Summary Quality Report as Away for Making Statistical Data Accessible and More Useful for all Users

Paper Submitted to Participate at European Conference on Quality in Official Statistics (Vienna,2-5 June 2014)
Waleed Mohamed
Statistician at National Statistics Office at Egypt (CAPMAS)
Email: kant2012xp@hotmail .com
Pres_capmas@capmas.gov.eg

Abstract

Summary quality report   provides users with both quantitative and qualitative information based on European standards. The reports also contain a summary of methods and explain areas where caution in interpretation may be required, structured around the European Dimensions of Quality. The objective of the reports is to help users understand our statistics and to use them appropriately.

There are two types of quality reports: (1) Summary Quality Report (SQR): static quality information for all releases, to be updated when there are changes in methods used.(2) Basic Quality Information (BQI): dynamic, release-specific quality information updated for each output.

The national statistics at Egypt (CAPMAS) considered SQR is very important because such tools not only help to improve the quality of our outputs by reporting the measurement of quality variables but also make figures transparent and easily accessible to the public; So this paper aim to show the Importance of Quality Reporting, purpose of SQR and Example for "Summary Quality Report for the Income, expenditure& Consumption Survey 2008/2009" at national statistics office at Egypt(CAPMAS).

Keywords: accessibility; the European Dimensions of Quality; Documentation; User Satisfaction.

1. Introduction

The national statistical office at Egypt (CAPMAS) is committed to providing its users with information about the methods used to produce the statistics and the quality of its data. This is consistent with the requirements of the Code of Practice for Official Statistics, which states the requirement to ensure that 'users are informed about the quality of statistical outputs'. The SQR is considered A valuable tool that ensures that processes and statistical results meet user priorities' and needs ,besides that it give them a clear picture on the quality of the data they use.

This paper will focus on SQR: definition, Purpose SQR, efforts of CAPMAS in data quality development as well as Example about "SQR for the Household Income, expenditure& Consumption Survey (HIEC).

2- Definition of SQR

A brief document describing the survey or statistical compilation quality in a product-oriented way, following the six Eurostat criteria of quality (relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence) and addresses relevant Key Quality Measures. Besides that it also provide us with a summary of methods used to compile outputs and describe the documentation of metadata for the statistical survey.

3- Purpose of the SQR

The purpose of the SQR is to give a summary for Users so that they can judge suitable uses of the data, and whether the data are suitable for their intended uses. In essence, this consists of static information, where the content will not change between outputs, except where there are changes in methods used to compile the output.

4- The summary Quality Report (SQR)

The SQR consists of two parts.

4.1 The first section: is a Summary of Quality, structured around the ESS dimensions of quality, pulling together the qualitative information for the output about each dimension.

4.2 The second section: is a summary description of the main processes used to compile the output and (Documentation (IECS) based on DDI &DCMI standards) besides that A brief description of the sampling design, including a description of the sampling frame and how it was constructed.

5- Efforts of (CAPMAS) at in data quality development

The National statistical office at Egypt (CAPMAS) in 1989 established the General Administration for statistical quality to assess and monitor the quality of data in all phases of the statistical work. This section is funded by the United Nations and this continued until the mid-nineties and then stopped work for several years. In 2004, the administration was reactivated again with full support of senior management.

- **Tasks of the quality control Department at CAPAMS can be summarized as follows:**

  (1) For field work, the quality control Department emphasizes on the quality of the stages of the statistical work for any statistical research produced by CAPMAS, as well as all phases of the work of the General Census of Population and Housing.

  (2) Making evaluation in phase Accounting and numbering of roads and blocks by high expertise's who write reports about any wrong finding.

(3) Using CALL CENTER to follow-up the complaints of Households that has not visited by field researcher.

(4) Withdrawing a random sample of size 5% and re-collecting data again to check the consistency with data collected in the first time.

(5) Making evaluation for field researchers' during training.

(6) Writing summary quality report by ending the work and publishing the results.

6- SQR for the Egyptian Household Income, expenditure& Consumption Survey 2008/2009" .

The Household Income, Expenditure and Consumption Survey (HIECS) is of great importance among other household surveys conducted by statistical agencies in various countries around the world since it provides us with a large amount of data that can used it in measuring the standards of living for households and individuals, as well as establishing databases that serve in measuring poverty, designing social assistance programs, and providing necessary weights to compile consumer price indices, considered to be an important indicator to assess inflation. In Egypt this survey has been conducted each 5 years until 2008 /2009 started to be conducted each 2 years.

• Objectives of the Household Income, expenditure& Consumption Survey (HIECS):

1 – To identify the average of consumer spending according to different characteristics of the population such Social, economic and demographic...etc

2- To measure average household and per-capita expenditure for various expenditure items along with socio-economic correlates.

3- To define percentage distribution of expenditure for various items used in compiling consumer price indices which is considered important indicator for measuring inflation.

4- To estimate the annual consumption, values of commodities and services consumed by households during the survey period to determine the levels of consumption and estimate the current demand which is important to predict future demands.

5- To study the relationships between demographic, geographical, housing characteristics of households and their income.

6- To Identify the average per capita income of the family according to economic activity and the Occupation of the head of the family.

7- To provide data necessary for national accounts especially in compiling inputs and outputs tables.

7- Summary Quality Report for HIEC survey

In this subsection we present and explain the quality elements of HIEC survey. These quality elements include :

7.1 Relevance

Relevance means the degree to which the statistical outputs meet users' needs.

The Household Income, expenditure& Consumption Survey (HIECS) meets the needs of a range of users. These are summarized below:

- The Data which obtained from (HIECS) survey is considered a major source for estimates of Household Expenditure in the Egypt National Accounts.

- Users outside government include independent research institutes, academic researchers, and business and market researchers.

- The main results from the survey are published by CAPMAS in the annual report at website and the dataset is available at web site.

- The most important users for income and spending survey " the Ministry of Finance , the Ministry of Insurance and Social Affairs , the Ministry of Housing, Utilities and Urban Communities , the Ministry of Health and Population , the Ministry of Electricity and Energy, the Ministry of Supply and Internal Trade, Ministry of Local Development .

- Issuing the Human Development Report for Egypt in 2010 and poverty index in Egypt, which depends on three dimensions (education, health, standard of living).

7.2 Accuracy

Accuracy means the closeness between an estimated result and the (unknown) true value.

This standard describes all the procedures that followed to ensure the accuracy and quality of data such as:

- The field staff was selected from among the efficient experienced persons working in CAPMAS and new graduates specially females.

- Intensive training program for supervisors was conducted at CAPMAS in Cairo and locally in governorates for interviewers and field editors.

- Supervision program was implemented (each 15 days) in all governorates to check the field work to overcome the field problems.

- Data were collected by using personal interview method for household in dwelling and it had been obtained from the head of household or wife or any eligible person in case of their absence.

- Cleaning and harmonizing raw data by Economic Research Forum.

- Estimates of Sampling Error using the Ultimate Cluster Method as applied in the CENVAR Module of the Integrated Microcomputer Processing System (IMPS) Package. In addition to the estimate of sampling error, the output includes estimates of coefficient of variation, design effect (DEFF) and 95% confidence intervals.

- It was put into consideration during the survey implementation to assign the quality control general division a core role in controlling the quality of the fieldwork to ensure data accuracy and avoid any errors in suitable time, as well as taking all the necessary measures to guarantee that mistakes are not repeated, with the application of the principle of reward and punishment, and announce the results to all those working in the survey.

7.3 Timeliness and punctuality

Timeliness: refers to the lapse of time between publication and the period to which the data refer.

Punctuality: refers to the gap between planned and actual publication dates.

The period of income and expenditure and consumption 2008/2009 extend over the 12 months starting from the month of April 2008 and end in the month of March 2009. Households are observed only two weeks, this is to collect information on food expenditure, the total research period Was divided to 4 consecutive quarters, the first quarter includes the first three months (April, May and June); the second quarter includes following three months and so on until the fourth quarter.

The reference period over which data was collected varies according to the type of data item as follows:

- **15 days**: for expenditure on food and beverages.
- **Month**: for expenditure on alcoholic beverages, tobacco and narcotics, housing and its accessories goods and services for routine household maintenance, health, operation of personal transport equipments, transport services, restaurants and hotels, personal care and other services.
- **Quarterly**: for expenditure on health, Communication.
- **Annually:** for expenditure on clothing and footwear, housing and its accessories, furnishings, household equipments and routine maintenance of the house, health, transport, communication, recreation and culture, education, restaurants and hotels and miscellaneous goods and services.

It is worth noting that in some cases the groups of commodities or services include more than one period such as health which has monthly, quarterly and annually items, on which the expenditure ends by the end of the survey period, depending upon the consumption frequency of these items.

The collection period was shortened to lighten the burden on the surveyed households and encourage them for more cooperation.

7.4-Accessibility and Clarity

Accessibility is the ease with which users are able to access the data , also reflecting the format(s) in which the data are available and the availability of supporting information.

Clarity refers to the quality and sufficiency of the metadata, illustrations and accompanying advice.

Income and expenditure and consumption results are published as a number of standard web tables in Excel format, and these tables are available at CAPMAS website**www.capmas.gov.eg**., and for metadata in the URL :http://www.censusinfo.capmas.gov.eg/Metadata-ar-v4.2/index.php/ catalog/ 84/study-description#page=data_collection&tab=study-desc

7.5 –Comparability and Coherence :

Comparability is the degree to which data can be compared over time and domain, such as geographic level. Coherence is the similarity degree of data derived from different sources or methods, and refers to the same phenomenon. Several considerations have been taken into account to render **Comparability and Coherence** for HIECS such as:

1- Using the 2007 Administrative Classification in coding statistical data geographically according to governorates, kism, Marks and villages as well as the new cities, established inside all governorates.

2- Use the latest classification (individual consumption according to the purpose COICOP) It is used for analysis at national level; especially for purposes of compiling consumer price indices. It consists of 12 divisions and 47 groups.

3- Followed the International Standard Industry Classification - Revision 4 (ISIC4) in the coding process at the level of four digits. Hence classification of industry cannot be made lower than the four-digit level. In this respect, the industry of persons working outside establishments has been deduced from their occupation.

4- The occupational classification for 2005 which is derived from the 1988 International Standard Classification of Occupation (ISCO) is used in coding occupations into six digits after being slightly modified to better suit Egypt's conditions.

5- Previous coding for all items on the tables of expenditure and consumption as well as answers the questions for characteristics of household members except the questions of economic activity and profession these questions were coding according to the guide of the profession and activity.

6- Moreover code data of location on the cover of the questionnaire according to guide of administrative units .In addition, all the data have the previous coding to avoid coding errors at the stage of automated data processing, There is no doubt that all above procedures foster confidence and consistency of data .

8- <u>Summary of main processes used to compile the output & Documentation metadata (HIECS)</u>.
   <u>In this section we discuss:</u>

   - **<u>Summary of Methods Used to Compile the Output</u>**

Target group	Household
Frequency	Income and expenditures survey is run every two years after it was run every 5 years and the cost is about 2.5 million pounds.
Sample size	The sample is about 48 658 households (2000 family almost Interviewed every fifteen days) distributed 47.2% urban and 52.8% Rural. The HIECS sample is approximately self weighting at national level and strictly self-weighting at the governorate level. It should be easy to attach a weight to each sample household record in the computer files, and the tabulation programs can weight the data automatically. The sampling probabilities at each stage of selection will be maintained in an Excel spreadsheet so that the overall probability and corresponding weight can be calculated for each sample cluster. The selection probabilities of urban and rural samples are presented as follow: **(1) <u>First stage sample for each governorate (urban/rural substrata)</u>** $$p_{\alpha} = \frac{\lambda M_{\alpha}}{\sum\limits_{\alpha} M_{\alpha}} \text{ , where}$$ P_{α} = Probability of selection of the αth PSU (EA) in the sample

	M_α = Number of census households of the αth PSU. λ = Number of PSUs to be selected from the substratun **(2) Second stage sample** $$P_{\beta\|\alpha} = \frac{19(20)}{M_\alpha}$$, Where, $p_{\beta\|\alpha}$ = the conditiona l probabilit y of selecting the βth household given that the αth PSU was selected Thus the overall sampling probability is $P_{\alpha\beta} = P_\alpha \times P_{\beta\|\alpha}$ The weight, for all the households in a given cluster, is equal to the reciprocal of the overall sampling probability stated above. • **Sampling Error Estimation** The sampling error of major survey estimates has been derived using the Ultimate Cluster Method as applied in the CENVAR Module of the Integrated Microcomputer Processing System (IMPS) Package. In addition to the estimate of sampling error, the output includes estimates of coefficient of variation, design effect (deff) and 95% confidence intervals.
Sampling frame	Due to recent census data (the reference point for the 2006 census is 21/22 November 2006), and in response to cost-benefit considerations, it was decided not to update the frame in the final stage for preview.

Procedures of quality outputs	• Applying the recent international recommendations of different concepts and definitions of income and expenditure considering maintaining the consistency with the previous surveys in order to compare and study the changes in pertinent indicators.
	• Evaluating the quality of data in all different Implementation stages to avoid or minimize errors to the lowest extent possible through:
	  - Implementing field editing after finishing data collection for households in governorates to avoid any errors.
	  - Setting up a program for the Survey Technical Committee Members and survey staff for visiting field work in all governorates (each 15 days) to solve any problem in time.

	- Re-interviewing a sample of households by Quality Control Department and examining the differences with the original responses. - For the purpose of quality assurance, tables were constructed for each survey round where internal consistency checks were performed to study the plausibility of mean household expenditure on major expenditure commodity groups and its variability over major geographic regions.
Data Editing	• **Office Editing:** Office editing is one of the main stages of the survey. It started as soon as the questionnaires were received from the field and accomplished by selected work groups. It includes: a- Editing of coverage and completeness b- Editing of consistency c- Arithmetic editing of quantities and values. • **Data Coding:** Specialized staff has coded the data of industry, occupation and geographical identification. • **Harmonized Data**: Cleaned data files are then all merged to produce one data file on the individual level containing all variables subject to harmonization. - Harmonized data is saved on the household as well as the individual level, in SPSS and converted to STATA format.

9- Documentation metadata of (HIECS)

Documentation **(HIECS)** is organized according to two standards Data Documentation Initiative (DDI) and Dublin core Metadata initiative (DCMI) each one of them includes many of the elements.

For more information about documentation (HIECS) visit the CAPMAS site.http://www.censusinfo.capmas.gov.eg/Metadata-ar-v4.2/index.php/catalog/84/ and visit Economic Research Forum website: http://www.erfdataportal.com/index.php/catalog/49.

10- Strengths and Limitations:

10.1-Strengths: The income and expenditures survey depends on appropriate sample size , It also covers a large number of variables related to employment and food and expenditures in different Activities of life, which makes it a very rich source of information about the conditions of Egyptian families and the problems which they suffer from it.

10.2-Limitations: Some components of income for households can be collected through the income surveys of institutional or administrative systems such as tax records and social insurance records....etc, as known These data are usually good quality and when using it we need to discuss issues of coverage of households as well as the types of income and reference periods and harmonization of time, definitions and units of analysis.

Conclusions

Summary quality report is A good tool for users to easily access quality information for the statistical survey based on ESS standard. In addition, SQR sheds lights on Documentation of the statistical operations including: definitions of concepts, methods of data collection, processing, dissemination, data about coverage, response rates, number of missing values, estimates of errors...etc

So CAPMAS is planning to ensure that SQRs are linked, and highly visible, on statistical output internet pages. This should allow all users of CAPMAS statistics to be able to easily access quality information and decide how the data can be used correctly**.**

References

1- Götzfried , A. and Linden, H. 2010, "Quality reporting within the Eurostat and ESS metadata systems simplification and improvement of ESS data quality reporting", Paper presented at the Work Session on Statistical Metadata (METIS), Geneva, 9-11 March 2010. http://www.unece.org/stats/documents/2010.03.metis.htm

2- Sarah ,G. Jacqui J. 2010 " User Views on Quality Reporting" paper presented at the European conference on quality in official statistics (Q2010), Helsinki, Finland 4-7 May 2010.

http://q2010.stat.fi/index.

3- Central Agency for Public Mobilization and Statistics:"Income Expenditure & consumption survey (2008/2009):"Survey Methodology, volume 1 , August 2009 .