6330
	Pensions	110306	112020
	Social Care	83016	84683
	Other transfers(unemployment benefits,sickness and maternity allowances,scholarship)	6991	7102
	Remittances	91147	92317
	Supported by another person (excluding remittances)	911268	922464
	Other sources	120026	124628
	Skipped	694	0
41.Second livelihood	Missing	14591	0
	Work (including work in your own land,work in your own or family business)	35541	18720
	Property or other investments (rent, shop, etc.)	4287	4086
	Pensions	20122	15317
	Social Care	15981	11741
	Other transfers(unemployment benefits,sickness and maternity allowances,scholarship)	1981	1737
	Remittances	62869	60314
	Supported by another person (excluding remittances)	153670	113060
	Other sources	502019	487950
	Skipped	927287	1026900

NOTE: In the before section 1423 Cases from Collective living are not included in this counting

ANNEX 4

Age pyramid by area of living