

Cambodia - Cambodia Socio-Economic Survey 1997

National Institute of Statistics

Report generated on: May 1, 2021

Visit our data catalog at: <http://nada.nis.gov.kh/index.php>

Overview

Identification

ID NUMBER
KHM-NIS-CSES-1997-v1

Version

VERSION DESCRIPTION
- v01: Edited and anonymized data.

PRODUCTION DATE
1997-06-10

Overview

ABSTRACT

The immediate objective of the Survey is the development of institutional capacity of the National Institute of Statistics (NIS) of the Ministry of Planning (MOP) to implement a demand driven multi-purpose living standards household survey based data collection system which produces regular, timely and relevant feed back to government policy makers. The project has provided technical assistance for the conduct of two large scale multi-objective national household surveys, the first one in 1997 and the second to be conducted in 1998/99. The primary objective of Cambodia Socio-Economic Survey (CSES) 1997 was to obtain data for the measurement of living standards in geographic stratification and different segments of the Cambodian society. The other objectives were to provide information needed by a variety of users such as government institutions, donor agencies, non- government organizations; to assist NIS to train its staff in planning, designing and conducting a household based survey system and institutionalize survey taking capability. The expansion of the scope of the survey to meet the data needs of a wide variety of users and thus minimize the duplication of household surveys and promote the acceptance of CSES as the national household survey programme was also an important objective.

KIND OF DATA
Sample survey data [ssd]

UNITS OF ANALYSIS
- Individual

- Household

Scope

NOTES

The objective of Cambodia Socio-Economic Survey (CSES) 1997 was to obtain data for the measurement of living standards in geographic stratification and different segments of the Cambodian society. The other objectives were to provide information needed by a variety of users such as government institutions, donor agencies, non- government organizations; to assist NIS to train its staff in planning, designing and conducting a household based survey system and institutionalize survey taking capability. The expansion of the scope of the survey to meet the data needs of a wide variety of users and thus minimize the duplication of household surveys and promote the acceptance of CSES as the national household survey programme was also an important objective

TOPICS

Topic	Vocabulary	URI
-------	------------	-----

Demographic Characteristics	CESSDA	http://www.nesstar.org/rdf/common
Education	CESSDA	http://www.nesstar.org/rdf/common
Economic Characteristics	CESSDA	http://www.nesstar.org/rdf/common
Health	CESSDA	http://www.nesstar.org/rdf/common
Housing and Environment	CESSDA	http://www.nesstar.org/rdf/common
Household Consumption Expenditure	CESSDA	http://www.nesstar.org/rdf/common
Household Assets and Liabilities	CESSDA	http://www.nesstar.org/rdf/common
Fertility, Mortality and Child Care	CESSDA	http://www.nesstar.org/rdf/common
Household's Access to Water, Firewood and Fodder	CESSDA	http://www.nesstar.org/rdf/common

Coverage

GEOGRAPHIC COVERAGE (1)

The sample was designed to provide estimates of the indicators at :

National (24 provinces)

GEOGRAPHIC COVERAGE (2)

Phnom Penh, Other Urban and Other Rural

GEOGRAPHIC COVERAGE (3)

Plain, Tonle Sap, Coastal, and Plateau/Mountain

GEOGRAPHIC UNIT

Group of Provinces

01. Banteay Meanchey

02. Battambang

03. Kampong Cham

04. Kampong Chhnang

05. Kampong Speu

06. Kampong Thom

07. Kampot

08. Kandal

09. Koh Kong

10. Kratie

11. Mondul Kiri

12. Phnom Penh

13. Preah Vihear

14. Prey Veng

15. Pursat
16. Ratanak Kiri
17. Siem Reap
18. Preah Sihanouk
19. Stung Treng
20. Svay Rieng
21. Takeo
22. Oddar Meanchey
23. Kep
24. Pailin

UNIVERSE

Select sample households from non-institutional households (All regular residents in Cambodia) in Cambodia.

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
National Institute of Statistics	NIS

OTHER PRODUCER(S)

Name	Affiliation	Role
World Bank		Project Executing Agency

FUNDING

Name	Abbreviation	Role
United Nations Development Program	UNDP	Funding
World Bank	WB	Funding
the Swedish International Development Cooperation Agency	SIDA	Funding

OTHER ACKNOWLEDGEMENTS

Name	Affiliation	Role
Mr. Nicholas Prescott	World Bank	Technical Adviser on Survey Design and Implementation
Mr. R. B. M. Korale	Senior Statistics Adviser	Technical Direction and Training Cambodian Statisticians

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
Lay Sophat	LS	NIS	Supervisor
Saint Lundy	SL	NIS	Quality Control

DATE OF METADATA PRODUCTION
2009-09-28

DDI DOCUMENT VERSION
Version 1.2 (July 2010). Revised version of original ddi document.

Version 1.1 (Feb 2010). Revised version of original ddi document.

Version 1.0 (Sep 2009). Original ddi document.

DDI DOCUMENT ID
KHM-NIS-CSES-1997-v11

Sampling

Sampling Procedure

A two stage stratified sampling design with the villages as the first stage units (PSU's) and households as the second stage units(SSU's) was used in the sampling strategy:

1. First Stage Selection

In the first stage the villages or primary sampling units (PSU's) were drawn from each domain. Within the three domains the villages were arranged by geographic codes with the villages grouped within communes and the communes within districts and districts within the provinces providing for some implicit stratification. The villages that had geographic codes also had the reported number of households based on the frame. The latter was used as the measure of size (MOS) in deriving the cumulated list for sampling. The sample villages were selected using the systematic sampling method with a random start with probability proportional to size method (PPS). The selection of sample villages was carried out through the use of a computer program.

2. The Second Stage Selection

For each selected village (PSU) a field listing was undertaken and let the actual number of households listed in the PSU be M_{hi}

*,

then the probability of selecting a household in the i th PSU in the h th domain is

$$ph(j/i) = nh / M_{hi}$$

where nh is equal to 10 in domains 1 and 2 and 15 for domain 3. Circular systematic random sampling with a random start was used to select households. The sampling interval would be equal to the current estimate of households in the PSU ascertained through the listing operation divided by 10 in the urban domains and 15 in the rural domain..

Please see Sample Selection in report or technical report of external resources

Deviations from Sample Design

The sampling design for the CSES 1997 considered several factors including the precision of data required by the users, the capacity of the national statistics office to conduct the survey, and most importantly the time constraint imposed to complete survey field work before the end of July 1997. Taking into account these factors, and specially the experience gained from the two socio-economic surveys conducted in 1993/94 and 1996, including estimates of feasible work loads, a sample of 6000 households to be selected from 474 villages was considered to be sufficient and manageable.

The design also took into consideration the need for separate analyses of three geographical domains, namely Phnom Penh, other urban areas aggregated together, and the rural area. In deciding the sample allocation to the three domains, it was decided that a size of around 1000 households would be adequate for the first two domains and the rest should be allocated to Domain 3 - Rural area, since it was envisaged that more detailed analysis of the poverty groups in this domain would be undertaken.

Response Rate

Despite the length of the questionnaire, the respondents had cooperated with the survey staff and provided answers to both questionnaires and it was possible to achieve a 100% response rate. At this stage it is not possible to comment on item non-response, and completeness of information provided by the respondents, and the respondent's fatigue arising from the length of the interviews which may have had a bearing on these issues.

Weighting

The estimates have been formed by weighting the data from the sample households to provide estimates that relate to all households in each domain. The weighting factors were calculated based on the probabilities of selection for the sample

The design weights are used to compensate for differences in the selection probabilities. The weight for the PSU is inversely proportional to its selection probability.

The probability of selection of j th household in normal size PSU's and blocks in the h th domain is

$$ph(i) \times ph(j/i) = ph(ij) \quad (\text{Eq. 3})$$

$$\text{where } ph(i) = ah \cdot Mhi / Mh$$

$$\text{and } ph(j/i) = nh / Mhi^*$$

Thus the design weights $whij$ for these units are

$$wh_{ij} = 1 / ph(ij)$$

$$M_h \times M_{hi}^*$$

$$= \frac{a_h \times M_{hi} \times n_h}{M_h \times M_{hi}^*} \quad (\text{Eq. 4})$$

$$a_h \times M_{hi} \times n_h$$

For the large PSU's which were segmented, the probability of selection of the jth household in the sth segment in the ith PSU in the hth domain is

$$ph(i) \times ph(s / i) \times ph(j / is) = ph(isj) \quad (\text{Eq. 5})$$

$$\text{where } ph(i) = a_h M_{hi} / M_h$$

$$ph(s / i) = 1 / s_i$$

$$\text{and } ph(j / is) = n_h / M_{his}^* \quad (\text{Eq. 6})$$

The design weight for such large PSU is

$$wh_{isj} = 1 / ph(isj)$$

$$M_h \times M_{his}^* \times s_i$$

$$= \frac{a_h \times M_{hi} \times n_h}{M_h \times M_{his}^* \times s_i} \quad (\text{Eq. 7})$$

$$a_h \times M_{hi} \times n_h$$

The design for CSES is not self weighting and therefore it is necessary to compute weight for each PSU, block or segment selected in the sample and these weights have to be used in the estimation procedure.

Questionnaires

Overview

The CSES 1997 questionnaire comprises 4 forms, namely:

Form 1: Listign of Households in the Village

Form 2: Village Questionnaire

Form 3: Core Questionnaire for Households

Form 4: Social Sector Household Module

Data Collection

Data Collection Dates

Start	End	Cycle
1997-06-01	1997-06-30	N/A

Data Collection Mode

Face-to-face [f2f]

Data Collection Notes

Each interviewer was assigned selected villages based on the sampling procedure. In order to complete the data collection activity within the planned time frame, each enumerator was assigned about 30/ 45 households in three or four villages. The questionnaires were filled by the method of personal interview.

A pre-listing of households was undertaken by the enumerator to generate the current list of households, which was essential to select the sample households based on the systematic sampling procedure. In addition to preparing a current list of buildings, housing units and households certain additional information such as the number of household members, principal economic activity of the household was also collected.

After the selection of sample households, the selected households were revisited to interview one or more responsible members of the household to fill in the core and social sector questionnaires. Before or after the household interviews, the enumerator interviewed the head of the village and other key informants to canvass information for the village questionnaire.

The field control procedures provided for the supervisors to inspect and make on the spot checks while the interview was being conducted and they were also required to re-interview a sub-sample of the households already interviewed by the enumerators under his supervision. To ensure effective supervision through inspections and re interviews, adequate funds were allocated for the payment of honoraria to supervisors for their supervisory duties. Some of the core group staff functioned as area coordinators and they were in over all charge of supervision as well as the coordination of the areas assigned to them. There was also a visit of the Minister of Planning and the Under Secretary of State MOP, Project Staff and Senior NIS Staff in Mid June 1997 to encourage the field staff and to study the operational issues and problems encountered in field work.

Despite the length of the questionnaire, the respondents cooperated with the survey staff and provided answers to both questionnaires and it was possible to achieve a 100% response rate. At this stage it is not possible to comment on item non-response, and completeness of information provided by the respondents, and the respondent's fatigue arising from the length of the interviews which may have had a bearing on these issues.

Questionnaires

The CSES 1997 questionnaire comprises 4 forms, namely:

Form 1: Listign of Households in the Village

Form 2: Village Questionnaire

Form 3: Core Questionnaire for Households

Form 4: Social Sector Household Module

Data Collectors

Name	Abbreviation	Affiliation
National Institute of Statistics	NIS	Ministry of Planning

Supervision

The supervisor is responsible for

- (i) administering the Village Questionnaires (Form 2),
- (ii) preparing the two Household Questionnaires for each village (for example, completing certain information on the Cover Page of each questionnaire, as described in this manual),
- (iii) checking all completed questionnaires to ensure that they have been filled up completely and well, and
- (iv) for making random visits to households that have been interviewed by interviewer to make sure that the answers are consistent with the completed questionnaire.
- (v) The supervisor is also expected to occasionally observe interviewers while they are conducting household interviews, especially during the first one or two weeks of the field work.

The district-level supervisor is responsible for checking the village questionnaires and for monitoring the survey's overall progress in those villages.

Data Processing

Data Editing

All completed questionnaires were brought to NIS for processing. Although completed questionnaires were checked and edited by supervisors in the field, specially because of the length of questionnaires and the complexity of the topics covered the need for manual editing and coding by trained staff was accepted as an essential priority activity to produce a cleaned data file without delay. In all, 39 staff comprising 35 processing staff and 4 supervisors were trained for three days by the project staff. An instruction manual for manual editing and coding was prepared and translated into Khmer for the guidance of processing staff. Manual processing of questionnaires commenced in mid August 1997.

In order to produce an unedited data file, keying in the data as recorded by field enumerators and supervisors, (without subjecting data to manual edit as required by the Analysis Component Project staff), it was necessary to structure manual editing as a two-phase operation. Thus in the first phase, the processing staff coded the questions such as those on migration, industry, and occupation which required coding. Editing was restricted to selected structural edits and some error corrections. These edits were restricted to checking the completeness and consistency of responses, legibility, and totaling of selected questions. Error corrections were made without canceling or obliterating the original entry made by the enumerator, by inserting the correction close to the original entry.

Much of the manual editing was carried out in the second phase, after key entry and one hundred percent verification and extraction of error print outs. A wide range of errors had to be corrected which was expected in view of the complexity of the survey and the skill background of the enumeration and processing staff. The manual edits involved the correction of errors arising from incorrect key entry, in-correct/ failure to include identification, miss-coding of answers, failure to follow skip patterns, misinterpretation of measures, range errors, and other consistency errors.

Other Processing

An in-house survey processing centre was established at the NIS to process the CSES 1997. A net work of 12 PCs with 2 high capacity PCs as servers was installed and NIS staff were trained to use the network system. The network can be strengthened with additional workstations to process a survey sampling of 15,000 households referred to in the project document.

Entire data processing was done on microcomputers and data entry and editing was carried out using Integrated Micro-Computer Processing System(IMPS) package developed by the US Bureau of the Census. Statistical Package for Social Sciences (SPSS) was used to obtain tabulations.

At the end of August 1997, the keyers and verifiers were trained for three days and key entry operations commenced. In all 30 key entry and verification staff and 3 supervisors were trained by the Data Processing Specialist to use the data entry screens prepared using IMPS software.

Four data entry systems were created to input the data from the four questionnaires. The data entry system for the listing form contains one record type with a maximum length of 49. The system for the village questionnaire contains 15 record types with a maximum record length of 105. The system designed for the core questionnaire contains 17 record types with a maximum record length of 116. The data entry system designed for the social sector module contains 12 record types with a maximum record length of 94. After keying in the data one hundred percent verification was done on all card types. In spite

of this safeguard to minimize errors it was found that verifiers had not only failed to detect errors but had introduced errors during verification. The set of consistency edit checks prepared for the survey when applied for a sample of three villages, the error printouts were so voluminous that it was decided to clean the files in stages, selecting a single record, question or a topic at a time. The first computer edit was applied to check the basic structure of the data and to check the skipping patterns. The errors were corrected manually and the data file was updated using IMPS programs. After completing the structural edit, the data file was re-edited for validity of records. Consistency edits were designed to detect responses that appeared to be inconsistent with other responses or in conflict with definitions and processing rules. It was necessary to run several edit checks to clean some data items. For tabulation several sub-master files were created for most data items. The inflation factors that should be assigned to each village were applied to the data at the tabulation stage.

Data Appraisal

Estimates of Sampling Error

The results obtained from the survey are subject to sampling errors. Sampling errors in surveys occur as a result of limiting the survey observations to a subset rather than the whole population. These errors are related to the sample size selected and sampling design adopted in the survey. In order to maintain these errors within acceptable levels, the efficient sampling design with the sample allocation described earlier was adopted.

In addition to sampling errors, the estimates are also subject to non-sampling errors that arise in different stages of any survey operation. These include

- errors that are introduced at the preparatory stage
- errors committed during data collection including those committed by interviewers and respondents
- processing errors

The first item includes errors arising from questionnaire design, preparation of definitions and instructions, preparation of table formats etc. The other two categories are clear from the terminology used. The use of trained enumerators and processing staff and careful organization and thorough supervision are essential to control and minimize these errors.

As already referred to, it was possible to obtain responses from all the villages and

households that were sampled, and thus it was not necessary to adjust the data for non-response. Thus the bias that is introduced into the estimates as a result of non-response was avoided.

The standard error of a survey estimate provides a measure of how far the survey estimate is likely to vary from the true population value (i.e. parameter) as a result of having collected the data on a sample basis rather through a complete census. The standard error $se(r)$ of a survey estimate is by definition

$$se(r) = \sqrt{\text{var}(r)}$$

The relative standard error or coefficient of variation (cv), on the other hand provides a measure of the relative variance of a survey estimate; that is the magnitude of the estimated sampling error relative to the magnitude of the estimate itself. The cv that is expressed as a proportional error enables the data user to compare the relative reliability or precision with which different types of survey characteristics have been measured eg. Means versus proportions, where direct comparisons of standard errors are uninformative since the magnitude of the standard error is dependent upon the magnitude of the estimate

Other forms of Data Appraisal

The results provide estimates at the level of the three domains Phnom Penh, other urban areas, and the rural sector into which the entire geographical area covered by the survey was divided. The survey design has provided for statistically reliable estimates for most characteristics at these levels of stratification.

The expenditure data from CSES 1997 presented here are not strictly comparable with the data from the SESC 1993/94, which canvassed very detailed data on consumer expenditure. SESC 1993/94 collected data on over 450 items of consumption expenditure, the type of information required to establish weights in the construction of consumer price indices. At that level of disaggregation it is possible to achieve results closer to actual consumption levels. Such surveys are required infrequently once in 5 -7 years because of costs and time involved in designing, conducting and processing such surveys. CSES 1997 had used a shorter list comprising 33 commonly used consumer items that were considered to be adequate to monitor consumption expenditure over time. In addition to this issue arising from differences in the scope of the two surveys, the researchers should take note of the decline in household size and changes in household structure which are important determinants of household expenditure.