Fafo AIS

The 2009 Gaza Assessment Survey

Sample design

Post fieldwork version

Table of contents

Tables

Introduction

This document describes the sample of the 2009 Gaza Assessment Survey. Its main aim is to document the sampling procedures and the procedures for handling non-response in the survey.

Requirements of the sample

The design of the sample was – like any other sample – subject to a number of constraints. The main design characteristics for the sample were the following:

1. The population selected for this study was all Palestinian households and individuals in the Gaza Strip.
2. The budget allowed for a sample of around 2,000 households.
3. The survey contains two questionnaires: the main questionnaire and the Randomly Selected Individual (RSI) questionnaire. The main questionnaire was designed to collect data about the overall condition of the dwelling and infrastructure, the economic situation and assistance, food consumption during the Gaza war, and also to provide basic demographic and socio-economic information concerning each household member. One knowledgeable adult household member answered the main questionnaire. To complete the RSI questionnaire, which posed questions about opinions and attitudes, one household member should be selected randomly among all household members aged 18+.

The sample frame

The sample frame was designed according to data from the 1997 census, which was implemented by the Palestinian Central Bureau of Statistics (PCBS). The sample frame is a list of clusters, which are provided by the PCBS. The sample frame consists of 132 clusters selected from 15 strata, covering the whole Gaza Strip area. We used these clusters as PSUs in the first stage of the sampling process.

The sample frame is a list of sampled clusters in all the strata. It is organised as a file with the following variables:

Table 1: List of variables in the sample frame

Variable	Explanation
Individuals sample size	Number of households sampled in each stratum
Sample of enumeration areas	Number of clusters sampled in each stratum
Total enumeration areas	Total number of clusters in each stratum
Household	Total number of households in each stratum
Population size 2006	An estimation of population in each stratum in 2006
Stratum description	Urban, rural, refuge camp of 5 governorates
Stratum code	Strata code

The file contains a total of 132 clusters (enumeration areas) in 15 strata. The strata design contains five governorates (North of Gaza Strip, Gaza, Deir el-Balah, Khan

Yunis, and Rafah). Each governorate was split into three strata (urban, rural, and refugee camp).

Table 2: Distribution of the sample and frame:

Number of sampled hosueholds	Sampled enumeration areas	Total number of enumeration areas	Total number of households	Population size 2006	Stratum description	Stratum code
240	15	101	24,251	174,604	Urban north of Gaza	551
16	1	9	1,581	10,122	Rural north of Gaza	552
128	8	49	12,980	93,455	Refugee camps, north of Gaza	553
576	36	290	58,526	409,680	Urban Gaza City	601
16	1	7	1,343	8,864	Rural Gaza City	602
144	9	65	13,618	87,158	Refugee camps, Gaza City	603
96	6	40	9,060	65,234	Urban Deir al Balah	651
16	1	6	942	6,500	Rural Deir al Balah	652
208	13	82	20,145	136,983	Refugee camps, Deir al Balah	653
288	18	138	28,889	199,336	Urban Khan Yunis	701
48	3	30	4,751	31,356	Rural Khan Yunis	702
80	5	32	7,337	49,161	Refugee camps, Khan Yunis	703
96	6	47	10,143	71,003	Urban Rafah	751
32	2	13	2,455	15,960	Rural Rafah	752
128	8	55	12,232	84,400	Refugee camps, Rafah	753
2112	**132**	**964**	**208,253**	**1,443,816**	**Total**	

Sample design

The key elements of the sampling are the following:

1. PSUs were households in Gaza.
2. PSUs were explicitly stratified according to the strata description. Each governorate was split into three strata according to strata description (urban, rural, and camp). Therefore, five governorates were divided into 15 strata.

3. Cluster sample (a list of clusters) was provided by the PCBS, and was used as PSUs in the first stage of sampling.
4. The allocation of the cluster sample makes the sample approximately self weighting, so that the sample was proportional to the size of strata, i.e., the total number of households in each stratum (an estimation of population size in 2006, based on 1997 census).
5. All the clusters sampled were be re-listed.
6. From the updated list the selection of households was made with linear systematic sampling.
7. A predetermined number of households were selected from each sampled cluster in each stratum. This number was determined before re-listing and did not change after re-listing.
8. One household member was randomly selected from among all household members aged 18+ in each selected household, to answer the RSI questionnaire.

Sample selection procedures

Cluster sample

The selection of clusters was conducted by PCBS within each stratum in Gaza.

Re-listing of PSUs/houses

The sample frame was designed according to the data from 1997 census. Furthermore, after the Gaza war, many houses were destroyed in some area and many households moved. The selected clusters were re-listed to get the accurate total number of households in each cluster. The purposes of re-listing were two: first, to enable selection of households, and second, to enable interviewers to locate the selected households.

All households within a sampled cluster were listed.

Selection of households

The sample is considered as a sample of households. Linear systematic sampling was used to select households from the list of re-listed households. In each cluster, 16 households were selected.

Substitution

No substitution of selected PSUs/houses or households took place.

Random selection of an individual aged 18 or above within the household

The interviewer was responsible for selecting the RSI. The RSI selection was from a subset of household members aged 18 or above, and who stayed at least one day every week with the selected household. The random selection entailed two steps. The interviewer first listed and sorted all eligible household members by sex and age, i.e. listed the males first and then the females, the older first and then the younger. The second step consisted of the random selection from the pre-sorted list, with the help of random number table, i.e. a Kish table, attached to the questionnaire. The Kish table scheme is probably the most common way of selecting individuals at random within the households. We used the original Kish set of 8 tables, which is reproduced below:

		Number of eligible					
Proportions assigned	Table #	1	2	3	4	5	6+
1/6	1	1	1	1	1	1	1
1/12	2	1	1	1	1	2	2
1/12	3	1	1	1	2	2	2
1/6	4	1	1	2	2	3	3
1/6	5	1	2	2	3	4	4
1/12	6	1	2	3	3	3	5
1/12	7	1	2	3	4	5	5
1/6	8	1	2	3	4	5	6

Source: Kish 1965: 399

To use the table, the interviewer must know which table to use, and how many eligible members there are in the household. Thus, if table 4 is to be used, and there were six eligible members in the household, person number 3 was selected. If table 7 were to be used with six eligible members, person number 5 would be chosen.

Furthermore, as in the case of the original Kish table, the tables were allocated to the interviews in different proportions. Thus, in a sample of 1200 households, table 1, 4, 5, and 8 should each be allocated to 200 households, while table 2, 3, 6 and 7 should be used for 100 households each. Each questionnaire was marked with the table number to use.

Inclusion probabilities and weights

It follows from the above that the sample is a two stage sample.

Notation

In order to describe the sample precisely and calculate inclusion probabilities we need to introduce some notation. This is done in Table 3. In general the notation uses subscripts to indicate the sample stage, and superscripts to indicate the source of the data used. Thus $N_{h,c}$ means the population in stratum h, cluster c.

Table 3: Notation used

Symbol	Meaning
N	Household count (initial estimate)
N^{l}	Household count as listed
N	Number of households Uppercase: Total numbers in population Lowercase: Sample numbers
$N_{h,d}^{\geq 18}$	Number of eligible household members for selection of RSI, i.e. aged 18 or older and live at least one day per week with the household

Symbol	Meaning
m	Sample number of PSUs /houses
p	Inclusion probability
h	Index of stratum
c	Index of PSU
f and i	Index of household (f used to indicate household in the sampling stage, i used to indicate the list of all households from 1 to n in the sample)
d and r	Index of RSI (d used to indicate RSI in the sampling stage, r used to indicate the list of all eligible household members from 1 to N in the household)

Selection of PSUs

The inclusion probability for a cluster *c* in stratum *h* is the following.

Equation 1: Inclusion probability for cluster

$$p_{h,c} = \frac{N_{h,c} m_h}{N_h}$$

Equation 2*: Inclusion probability for household*

$$p_{h,c,f} = \frac{n_{h,c}}{N^{l}_{h,c}}$$

Note that the listed number of households is used, rather than the initial estimate of households from the census. The $n_{h,c}$ is pre-determined number of households to be selected in each PSU, which is 16 in each cluster.

The overall inclusion probability for a household then becomes:

Equation 3: Overall inclusion probability for household

$$p_i = p_{h,c} \cdot p_{h,c,f} = \frac{m_h N_{h,c} n_{h,c}}{N_h N^{l}_{h,c}}$$

Selection of RSIs

The inclusion probability for RSI *d* within the N adults (members 18+) of household *i* is:

Equation 4: Inclusion probability for RSI

$$p_d = \frac{1}{N^{\geq 18}_{i,d}}$$

Since only one RSI is selected.

The overall inclusion probability for a random selected individual then becomes:

Equation 5: Overall inclusion probability for RSI

$$p_r = p_i \cdot p_d = p_{h,c} \cdot p_{h,c,f} \cdot p_d = \frac{m_h N_{h,c} n_{h,c}}{N_h N_{h,c}^{l} N_{i,d}^{\geq 18}}$$

Sampling weights

There are two types of sampling weights. The expansion weights create estimates equivalent to real numbers in the population, while the relative weights retain the sample size and only adjust the relative contribution of each unit of analysis (household or individual). Only the expansion weights, which are the inverse of the sampling probability, are calculated in this survey.

Thus, the expansion sampling weight for household *i* is:

Equation 6

$$W_i^e = \frac{1}{p_i}$$

The expansion sampling weight for RSI *r* is:

Equation 7

$$W_r^e = \frac{1}{p_r}$$

The sampling weight for a household is then adjusted for the total number of households in each stratum, and the total number of households in Gaza. The sampling weight for the RSI is then adjusted for the estimated gender and age distribution of adults in Gaza in 2006, and the estimated overall population figure for Gaza in 2006.

Non-response and non-response corrections

The response rate achieved during the fieldwork of a survey is crucial for the quality of the survey results. When response rates are low, one may justifiably suspect biases in the results.

In general one can distinguish between two types of non-response: unit non-response and item non-response. Unit non-response pertains to the non-response of a whole unit, such as a household. In that case almost nothing is known about that household.

Item non-response pertains to the lack of information on a specific item for a unit, for instance that a person does not answer questions about income.

Here we will only consider unit non-response.

Unit non-response: the household

The results of interviews or attempted interviews can be studied using a classification of non-response in the questionnaire, derived from Hidiroglou, Drew and Gray (1993). The response categories in the framework are given in Table 4 and 5.

The framework is built around the observation that an interview can be missing for several reasons. First, it may be that the selected household does not belong to the sampling frame. This could be the case for instance for diplomats which were not considered eligible. Second, a selected dwelling is vacant; or a selected household, actually exists but could not be reached during the interview period. Further, a selected household may refuse, or not be found at home. Also, the classification has to take into account that there will be some situations where the interviewer cannot determine if a household exists or not, or cannot find the listed dwelling.

Table 4: Response categories and interview status for households

Category	Response type	Interview status
1 Interview completed	Interview is possible (response)	2004
2 Refusal converted (the respondent initially refused, but co-operated after a visit by the supervisor)	Interview is possible (response)	2
3 Partly completed	Interview is possible (non response)	14
4 Refusal	Interview is possible, non-response	25
5 Dwelling vacant	No interview possible	16
6 Dwelling under construction	No interview possible	0
7 Dwelling does not exist	No interview possible	18
8 Status not determined (the field work team could not find out if a household was living at the address)	Not clear, usually distributed over possible and not possible interview	1
9 Not eligible	No interview possible	1
10 No contact (the household exists, but could not be found at home)	Interview is possible, non-response	31

Table 5: RSI Response categories and interview status

1 Interview completed	Interview is possible (response)	1821
2 Refusal converted (the respondent initially refused, but co-operated after a visit by the supervisor)	Interview is possible (response)	11
3 Partly completed	Interview is possible (non response)	20
4 Refusal	Interview is possible, non-response	60
5 Not eligible	No interview possible	54
6 No contact (the household exists, but could not be found at home)	Interview is possible, non-response	52

A number of rates can be computed from the above framework as indicated by Table 6.

Table 6: Calculation of non-response rates in Gaza

Item	Calculated as	
Total- n	All households/dwelling units drawn in the sample	2112
Resolved – n	Total minus the units with indeterminate status (code 8)	2111
In scope – n	Resolved minus the not existing, not eligible or vacant units	2076
Completed interview –n	Interviews with at least first visit	2006
Completed RSI interview	Interviews completed with RSI questionnaires	1832
Resolved rate	Resolved/Total	99.95%
In scope rate	In scope/Resolved	98.3%
Non-existence rate	Non-existent units/resolved units	0.9%
Temporary out of scope rate	(Vacant + Not eligible) / resolved	0.8%
Response rate	Completed interviews/In scope	96.6%
Refusal rate	Refusals/In scope	1.2%
Refusal conversion rate	Refusals converted/(Refusals + converted)	7.4%
No contact rate	(Not determined + No contact)/ (Not determined + In scope)	1.5%
Non-response rate	(Not determined + Refusals + No contact)/(Not Determined + In scope)	2.7%
RSI in scope	Household in scope minus the not eligible	2022
RSI Response rate	Completed RSI interviews/In scope	90.1%
RSI Refusal rate	Refusals/In scope	4.2%
RSI Refusal conversion rate	Refusals converted/(Refusals + converted)	13.3%
RSI No contact rate	(Not determined + No contact)/ (Not determined + In scope)	4.2%
RSI Non-response rate	(Not determined + Refusals + No contact)/(Not Determined + In scope)	8.4%

The framework allows for showing both non-response due to imperfections in the sample frame and imperfections caused by problems during fieldwork.

References

Deming, W.E. 1960. *Sample Design in Business Research*, New York: Wiley.

Hidiroglou, M., J. Drew and G. Gray. 1993. ‘The measurement of non-response in surveys‘, *Survey Methodology*, Vol. 19: 81-94.

Kish, L. 1965. *Survey sampling*, New York: Wiley.