



# **Palestinian Central Bureau of Statistics**

**The Palestinian Expenditure and Consumption Survey (PECS) 1998**

## **Data User Guide**

## Concepts and Definitions

<b>Household:</b>	One person or group of two or more persons with or without a family relationship who live in the same dwelling unit, who share meals and make joint provisions for food and other essentials of living.
<b>Head of household:</b>	The person who usually lives with the household and is recognized as head of household by its other members. Often, he/she is the main decision-maker and is responsible for financial support and welfare of the household.
<b>Expenditure:</b>	<p>Includes the following:</p> <ol style="list-style-type: none"><li>1. Cash spent on purchase of goods and services for living purposes.</li><li>2. The value of goods and services payments or part of payments received from the employer.</li><li>3. Cash expenditure spent as taxes (non-commercial or non-industrial), gifts, contributions, interests on debts and other non-consumption items.</li></ol>
<b>Consumption:</b>	<p>Includes the following:</p> <ul style="list-style-type: none"><li>. Cash spent on purchase of goods and services for living purposes.</li><li>. The value of goods and service payments or part of payments received from the employer.</li><li>. Own-produced goods and food, including consumed quantities during the recording period.</li><li>. Estimated rent value of the dwelling.</li></ul>
<b>Recording period:</b>	The temporal point during which data on household expenditure and consumption are recorded. It includes one month for record keeping, and twelve months for the last visit interview on durable goods.
<b>Dependency rate:</b>	The ratio of the persons defined as economically dependent to those in the ages defined as economically productive; arbitrarily defined as the ratio of the elderly and young (those 65 years and over plus those under 15 years of age) to the population of the “working age” (those 15-64 years of age).
<b>Economic dependency ratio:</b>	<p>It is number of economic dependent household members divided number of economic independent household members.</p> <p><i>Economic dependent:</i> household members under 15 years of age old and members 15 years of age or more who are not income earners. <i>Economic independent:</i> household members’ 15 years of age or more who are income earners.</p>
<b>Level of Living:</b>	Food Consumption divided by Total Consumption. This

indicator is based on the following assumption:

The Level of Living is identified by the proportion of consumption on food out of the total consumption (Angles Law of Poverty).

It is distributed to three categories:

- Better-off: "Food Consumption to Total Consumption less than 30%"
- Middle category: "Food Consumption to Total Consumption between 30-44%"
- Worse-off: "Food Consumption to Total Consumption between 45-100%"

**Governorates:**

Governorates are divided into the West Bank Governorates and Gaza Governorates. The West Bank includes North Governorates (Jenin, Tulkarm, Qalqilya and Nablus), Middle Governorates (Ramallah, Jerusalem and Jericho) and South Governorates (Hebron and Bethlehem). Gaza strip includes North Gaza and Gaza city, Middle Gaza and South Gaza.

**Composition of the Household:**

The relationship that connects the households together. It may be of the following nature: one individual, couple, parent/s with unmarried children, parent/s with unmarried and married children with grandchildren. A household constituting of one of the aforementioned cases living together with relatives, a group of Kinship related individuals living in one shelter and sharing living activities.

**Main Source of Income:**

The most regular and stable source of income, Sources of income are: agriculture and raising livestock, fishing and other similar activities, employment in the public or private sectors, social support, salaries/wages from working in the Israeli sector, transfers from inside the West Bank and Gaza, transfers and remittances from abroad, inheritance or pension, rent of property, etc.

## The Questionnaire

The PECS questionnaire consists of two main sections:

**First section:** Certain articles / provisions of the form filled at the beginning of the month, and the remainder filled out at the end of the month. The questionnaire includes the following provisions:

**Cover sheet:** It contains detailed and particulars of the family, date of visit, particular of the field/office work team, number/sex of the family members.

**Statement of the family members:** Contains social, economic and demographic particulars of the selected family.

**Statement of the long-lasting commodities and income generation activities:**

Includes a number of basic and indispensable items (i.e, Livestock, or agricultural lands).

**Housing Characteristics:** Includes information and data pertaining to the housing conditions, including type of shelter, number of rooms, ownership, rent, water, electricity supply, connection to the sewer system, source of cooking and heating fuel, and remoteness/proximity of the house to education and health facilities.

**Monthly and Annual Income:** Data pertaining to the income of the family is collected from different sources at the end of the registration / recording period.

**Second section:** The second section of the questionnaire includes a list of 54 consumption and expenditure groups itemized and serially numbered according to its importance to the family. Each of these groups contains important commodities. The number of commodities items in each for all groups stood at 707 commodities and services items. Groups 1-21 include food, drink, and cigarettes. Group 22 includes homemade commodities. Groups 23-45 include all items except for food, drink and cigarettes. Groups 50-54 include all of the long-lasting commodities. Data on each of these groups was collected over different intervals of time so as to reflect expenditure over a period of one full year.

## Data Sets Linkage

There are 3 data files in this CD-ROM, with one key variable in each file to allow the linkage between these files, namely: SERIAL variable (household number). The following table describes the files names, content and identification variables.

File Name	Content	Key Variables
ROSTER	Household's Roster	SERIAL: Unique household number
PRODUCTS	Products Data	SERIAL: Unique household number
IDENTIFICATION	Identification Data	SERIAL: Unique household number

## The Study Population

The target population in this sample survey comprises all private household living in the West Bank and Gaza strip, excluding nomads.

## Sample and Frame

The target population in this sample survey comprises all households living in the West Bank and Gaza Strip, excluding nomads and students.

The sample design is a stratified two-stage design for households selected to be interviewed. At the first stage a sample of *cells* (PSUs) was selected from the PCBS master sample frame. At the second stage, a sample of *households* was selected after a complete household listing of the sampled cells.

### Sample Design

#### Stratification

Four levels of stratification have been made:

1. Stratification by District.
2. Stratification by place of residence which comprises:  
(a) Municipalities                      (b) Villages                      (c) Refugee Camps
3. Stratification by locality size.
4. Stratification by cell identification in that order.

### Sample Size

The sample size is about 3,196 households allowing for non-response and related losses .

### Target cluster size:

The next important issue in the sample design is the target cluster size or “sample-take” which is the average number of households to be selected per PSU. In this survey, the sample take is around 10 households.

### Self-weighting design:

At the first stage, clusters or “cells” have been selected with PPS probability proportional to estimated measure of size ( $M_i$ ) for unit ( $I$ ):

$$f1_i = \frac{aM_i}{\sum M_i}$$

Where the summation covers all clusters in the population; a-300 is the total number of selected clusters. It is highly desirable for the PECS to have a constant overall sampling rate ( $f$ ), i.e. to have a self- weighting sample. This requires the second stage probability for the selection of households and persons within any sample cluster  $i$  to be as follows:

$$f2_i = \frac{f}{f1_i} = f \frac{(\sum M_i)}{a} \frac{1}{M_i} = \frac{(b)}{M_i}$$

Where  $b$  is a constant (independent of  $i$ ) to be determined to obtain the required sample size,  $n = 3,000$  households. Since the measure of size  $M_i$  are likely to differ from the actual number  $L_i$  of households listed in any cluster  $i$ , the actual number of households selected with the above  $f2_i$  shall vary from one cluster to another and are presented as:

$$b_i = f2_i * L_i = \frac{(L_i)}{M_i} * b$$

Adding all clusters in the sample results in the required constant  $b$ , to achieve the target sample size  $n$  as:

$$b = \frac{n}{\sum_a (L_i / M_i)}$$

Hence to control the overall sample size,  $b$  is determined after completing the listing in all sample areas.

The above procedure allows for variation in sample sizes  $b_i$  at the level of individual clusters, in order to provide a self-weighting sample. Households within each sample cluster shall be selected systematically from the lists prepared for that purpose, using the sampling interval,

$$I_i = \frac{1}{f2_i} = \frac{(L_i)}{b_i} = \frac{(M_i)}{b}$$

Where:

$a$	Number of cells in the sample (equals 300)
$M_i$	Number of housing units in cell $i$
$L_i$	Number of listed of households in cell $i$
$n$	Proposed sample size (n= 3,196 HHs)
$b$	Average sample take
$b_i$	Sample take in cell $i$
$f$	Sample rate
$f1_i$	First-stage sampling rate
$f2_i$	Second-stage sampling rate

Which is fixed for each cluster but varies between clusters depending on the measure of size ( $M_i$ ) with which the area was selected at the first stage.

The sample-take  $b_i$  must be allowed to vary depending on the actual number of households  $L_i$  found after listing. However, provision must be made to avoid extreme variation in cluster sample size. This could be done by using the above procedure to compute the ratio  $(b_i / b)$  for each cluster in the sample. If this ratio lies outside the range say 0.5 - 4.0, adjust  $b_i$ , i.e.

the interval  $L_i$  to be applied for the selection of households in the cluster, so as to keep the ratio within the above range.

### Sample Rotation:

The total number of (480) cells have been divided into (24) groups (subsided sample), each one constituted of (20) cells. A sub-sample of (300) cells is used year round by a monthly sample constituted of two minor samples (30 cells). The survey includes independent cells and not cross section ones, each of these is formed from (300) households for each month (round).

#### (Replication)

Month	A	B	C	D	E	F	G	H	I	J	K	L
1	×											
2		×										
3			×									
4				×								
5					×							
6						×						
7							×					
8								×				
9									×			
10										×		
11											×	
12												×

### Estimations Procedure

The sample is self-weighting by design. To estimate a given total  $Y$  for a given sub-population  $A$ , we introduce the following formula:

$$(1)_A \quad Y_A = \sum_h \sum_i \sum_j W_{hij} Y_{hij}$$

But since  $W$  is constant for all  $j$  within  $i$ , then: the estimating formula becomes:

$$(1)_B \quad \hat{Y}_A = \sum_h \sum_i W_{hi} Y_{Ahi}$$

Where,

$\hat{Y}_A$  = Estimated total for variable  $Y$  in sub-population  $A$

$h$  = The sub-stratum within the estimation domain

$i$  = The sample PSU (cell)

$j$  = The unit of analysis or element

$A$  = Subset of elements possessing a given attribute, that is, belonging to a given sub-population  $A$

$Y_{hij}$  = Observed value of variable “y” for  $j$ -the element of  $i$ -the sample PSU in stratum  $h$

$W_{hij}$  = Final (adjusted) sampling weight for the element

$y_{Ahi}$  is the unweighted PSU total within  $h$  for sub-population  $A$

The estimator for a given ratio for sub-population  $A$  is the following:

$$(2) \quad \hat{R}_A = \frac{\hat{Y}_A}{\hat{X}_A}$$

Where:

$\hat{R}_A$  = Estimate for the ratio of two variables,  $Y/X$ , in sub-population  $A$

$\hat{X}_A$  = Estimated total for variable  $X$  in sub-population  $A$ , given by formula (1)<sub>B</sub>

$\hat{Y}_A$  = Estimated total for variable  $Y$  in sub-population  $A$ , also given by formula (1)<sub>B</sub>

Means and proportions are special types of ratios. In the case of the mean, the variable  $X$ , in the denominator of the ratio, is defined to equal 1 for each element, so that denominator is the sum of weights in the sub-population.

In the case of the proportion, the variable  $X$  in the denominator is also defined to equal 1 for all elements. In addition, the variable  $Y$  in the numerator is binomial and is defined to equal either 0 or 1, depending on the absence or presence of a specified attribute in the element observed.

## Calculation of Variance

It is very important to calculate standard errors for the main survey estimates so that the user can have an idea of their reliability or precision.

The variance calculation will use the method of ultimate clusters. Within any domain of estimation, for a sub-population  $A$ , and for a characteristic  $Y$ , the formulas are:

(a) The variance of an estimator of a total is estimated by:

$$(3) \quad V\left(\hat{Y}_A\right) = \sum_h \left[ \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \hat{Y}_{Ahi} - \frac{\hat{Y}_{Ah}}{n_h} \right)^2 \right]$$

Where:

$$(4) \quad \hat{Y}_{Ahi} = \sum_{j \in A} W_{hij} Y_{hij}$$

and:

$$(5) \quad \hat{Y}_{Ah} = \sum_i \sum_{j \in A} W_{hij} Y_{hij}$$



The expression in (3) is an unbiased estimator of the variance.

(b) The variance of an estimator of a ratio is estimated by:

$$(6) \quad V(\hat{R}_A) = \frac{1}{\hat{X}_A^2} \left[ V(\hat{Y}_A) + \hat{R}_A^2 V(\hat{X}_A) - 2 \hat{R}_A COV(\hat{X}_A, \hat{Y}_A) \right]$$

Where:

$$COV(\hat{X}_A, \hat{Y}_A) = \sum_h \frac{Dom}{n_h - 1} \sum_{i=1}^{n_h} \left( X_{Ahi} - \frac{\hat{X}_{Ah}}{n_h} \right) \left( Y_{Ahi} - \frac{\hat{Y}_{Ah}}{n_h} \right)$$

$V(\hat{Y}_A)$  and  $V(\hat{X}_A)$  are calculated according to formula (3);

$\hat{X}_A$  is calculated according to formula (1); and

$\hat{R}_A$  according to formula (2).

## Reference Date

Twelve calendar months, between January and December 1998.

## Data collecting

### Pilot Survey

The pilot survey was designed for testing the questionnaire, fieldwork procedures and instructions, data entry template and instructions, and the estimated costs. PECSII carried out by PCBS in West Bank and Gaza Strip through the period "January-December 1997, was considered as a pilot survey for this survey (PECSIII). PECSII survey results were thoroughly evaluated, and modification in the plan, questionnaire, fieldwork procedures,..etc were made accordingly.

### Field Operations

Four teams of female interviewers, three in the West Bank and one in Gaza Strip carried out data collection. Each team consisted of a supervisor, a field editor, and 4-5 interviewers. Fieldwork teams were distributed to different districts according to sample allocation. All field staff received a training session combining general theoretical and practical components. Interviewers, supervisors and editors for the survey were selected from those who worked on PECS1 & PECSII. Fieldwork procedures and organization were designed to ensure adequate supervision and the collection of high quality data. To this end, several quality control measures were used though out fieldwork.

An interviewer undertook between 120 and 150 household visits in a month. The households were asked to keep daily records in a diary during their recording month.

### Response Rate

Selection of (3,196) households representing the West Bank and Gaza Strip, (2,218) households in the West Bank and (978) households in Gaza Strip.

After completing the last visit to the households during the survey period, The results of questionnaires were as follows:

Region	Number of Questionnaires			Number of incomplete Questionnaires			
	Total	Uninhabited	Completed	Refused	Travel	Movement	Other
West Bank	2,218	89	1,965	76	29	23	36
Gaza Strip	978	34	886	40	8	3	7
Palestinian Territory	3,196	123	2,851	116	37	26	43

Excluding the uninhabited housing units, the sample of survey become (3,073) households, (2,129) households in the West Bank, And (944) households in Gaza Strip. (2,851) households completed the questionnaire: (1,965) households in the West Bank and (886) households in Gaza Strip. Therefore the percentage of non response rate was as indicated below:

Result of Interview	Palestinian Territory (%)	West Bank (%)	Gaza Strip (%)
Non response	7.8	8.3	6.5
Refused	4.1	3.9	4.5
Not complete*	2.2	2.6	1.2
Other	1.5	1.8	0.8

\*Not complete include the movement and travel of Household.

## **Data Quality**

Sample surveys are exposed to two types of errors: statistical errors resulting from studying a given part of the society and from not including all the sections of the society. Since the households' expenditure and consumption survey is conducted by using a sample method, statistical errors are unavoidable. Therefore, a potential sample of suitable design has been employed whereby each unit of the society has a high chance of selection. Upon calculating the rate of bias in this survey, it appeared that the data is of high quality. The second type of errors is the non-statistical errors that relates to the design of the survey, mechanisms of data collection, management and analysis of data. Members of the work Commission were trained on all possible mechanisms to tackle such potential problems, as well as on how to address cases in which there were no responses (4.1%).