Fafo AIS

The 2009 West Bank Opinion Poll

Sample design

Post fieldwork version

Table of contents

Tables

Introduction

This document describes the sample of the 2009 West Bank Opinion Poll. Its main aim is to document the sampling procedures and the procedures for handling non-response.

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 the West Bank.
2. The budget allowed for a sample of around 1,500 households.
3. The survey was designed to ask one household member in each selected household various questions about political opinions, trust in institutions, political preferences, etc. The household member to be interviewed was to be selected randomly among all household members aged 18+.

The sample frame

The sample frame was designed according to the data from the 1997 census, which was implemented by the Palestinian Central Bureau of Statistics (PCBS). The sample frame is a list of buildings in the selected clusters, which are provided by the PCBS. The sample frame consists of 144 clusters selected from 31 strata, covering the whole West Bank area, and East Jerusalem. We use 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, camp of 11 governorates
Stratum code	Strata code

In total the file contains 144 clusters (enumeration areas) in 31 strata. The strata design contains 11 governorates (Jenin, Tubas, Tulkarm, Nablus, Qalqiliya, Salfit, Ramallah & Al Bireh, Jericho, Jerusalem, Bethlehem, and Hebron). Each governorate was split into three strata according to strata description (urban, rural, and refugee camp), except Qalqiliya and Salfit.

Table 2: Distribution of the sample and frame

Number of sampled households	Sampled enumeration areas	Total number of enumeration areas	Total number of households	Population size 2006	Stratum description	Stratum code
55	5	93	17,420	102,777	Urban Jenin	11
66	6	190	24,876	146,769	Rural Jenin	12
44	4	15	2,180	12,210	Refugee camps, Jenin	13
11	1	15	2,727	16,087	Urban Tubas	51
11	1	38	4,173	26,292	Rural Tubas	52
22	2	6	1,009	5,750	Refugee camps, Tubas	53
44	4	80	14,271	79,920	Urban Tulkarm	101
33	3	86	12,325	71,485	Rural Tulkarm	102
77	7	18	3,506	21,388	Refugee camps, Tulkarm	103
77	7	135	24,984	139,913	Urban Nablus	151
77	7	169	26,406	161,079	Rural Nablus	152
121	11	35	5,801	35,388	Refugee camps, Nablus	153
33	3	52	10,193	59,121	Urban Qalqiliya	201
22	2	49	5,900	38,351	Rural Qalqiliya	202
11	1	17	3,229	18,082	Urban Salfit	251
22	2	53	7,549	46,047	Rural Salfit	252
55	5	123	19,055	99,084	Urban Ramallah & Al Bireh	301
77	7	177	27,413	172,703	Rural Ramallah & Al Bireh	302
66	6	19	3,052	18,614	Refugee camps, Ramallah & Al Bireh	303
11	1	19	3,582	20,416	Urban Jericho	351
11	1	17	2,281	14,825	Rural Jericho	352
33	3	9	1,470	8,379	Refugee camps, Jericho	353
143	13	250	53760	301055	Urban Jerusalem	401
33	3	88	13025	80753	Rural Jerusalem	402

Number of sampled households	Sampled enumeration areas	Total number of enumeration areas	Total number of households	Population size 2006	Stratum description	Stratum code
33	3	21	4407	25282	Refugee camps, Jerusalem	403
33	3	73	12,401	62,004	Urban Bethlehem	451
44	4	127	16,205	103,709	Rural Bethlehem	452
55	5	17	2,619	14,404	Refugee camps, Bethlehem	453
154	14	301	55,120	363,791	Urban Hebron	501
66	6	218	23,386	163,704	Rural Hebron	502
44	4	14	2,359	15,097	Refugee camps, Hebron	503
1584	**144**	**2524**	**406,684**	**2,444,476**	**Total**	

Sample design

The key elements of the sampling are the following:

1. PSUs were dwellings in the West Bank.
2. PSUs were explicitly stratified according to the strata description. Each governorate was split into three strata: urban, rural, and camp.
3. Cluster sample (a list of sampled clusters) from each stratum 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 the strata, i.e., the total number of households in each stratum (an estimation of population size in 2006, based on 1997 census).
5. In each sampled cluster, 11 building structure were selected from a list of addresses of all dwellings in the cluster, provided by the PCBS.
6. One household was randomly selected from all households in each selected building structure. The selection was made by the interviewer with a random selection table.
7. One household member was randomly selected from among all household members aged 18+ in each selected household, to answer the questionnaire.

Sample selection procedures

Cluster sample

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

Re-listing of PSUs/houses

There was no re-listing of houses in the selected clusters.

Selection of dwellings

In each sampled cluster, 11 building structures were selected from a list of addresses provided by PCBS. Maps provided by the PCBS comprised of drawings and numbered structures in each cluster. The building structures were selected by linear systematic selection from the list of structures. As the maps provided by the PCBS were old, possible new dwellings in the selected clusters and not listed by the maps were not included in the sampling.

Selection of households

The interviewer was responsible for selecting the household in each building structure: Firs, the interviewer counted (and listed) all households in the building. Then the interviewer selected one household randomly by using a "random sheet", and by counting households from the top down, and from right to left.

Substitution

No substitution of selected PSUs, dwellings or households took place.

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

The interviewer was responsible for the selection of the RSI. The RSI selection occurred from a subset of the 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 the eligible household members by sex and age, i.e. listed males first and then females, and the older first and then the younger. The second step consisted of the random selection from a pre-sorted list, with the help of random number table, i.e. a so-called Kish table, attached to the questionnaire. The Kish table scheme is probably the most common way of selecting individuals at random within 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 are six

eligible members in the household, person number 3 is selected. If table 7 is 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 1,200 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

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/ dwelling count (initial estimate)
N^{l}	Household count reported by the interviewer in each selected dwelling
n	Sample number of households/ dwellings
$N_{i,a}^{\geq 18}$	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 PSU
d	Index of dwellings
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)
a and r	Index of RSI (‘a’ 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 dwelling*

$$p_{h,c,d} = \frac{n_{h,c}}{N_{h,c}}$$

The $N_{h,c}$ is total number of dwellings listed in the map of each PSU, provided by PCBS. The $n_{h,c}$ is the pre-determined number of dwellings to be selected in each PSU, which is 11 dwellings within each cluster.

Equation 3*: Inclusion probability for household*

$$p_{h,c,d,f} = \frac{1}{N^{l}_{h,c,d}}$$

The $N^{l}_{h,c,d}$ is the total number of households reported by the interviewer in each selected dwelling.

The overall inclusion probability for a household then becomes:

Equation 4: Overall inclusion probability for household

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

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

$$p_a = \frac{1}{N^{\geq 18}_{i,a}}$$

Since only one RSI is selected.

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

Equation 6: Overall inclusion probability for RSI

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

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

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 is 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 reach during the interview period. Third, a selected household, which actually exists and is eligible, 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

Category	Response type	Interview status
1 Interview completed	Interview is possible (response)	1392
2 Refusal converted (The respondent initially refused, but co-operated after a visit by the supervisor)	Interview is possible (response)	29
3 Partly completed	Interview is possible (non response)	30
4 Refusal	Interview is possible, non-response	51
5 Dwelling vacant	No interview possible	20
6 Dwelling under construction	No interview possible	0
7 Dwelling does not exist	No interview possible	2
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	0
9 Not eligible	No interview possible	19
10 No contact (the household exists, but could not be found at home)	Interview is possible, non-response	41

A number of rates can be computed from the above framework as indicated by Table 5.

Table 5: Calculation of non-response rates in the West Bank

Item	Calculated as	
Total- n	All households/dwelling units drawn in the sample	1584
Resolved – n	Total minus the units with indeterminate status (code 8)	1584
In scope – n	Resolved minus the not existing, not eligible or vacant units	1543
Completed interview –n	Interviews with at least first visit	1421
Resolved rate	Resolved/Total	100%
In scope rate	In scope/Resolved	97.4%
Non-existence rate	Non-existent units/resolved units	0.1%
Temporary out of scope rate	(Vacant + Not eligible) / resolved	2.5%
Response rate	Completed interviews/In scope	92.1%
Refusal rate	Refusals/In scope	3.3%
Refusal conversion rate	Refusals converted/(Refusals + converted)	36.3%
No contact rate	(Not determined + No contact)/ (Not determined + In scope)	2.7%
Non-response rate	(Not determined + Refusals + No contact)/(Not Determined + In scope)	6.0%

The framework allows for showing both non-response that is due to imperfections in the sample frame and imperfections that is due to problems during interviewing.

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.