

2010 Gaza Strip Poll

Sample design

By Zhang Huafeng

Table of contents

Introduction.....	1
Requirements of the sample.....	1
The sample frame.....	1
Sample design.....	2
Sample selection procedures.....	2
Cluster sub-sample.....	2
Re-listing of PSUs/houses.....	3
Selection of households.....	3
Substitution.....	3
Random selection of an individual aged 15 or above within the household.....	3
Inclusion probabilities and weights.....	4
Notation.....	4
Selection of PSUs.....	5
Selection of RSIs.....	5
Sampling weights.....	6
Non-response and non-response corrections.....	6
Weight adjustment: population size.....	6
References.....	7

Tables

Table 1: List of variables in the sample frame.....	1
Table 2: Distribution of the sub-sample:.....	2
Table 3: Notation used.....	4

Introduction

This document describes the sample of Fafó's 2010 Gaza Strip Poll. Its main aim is to document the sampling procedures and the procedures for calculating the weight 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 living in Gaza.
2. The budget allowed for a sample of around 990 households.
3. The survey was designed to ask one household member in each selected household to answer the questions on economic status and the perception of various political issues. The household member to be interviewed should be selected randomly among all the household members aged 18+.

The sample frame

The sample frame was designed according to the data from the census in 1997, which was implemented by the Palestinian Central Bureau of Statistics (PCBS). The sample frame is a list of clusters, organized into 15 strata by PCBS. From that frame, a sample of 132 clusters selected from 15 strata, covering the whole Gaza area. Due to the small sample size allowed by the budget, we decided to sample 66 clusters from the 132 cluster sample we got from PCBS.

The sample frame from PCBS is a list of the sampled clusters. It is organised as a file with the following variables:

Table 1: List of variables in the sample frame

Variable	Explanation
Governorate code	Governorate, region
Governorate name	
Locality code	Locality code
Locality name	
Number of hh EA in 1997	Number of households in each cluster (Enumeration Area) in 1997
EA code	Code of clusters (Enumeration Area)
Strata code	Strata code
Strata description	Urban, Rural, camp of five governorates

In the sample from PCBS, there are 132 clusters (Enumeration Areas) in 15 strata. The strata design contains 5 governorates (North of Gaza Strip, Gaza, Deir el-Balah, Khan Yunis, Rafah). Each governorate was split into three strata according to strata description (urban, rural, and camp). In the final sub-sample, 66 clusters in 14 strata were selected.

Sample design

The key elements of the sampling are the following:

1. PSUs were households in the Gaza Strip.
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).
3. Cluster sample (a list of clusters) was provided by PCBS, and was used as PSUs in the first stage of sampling.
4. A sub-sample of 66 clusters was selected with PPS from the sample frame provided by PCBS.
5. In each stratum, 15 dwellings should be selected from each sampled cluster.
6. One household should be randomly selected from each selected dwelling.
7. One household member should be randomly selected from all the household members aged 18+ in each selected household, to answer the RSI questionnaire.

Sample selection procedures

Cluster sub-sample

The selection of clusters in the first stage was conducted by PCBS within each stratum in the Gaza Strip. Then a sub-sample of 66 clusters was selected out of 132 clusters provided by PCBS; PPS was applied when selecting the sub-sample.

Table 2: Distribution of the sub-sample:

Individuals sample size	Sample of enumeration areas	Total enumeration Areas	Household	Population size 2006	Stratum description	Stratum code
120	8	101	24,251	174,604	Urban, northern Gaza	551
15	1	9	1,581	10,122	Rural, northern Gaza	552
60	4	49	12,980	93,455	Refugee camp, northern Gaza	553
270	18	290	58,526	409,680	Urban, Gaza	601
60	4	65	13,618	87,158	Refugee camp, Gaza	603
45	3	40	9,060	65,234	Urban, Deir al Balah	651
15	1	6	942	6,500	Rural, Deir al Balah	652
105	7	82	20,145	136,983	Refugee camp, Deir al Balah	653

120	8	138	28,889	199,336	Urban, Khan Yunis	701
30	2	30	4,751	31,356	Rural, Khan Yunis	702
30	2	32	7,337	49,161	Refugee camp, Khan Yunis	703
45	3	47	10,143	71,003	Urban, Rafah	751
15	1	13	2,455	15,960	Rural, Rafah	752
60	4	55	12,232	84,400	Refugee camp, Rafah	753

Re-listing of PSUs/houses

The households living in all the 132 selected clusters in the first stage were relisted in 2009, and the number of re-listed households was used in the second stage of sub-sampling.

Selection of households

Substitution

No substitution of selected PSUs/houses or households was carried out.

Random selection of an individual aged 15 or above within the household

The interviewer is responsible for selection of the RSI. The RSI selection is from a subset of the household members aged 18 or above, and who live at least one day every week with the selected household. The random selection entails two steps. The interviewer should first list and sort all the eligible household members by sex and age, i.e. list male first and then female, the older first and then the younger. The second step is the random selection from a pre-sorted list, with the help of random number table, i.e. 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 will use the original Kish set of 8 table, which are reproduced below:

Proportions assigned	Table #	Number of eligible					
		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 are 6 eligible members in the household, person number 3 is selected. If table 7 is to be used with 6 eligible members person number 5 would be chosen.

Furthermore, in the case of the original Kish table, the tables are 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 must be marked with the table number to use.

Inclusion probabilities and weights

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
m	Sample number of PSUs /houses
p	Inclusion probability
h	Index of stratum
c	Index of cluster
s	Index of the sample in first stage
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 the first stage sample

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

Equation 2: Inclusion probability for the sub-sample

$$p_{s,c} = \frac{N_{h,c}^l m_s}{N_s}$$

Equation 3: Inclusion probability for household

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

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 same within each stratum, but different between different strata.

The overall inclusion probability for a household then becomes:

Equation 4: Overall inclusion probability for household

$$p_i = p_{h,c} \cdot p_{s,c} \cdot p_{h,c,f} = \frac{m_h m_s N_{h,c} n_{h,c}}{N_h N_s}$$

Selection of RSIs

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

Equation 5: Inclusion probability for RSI

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

Since only one RSI is selected.

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

Equation 6: Overall inclusion probability for RSI

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

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 7

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

The expansion sampling weight for RSI r is:

Equation 8

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

The sampling weights as such are not used in estimation of survey results, because the sampling weights are adjusted for the actual population size, as described below.

Non-response and non-response corrections

The non-response rate was around 6 percent in the survey; no non-response correction was made.

Weight adjustment: population size

The sample weight was adjusted based on the Gaza population from 2007 census. The age distribution of population size from 2007 census was projected using the Spectrum policy project software to predict the population size by age and gender in 2010, based on the assumption about fertility, mortality and migration in Occupied Palestinian Areas.

References

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

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

Lehtonen, R. and E. J. Pahkinen. 1995, *Practical methods for design and analysis of complex surveys*. Chicester: Wiley.

John Stover and Sharon Kirmeyer. 2008, *DemProj, A Computer Program for Making Population Projections*.