

THE 1967 CENSUS OF THE WEST BANK AND GAZA STRIP: A DIGITIZED VERSION

Dataset Documentation

Levy Economics Institute of Bard College

Joel Perlmann, *Project Director*

February 2012

Table of Contents

Introduction	3
Organization of the 1967 Census Enumeration and Reports	4
Creating the Digitized Dataset	5
Table Notes	6
The Geographic Area to Which the Table Refers	8
Other Census Datasets: Digitized Tables vs. Public-use Samples	8
Acknowledgments	10
Suggested Citation for Material Taken from This Site	11
Correspondence	11
References Cited	11

Introduction

In the summer of 1967, just after the Six-Day War brought the West Bank and Gaza Strip under Israeli control, the Israeli Central Bureau of Statistics (ICBS) supervised a census in these territories. The census included an impressive array of questions about individuals, households, and the quality of residences. Indeed, it included much the same range of questions that Israel had asked or would ask of its own population in the 1961 and 1972 censuses, respectively.

Questions deal with age, sex, religion, place of residence, educational attainment, occupation, industrial sector, income, household structure, health, female fertility, and housing conditions. Moreover, the enumeration also asked two crucial questions about refugee status: Had the individual lived prior to the 1948 War in the area that became the State of Israel? And, Was the individual living in or outside of a refugee camp at the time of the 1967 enumeration?

The ICBS prepared seven volumes of reports based on this enumeration. These constitute the first modern census reports on the Palestinian population (except for the reports based on the 1961 Israel population census, which is limited to Palestinian citizens of Israel at the time). Jordanian and Egyptian censuses during the preceding decade had canvassed the West Bank and Gaza Strip, respectively, but even the former had published only a few general tables. Yet these volumes have not been used extensively in the writing on the evolution of the occupied territories, particularly during the last quarter century. One reason for this situation is that they are not widely available. Also, even when in hand, the published volumes are not easy to use; they are subject to all the limitations of older volumes filled with numbers – indeed, all the limitations of older volumes published as quickly as possible in order to be of use at the time. They have not been made machine-readable until now.

The Levy Economics Institute therefore now offers the content of these published volumes in machine-readable form (in Excel and PDF format), free of charge, in the hope that the data can be exploited by researchers interested in a fuller understanding of the evolution of the social history of the Palestinian people in the territories. The material in the tables is of great value even on its own, but its value is radically enhanced when thought of in the context of the two later censuses of the Palestinians in the same area: the 1997 and 2007 censuses of the Palestinian Authority's Central Bureau of Statistics (PCBS). Tabular data from these censuses are available on the website of the PCBS, and public-use sample datasets (explained below) can also be obtained, either from an online archive of such documents (IPUMS International) or from the PCBS itself. Alone, the information from 1967 provides a snapshot of a population at a

particular moment in time – a critical historical moment, as things turned out, and 30 years prior to the next census benchmark for them. Together with the later material, the 1967 information provides a historical snapshot to which later ones can be compared.

In addition to the Excel tables, the online database includes a PDF version of the introduction to six of the seven volumes of reports. The introductions include the ICBS analysis of major findings, in addition to extensive descriptions of how the enumeration and data processing were carried out and discussions of possible errors, from sampling and other sources.

Organization of the 1967 Census Enumeration and Reports

The census was conducted in the West Bank, Gaza Strip, Golan Heights, and (a month later) East Jerusalem. However, we have omitted the few tables dealing with the Heights from the online collection, as they pertain to a relatively small and non-Palestinian population. The remaining tables in the six volumes of reports include 44 found in analytic introductory sections and another 221 that compose the main text of the volumes. Many of the text tables are complex, dealing with several population characteristics at once. And some are also very long, including detailed social data for each village, town, and city and subtotaled by regions.

Like the Israeli censuses of the time (and the American ones), the enumeration effort was divided into two parts: everyone was asked a brief series of questions (*the short-form questionnaire*) and a sample of the population were asked a much fuller series of questions (*the long-form questionnaire*). The difference is critical for understanding the military and administrative context for the work. While the ICBS directed the enumeration and the preparation of the reports, most of the work was carried out for the Israel Defense Force (IDF). The short form, completed for the entire population, included the name, sex, age, religion, and street address of every resident in every household or institution. The result amounted to a population register no doubt of use for the military. It is also the basis for the tabulations in volume 1 of the reports. The next four volumes were based on sample data from the long form. The final two volumes were officially carried out not for the IDF but for the Jerusalem Municipality; it covered the population of East Jerusalem (as defined after the War) from both short and long forms.

Israeli army personnel, under the supervision of the ICBS served as census takers; a curfew was declared for the day of the enumeration, and the population received ID cards after

being enumerated. A similar procedure had been followed in the first Israeli population census of 1948. The ICBS chose a more selected group of soldiers, with a better command of Arabic, to administer the more detailed long form to the sampled population; these soldiers received some 12 hours of training by the ICBS. "Enumeration Districts with an average of 55 households" were created and "the sample was a sample of enumeration districts (cluster sample) drawn systematically from geographical strata" (ICBS 1967-70, Vol. 3, p. xv). In the West Bank (excluding East Jerusalem) and Gaza Strip the goal was to sample 20 percent of the population; in East Jerusalem, 50 percent. The Introductions to volume 1 and volumes 2 and 3, respectively, describe the 100 percent and sample enumerations in more detail for the West Bank and Gaza Strip. The Introduction to volume 6 provides the same information for East Jerusalem (the long-form questionnaire was very slightly modified for that area). Copies of the census questionnaires, in English translation, may be found in the introductions. Apparently, for actual use in the enumeration, the short-form questionnaire was printed in Arabic and Hebrew; the long-form questionnaire only in Hebrew (or so the forms appear in the Hebrew versions of the introductions). The introduction to volume 1 also describes how the material was keypunched and then converted to computer tapes readable on an IBM 360/30 mainframe.

Creating the Digitized Dataset

Copies of the six volumes of ICBS reports were scanned to PDF files at the Harvard University Library. Then the process involved several additional stages.

- 1) I exploited a sophisticated optical character recognition (OCR) program, OmniPage Professional 17. The OCR program operated on each page of the scanned PDF images in order to create the first draft of the digitized Excel tables. In order to maximize the program's accuracy of recognition, only table cells were put through the OCR process; thus the program only had to recognize numerals. Row and column headings and table titles were manually entered later. The OCR program's capabilities were impressive; thus it focused on designated parts of a given page, treated the material therein as a table of numbers arranged in rows and columns, 'learned' from corrections, and converted the recognized material into Excel output. Nevertheless the challenges to the program were also impressive: the faded

- numbers, the rows and columns that were not quite straight on the page (and hence could not be correctly interpreted as cell entries of a given row), etc.
- 2) After exploiting the OCR program as fully as I could, I turned the material over to two Bard undergraduate research assistants who compared the OCR results in the Excel draft tables against the PDF page images. The first task was to be sure that the structure of the table was correct: that all rows and columns were included. The second task was to add the row and column headings table title. Then the RAs faced the difficult task of checking the digitized cell entries against the cell entries in the PDF image. A high proportion of the tables involved row and/or column totals. In these cases the RAs could exploit the Excel ‘sum’ function in order to recalculate totals and compare those to the totals found in the PDF for the original printed table. When they differed the entries that made up the total were checked one by one.
 - 3) Finally, the work of the two RAs was rechecked. First, two of the Levy Institute support staff compared the Excel tables to the PDF images in all six volumes. And finally, Bard student volunteers undertook an independent check of the RAs work in two of the volumes. By this last stage, the number of errors caught was very slight (and the errors were never systematic). Indeed, corrections to the original printed materials were of comparable frequency to those caught in the newly digitized version.

Table Notes

Markers referring to notes at the end of tables are found throughout the digitized tables. Each such marker is placed *in a separate narrow cell* to the right of the cell entry so marked. Had we left the marker in the same cell as the census result to which it refers, one of the major advantages of Excel would have been lost to the user: the ability to manipulate cells of the worksheet for calculations.

Markers for three *types of notes* appear in the digitized tables.

- 1) All table notes found in the original publications were copied. These were typically indicated by roman lowercase letters (e.g., a-f) and those have been preserved.
- 2) Errors found in the original are indicated by an x followed by a numeral (x1, x2, x3, etc.).
- 3) The ICBS introductions warned readers that “For lack of time, exact estimates of sampling errors have not yet been drawn up. Only very rough estimates were made to caution the reader in respect of estimates for which the sampling error might be considerable” (e.g., ICBS, 1967–70, Vol. 3, p. xxviii). In the published volumes, one or two sets of parentheses around the cell entry identify results subject to “high” and “very high” sampling error, respectively. In the digitized version, the note marker appears in a narrow adjacent cell, to the right of the cell entry. Cell entries subject to high and very high sampling error are marked by one and two asterisks, respectively.¹

The introductions to volumes 2–6 note further that cells without any special designation had a relative standard error—i.e., [standard error for the cell entry] / [cell entry]—below 30 percent. The criterion for distinguishing between the dangers of “high” and “very high” sampling error are given in volumes 3 and 6 as a relative standard error of 50 percent. For volumes 2 and 4, the situation is more complicated: no such criterion is printed. Instead, the discussion concerns the sample size of the denominator.² Users interested in pursuing these issues in detail should

¹ In the original, the explanation for the meaning of the warnings appears only in the introduction rather than in each table. For clarity’s sake, in the digitized tables the asterisks are treated as note markers for which the note appears at the end of every relevant table. In the note, the words in quotes, “subject to high [or very high] sampling error,” are copied from the ICBS introductions.

² In volume 5 the introduction simply refers to the introductions to volumes 2 and 3. But these two volumes differ as to the criterion for the difference between high and very high sampling error so that it is not clear which to follow for cell entries in volume 5. The reason for the difference in criterion between volumes 2 (and 4) and 3 (and 6) apparently has to do both with the complexity of the sampling design as it relates to the specific variables in relevant tables (including but not limited to whether the cell entry refers to households or persons). See especially the following pages of the introductions: volume 2, pp. xxviii–xxix; volume 3, p. xviii; volume 4, p. xxiv; volume 5, p. xiii; and volume 6, p. xxxiv. In each case, guidelines for confidence intervals immediately follow.

also notice the guidelines for calculating confidence intervals found in each volume.

The Geographic Area to Which the Table Refers

Most text tables in volumes 1–5 cover either the population in the West Bank or the population in the Gaza Strip. In the published version of volumes 1–5, only general headings inserted into the table of contents guide the reader to the fact that (for example) in volume 1, tables 2–5 and 6–9, cover the West Bank and Gaza Strip, respectively (while table 1 covers all the occupied territories). The table of contents to the digitized version preserves this information. However, to minimize confusion and the need for cross checking, each table indicates if the table refers to the West Bank or Gaza Strip only (directly below the volume and table number).

Volume 1 of the printed volumes appeared in 1967; it referred throughout to the “West Bank.” By contrast, the other volumes, which appeared between 1968 and 1970, referred to the area as “Judea and Samaria.” The Levy Institute’s digitized version uses only the former designation when inserting information on location. However, direct quotations (in the table titles or in row or column headings) were copied as found. Similarly, all six census volumes distinguish between The West Bank and the area defined by the municipality of East Jerusalem; the digitized version does likewise.³

Other Census Datasets: Digitized Tables vs. Public-use Samples

As already mentioned, there have been three machine-readable censuses for the West Bank and Gaza Strip from which detailed information is available: the 1967, 1997, and 2007 censuses—the first undertaken by the Israelis, the latter two by the Palestinian Authority. Tables from the latter two enumerations are available on the website of PCBS.⁴

Users should also appreciate the distinction between digitized *tables* and a digitized *individual-level dataset* made available to researchers. The former is a collection of tables like those published in a census report but now entered into a convenient format such as Microsoft

³ The Israeli government in the years 1967–70, when the census was conducted and the six report volumes published, had not formally annexed East Jerusalem. Nevertheless, it had defined municipal boundaries for that area, and created an administrative coordination between East and West Jerusalem.

⁴ There have also been a number of other surveys of the population, most notably by the Norwegian research group FAFO. See in particular Heiberg and Øvensen (1993).

Excel for easy manipulation by the user. The latter is *not* a report aggregating information; rather, it comprises the unaggregated responses that each individual in a representative sample of census respondents provided to each question. Since the samples are selected so as to include all members of households, information on household structure and well-being – even on the dwellings in which the households live – is also a staple of these datasets. Such datasets are commonly referred to as *census public-use samples*.

The size of these census public-use samples outstrips anything the social scientist is familiar with from other contexts; typically they include 1–5 percent of the people in large countries such as the United States and as much as 20 percent of the people in smaller countries, such as Israel or the Palestinian territories. For the intensive study of national populations – whether at one moment in time or across decades – these public-use datasets have proven invaluable.

The great value of the public-use census samples is found in the fact that the information has not been aggregated in any way; by contrast in census publications it always has been aggregated. Relying on the public-use census samples, researchers are free to explore any question they like, so long as the census asked a relevant question. No longer are inquiries limited to the tables selected by census officials – who are constrained by budgets, by the need to cover a little of everything, and (often enough) by some political sensitivities. Likewise researchers using public-use datasets need not limit attention to a variable's broadest categories (so often found in published tables); instead they may study a variable at its narrowest, most refined, level (e.g., geographic place, educational attainment, occupational distribution, fertility, etc). And crucially, researchers may chose to control for many variables at once (while exploiting refined categories of those variables) in an effort to understand outcomes with multivariate statistical techniques. Thus, fertility decline in city vs. village may be compared while controlling for the age, education, and religion of the mother and the father separately as well as for the household economic level and region of the country.

Public-use samples for the 1997 and 2007 PCBS censuses are available from the PCBS typically for a fee; these include 20 percent of the enumerated population (PCBS 2008). A version including 10 percent of the population is available free online at the University of Minnesota Population Center's project, IPUMS International, which archives such data from scores of countries (MPC 2011). Public-use samples are available for Israeli population censuses as well; these include residents of the State (Jews and Arabs) and Jewish residents on

settlements in the occupied territories (the public-use samples for the Israeli population censuses are also available from IPUMS International). However, the public-use sample for the census that is our subject here – the 1967 census of the West Bank, Gaza Strip, and East Jerusalem – is not as easily accessible. The ICBS recently recovered the computer tape, first produced for internal use in the 1960s, and prepared a version for computers of our own day. However, in most if not all countries, not all datasets are released to researchers without restrictions. The ICBS limits applicants to some datasets to researchers working in Israel who also have an affiliation with an institution of higher education there (ICBS 2011). Access to the public-use sample from the 1967 census is so limited.⁵

Users should therefore be aware that the actual dataset from which tables were first constructed has not been lost; it is to be hoped that in the future it will be made available to all. In the meantime, the value of the digitized tables on the Levy Institute website will be all the greater because it allows researchers to learn a great deal and sharpen the questions they would address to the public-use sample. Likewise, even those who can work with the public-use sample today are likely to find that first exploring the potential of the digitized tables will save them time and improve their own use of the public-use sample.

Acknowledgments

Bard student volunteers contributed more than 130 hours of work, part at an early stage, experimenting with methods of data entry, and part at the end in checking results; thanks to all and especially to Mujahed Sarsur who organized the volunteers. Wendy Thomas at the University of Minnesota Population Center kindly shared her knowledge of OCR programs and ran preliminary tests that allowed us to determine a viable strategy for the workflow. A subsidy from the Bard College President Leon Botstein enabled the hiring of the two careful research assistants (Liudmila Malyshava and Maya Perlmann). The Levy Institute contributed my own time and that of staff members Kathleen Mullaly and Jeannie Lam, who proofread tables; Barbara Ross, who copyedited everything else; Marie-Celeste Edwards, who brought the material onto the website; and Susan Howard, the Institute's administrator, who facilitated

⁵ Personal correspondence with the ICBS. I am grateful to both Shlomo Yitzhaki, who heads the ICBS, and Ari Paltiel for their patient replies to many queries. And I am gratified that my own formal and informal inquiries throughout the relevant Israeli bureaucracy and archives helped lead to the rediscovery of the relevant computer tape.

every stage. Finally, I am grateful to Levy Institute President Dimitri B. Papadimitriou, who recognized the value of the effort and supported it throughout.

Suggested Citation for Material Taken from This Site

For *tabular data*, the citation should include the volume and table number. For *the PDFs of the introductions* to the original reports, the citation should include the volume number and page(s) cited. For the *Dataset Documentation* the citation should include “Dataset Documentation.” In all three cases, the rest of the citation should be:

Joel Perlmann. *The 1967 Census of the West Bank and Gaza Strip: A Digitized Version*. Annandale-on-Hudson, N.Y.: Levy Economics Institute of Bard College. November 2011 – February 2012. [Digitized from: Israel Central Bureau of Statistics, *Census of Population and Housing, 1967 Conducted in the Areas Administered by the IDF*, Vols. 1–5 (1967–70), and *Census of Population and Housing: East Jerusalem, Parts 1 and 2* (1968–70). <http://www.levyinstitute.org/palestinian-census/.>]

Correspondence

Any correspondence should be addressed to perlmann@levy.org. Reports of any errors found in the work would be especially useful, so that corrections can be made.

References Cited

- Heiberg, M., and G. Øvensen, eds. 1993. *Palestinian Society in Gaza, the West Bank and Arab Jerusalem: A Survey of Living Conditions*. Oslo: FAFO.
- Israel Central Bureau of Statistics (ICBS). “Microdata for Research.” Accessed April 5, 2011, http://www.cbs.gov.il/mucpuf/muc_puf_e.htm.
- . 1967–70. *Census of Population and Housing, 1967 Conducted in the Areas Administered by the IDF*, Vols. 1–5 (1967–70) [vol. 1, *West Bank of the Jordan, Gaza Strip and Northern Sinai, Golan Heights: data from the full enumeration*; vol. 2, *Housing conditions, household equipment, welfare assistance, and farming in the administered areas*; vol. 3, *Demographic characteristics of the population in the administered areas: data from the sample enumeration*;

- vol. 4, *Labor force, part 1; vol. 5, Additional data from the sample enumeration*]. Jerusalem: ICBS.
- . 1968–70. *Census of Population and Housing: East Jerusalem [parts 1 and 2]*. Jerusalem: ICBS.
- Minnesota Population Center (MPC). 2011. *Integrated Public Use Microdata Series, International: Version 6.1* [Machine-readable database]. Minneapolis: University of Minnesota.
- Palestinian Central Bureau of Statistics (PCBS). 2008. *Population, Housing and Establishment Census, 2007: Census Final Results in The West Bank – Summary (Population and Housing)*. Ramallah: PCBS.
- . N.d. *Public Use Files [Population Housing and Establishment Census, 1997 and 2007]*. Accessed April 11, 2011, <http://www.pcbs.gov.ps/DesktopDefault.aspx?tabID=3538&lang=en>.