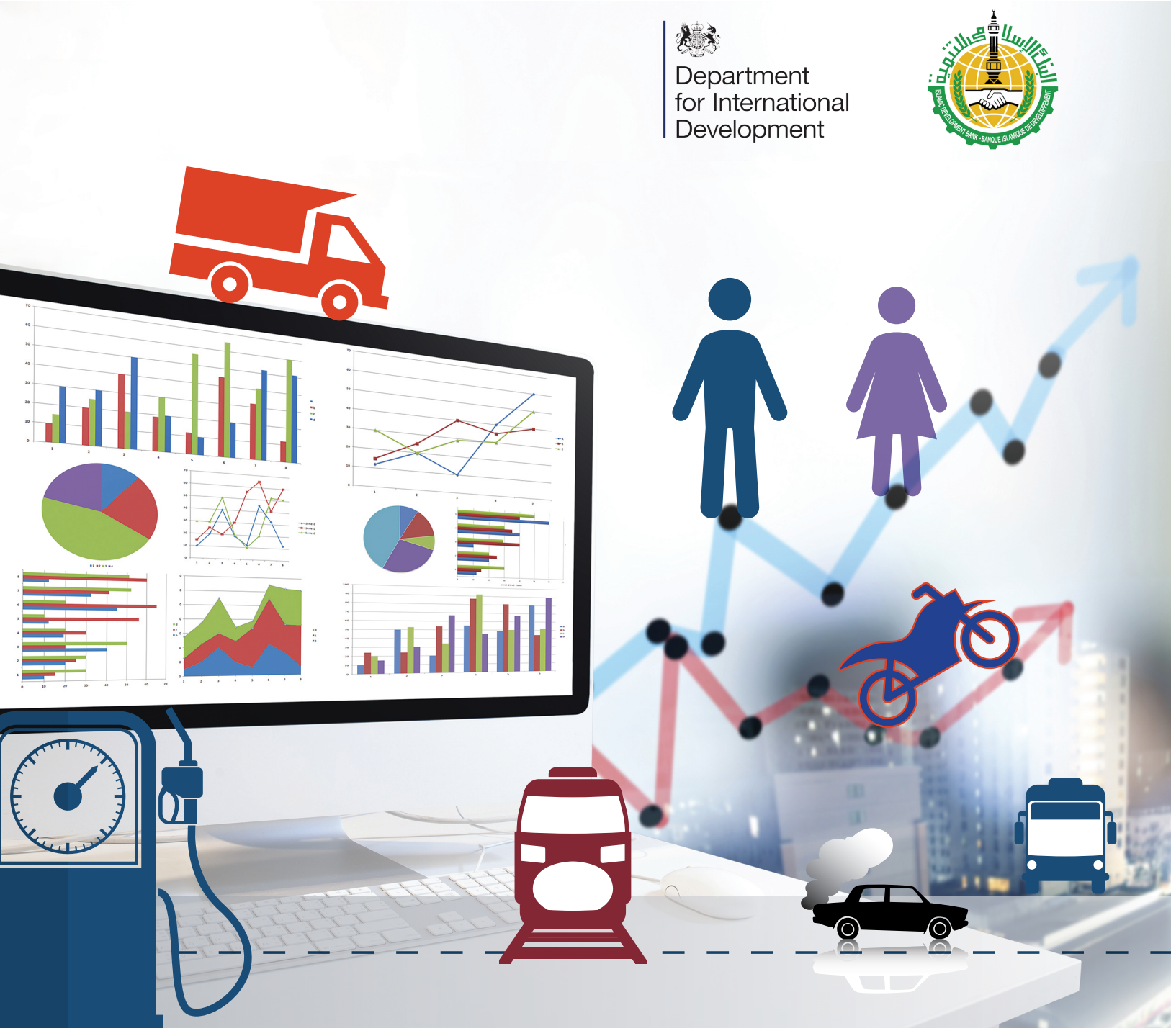
Economic and Social Commission for Western Asia

Department for International Development



UNITED NATIONS
ESCWA

Guidelines on Energy Consumption Surveys in the Transport Sector

Experiences in Selected Arab Countries

Strengthening Statistical Capacity of Arab Countries in Producing Energy Statistics and Energy Consumption Surveys

Guidelines on Energy Consumption Surveys in the Transport Sector

Experiences in Selected Arab Countries

Prepared by

Abdulhakeem Eideh[1], Therese El Gemayel[2] , and Wafa Aboul Hosn[3]

[1] Abdulhakeem Eideh is an associate professor of survey methodology at Al-Quds University, Palestine, Email: msabdul@science.alquds.edu

[2] Therese El Gemayel is an Energy and Environment Consultant and coordinator of the project on energy consumption in the transport sector survey (UN-ESCWA-IDB, 2015). Email: therese.gemayel@gmail.com

[3] Wafa Aboul Hosn is Chief Economic Statistics at the Economic and Social Commission for Western Asia (UNESCWA) and manager of the project on energy consumption in the transport sector survey (UN-ESCWA-IDB, 2015). Email: aboulhosn@un.org

Contents

Acknowledgements

The authors would like to thank the Islamic Development Bank and the Department for International Development of the UK for providing funds for the Economic and Social Commission for Western Asia (UNESCWA) to implement the project "Strengthening Statistical Capacity of Arab Countries in Producing Energy Statistics and Energy Consumption Surveys" in the three Arab countries: Egypt, Jordan and Palestine.

We thank the Director of the national statistical offices of Egypt (CAPMAS), of Jordan (DOS) and of Palestine (PCBS) who supported the survey at high level offices and the survey departments and the energy statistics units staff who worked intensively to conduct the survey in a short time and coordinated with the line ministries. The Director of the Statistics Division at ESCWA Juraj Riecan provided support and guidance.

Abbreviations

ECSTS	Energy Consumption Survey in Transport Sector
HH	Hansen and Hurwitz
HT	Horvitz-Thompson
IRES	International Recommendations for Energy Statistics
PPS	Probability Proportional to Size
PPSWR	PPS Sampling With Replacement
PPSWOR or πPS	PPS Sampling Without Replacement
SRSWR	Simple Random Sample With Replacement
SRSWOR	Simple Random Sample Without Replacement

Executive summary

The publication serves as reference guidelines on sampling frame, sample size determination, estimation of population total and population mean, for data collection on energy consumption in transport sectors, namely, road, railway, maritime, and air. The publication is divided into two main parts:

The first part contains the introduction which emphasizes the background and objectives of the guide. It describes the various collection methods of statistical data: administrative registers, sample surveys, definitions and concepts: survey sampling, sampling frame- frame population, survey planning, and censuses. It also describes the different sampling techniques and method of estimating population parameters: random and non-random sampling: notation and definitions, simple random sampling (with and without replacement), systematic random sampling, stratified random sampling, post-stratification, ratio estimators, regression estimators, single stage sampling- unequal probability of selection, and quota sampling.

The second part focuses on the preferred design for energy consumption, international recommendations for comparable energy statistics. It also describes the preparation and design phase of the survey on energy consumption in the transport sector that was conducted in Egypt, Jordan and Palestine for the year 2015.

PART I

1- Introduction

Data collection for the 2015 Energy Consumption Survey in Transport Sector (ECSTS) will be conducted between March and May 2015, collecting data for reference year 2014. The goal of the ECSTS is to provide statistical information about energy consumption, expenditures, and energy efficiency in Egypt, Jordan and Palestine, in all modes of transportations, namely: road, railway, maritime, and air. In addition, the goal is to link the resulted statistics to national policies on energy production and consumption based on collected statistics.

2- Collection of statistical data

Within the national statistical offices, three kinds of statistics are published – statistics based on sample surveys, statistics based on censuses and statistics based on administrative registers. It is most common to only differentiate between sample surveys and censuses, where the statistical office is responsible for the collection of the data. These two survey types are dominated by the work to collect data. However, data required to assess energy consumption in the field of transport can be obtained from sample surveys and administrative registers.

In this project, three types of data collection methods will be used, namely: administrative registers data bases, sample surveys (random and nonrandom samples), and censuses.

2.1. Administrative registers

According to Wallgren and Wallgren (2007), an **administrative register** is maintained to store records on all objects to be administered and the administrative process requires that it is possible to identify all objects. The following definition is valid for both administrative and statistical registers:

A **register** aims to be a complete list of the objects (e.g. vehicles) in a specific group of objects or population. However, data on some objects can be missing due to quality deficiencies. Data on an object's identity should be available so that the register can be updated and expanded with new variable values for each object. Complete listing and known identities are thus the important characteristics of a register (Wallgren and Wallgren, 2007).

The identity in register processing in energy use data in transport sector can be registration mark of vehicle identity number who is unique within a national administrative system.

Statistical register is used to describe registers within a system of statistical registers within a statistical office or other organization, such as ministries in charge of energy, energy observatories which are usually in charge of the preparation of the energy balances, Ministry of Transport, statistics institutes, etc. Such registers can be based either on a census carried out by the agency or on administrative registers from authorities and organizations outside the statistical office (Wallgren and Wallgren, 2007).

Data collection in a sample survey does not give rise to a register, as the micro data about the sample only consists of a small part of the surveyed population.

Register-based statistics refers to statistics that are based on existing registers (Wallgren and Wallgren, 2007).
Wallgren and Wallgren (2007) formulated four principles on how to use administrative data:

1. A statistical office should have access to administrative registers kept by public authorities. This right should be supported by law as the protection of privacy.
2. These administrative registers should be transformed into statistical registers. Many sources should be used and compared during this transformation.
3. All statistical registers should be included in a coordinated register system. This system will ensure that all data can be integrated and used effectively.
4. Consistency regarding populations and variables are necessary for the coherence of estimates from different register-based surveys.

2.2. Sample surveys

Sample surveys or finite population sampling is the most important branch of statistics and is distinguished from the rest of statistics by its focus on the **actual population** of which the sample is part of. Without sample surveys there is no data, and without data there is no statistics.

A **finite population** is a collection of distinct units such as vehicles, businesses, etc. **Sample survey** - partial investigation of the finite population, is concerned with selecting samples or subsets of the units, observing features of the sample units, and then using those observations to make inferences about the entire population or a quantity of a finite population (Sarndal et al., 1992 and Valliant et al., 2000).

According to Valliant et al. (2000), there are five general steps in a sampling investigation of a finite population:

(1) Define the scope and objectives of the study, including population to be studied and general information to collect.
(2) Choose tools and techniques for making observations, for example: a questionnaire, containing attitudes scales or questions asking for factual data like a person's income or the number of employees in a business; physical measurements such as height, weight and blood pressure; and expert inspection by customs officials of import shipments to determine whether contraband is being smuggled.
(3) Choose a sample.
(4) Gather data on the sample.
(5) Analyze the data and make inferences.

Nevertheless, existing information published by official sources are very often insufficient or incoherent for the proper definition of the sector's energy balance, mainly vehicles operated by non-organized structures (private vehicles, commercial vehicles, transport for other purposes, etc.) (ESCWA, 2013).

According to ESCWA (2013), it is necessary to conduct field surveys to set the baseline and validate data, which will help to develop indicators adapted to the local context, and eventually

design appropriate energy efficiency programs. However, surveys are usually expensive, and data handling must be performed according to valid and widely known methods leading to coherent and realistic statistics.

2.3. Definitions and concepts: survey sampling

Target population - U **:** The population to be investigated, and to which the conclusions refer. An enumeration rule exists that unequivocally defines the units belonging to the population. It must always be possible to determine in practice whether a certain unit does or does not belong to the target population. The results of the survey may be affected by the failure to include relevant units in the population and to exclude irrelevant ones (Sarndal et al., 1992).

The goal of a survey is to provide information about the target population in question or about subpopulations of special interest, for example, road passenger vehicles and road goods vehicles as two subpopulations of all vehicles. Such subpopulations are called domains of study or just **domains**. In short, target population - U is the set of elements about which information is wanted and parameter estimates are required**.**

For example, **in-scope vehicles** of the Canadian Vehicle Survey (CVS, 2009) include all motor vehicles, except buses (buses were included in the survey prior to 2004), motorcycles, off road vehicles (for example snowmobiles, dune buggies, amphibious vehicles) and special equipment (for example cranes, street cleaners, snowplows and backhoes), registered in Canada anytime during the survey reference period, that have not been scrapped or salvaged. The **population of interest** consists of vehicle-days, composed from the in-scope vehicles and the days within the survey reference period.

Variables of study: A value of one or more variables of study is associated with each population elements. For example:

- Gross vehicle weight.
- Carrying capacity.
- Type of fuel used.
- Number of fuel liters purchased for a specific vehicle during the survey week.

Finite population parameters: The goal of a survey is to get information about unknown population characteristics or parameters – functions of the study variable values. They are unknown, quantitative measures of interest to the investigator. For example:

- Number of vehicles by type of vehicle and vehicle model year.
- Number of vehicles by type of vehicle and type of fuel.
- Vehicle-kilometers and passenger-kilometers for trucks 15 tons or more.
- Final energy consumption of the sectors (road transport, railway transport, airline transport, maritime transport) and by fuel type (gasoil, gasoline, kerosene, etc.).

Sampling frame: In most surveys, access to and observation of individual population elements is established through a sampling frame – a device that associates the elements of the population with the **sampling units** in the frame **(**Sarndal et al., 1992).

Sample: "A sample is a subset of units in a population, selected to represent all units in a population of interest. It is a partial enumeration because it is a count from part of the population". From a population a sample - a subset of elements is selected. This can be done by selecting sampling units in the frame (Australian Bureau of Statistics, Statistical Language, 2013).

Measurements: The sample elements are observed, that is, for each element in the sample, the variables of study are measured and the values are recorded **(**Sarndal et al., 1992).

Point estimates: The recorded values are used to calculate point estimates of the finite population parameters of interest, for example, totals, means, ratios, regression coefficients, etc. **(**Sarndal et al., 1992).

Non-response rate: According to Sarndal et al. (1992), a **non-response rate** characterizes the sampled population from which resulted a "failure to obtain a measurement of one or more study variables for one or more elements selected for the survey". This observation is present in most of the conducted surveys.

Aggregated data is defined as "the result of transforming unit level data into quantitative measures for a set of characteristics of a population" (UNData, Glossary, 2008).

Questionnaire is a group or list of questions designed to collect information on one or more topics from a reporting unit or producer of official statistics (UNData, Glossary, 2008).

Statistical mean and variance are two different statistical measures used to explain data. The mean or arithmetic average is defined as the sum of all the observations divided by the total number of observations. The mean is a measure of center, which is "attempts to describe a whole set of data with a single value that represents the middle or center of its distribution" (Australian Bureau of Statistics, Statistical Language, 2013). Variance, on the other hand, measures spread of data around the mean. It describes "how similar of varied the set of observed values are for a particular variable" (Australian Bureau of Statistics, Statistical Language, 2013).

2.4. Sampling frame - frame population

According to Lessler and Kalsbeek (1992), sampling frame is any material or device used to obtain observational access to the finite population of interest. It must be possible with the aid of the frame to:
(a) Identify and select a sample in a way that respects a given probability sampling design;
(b) Establish contact with selected elements – by telephone, visit at home, mailed questionnaire, etc.

Units of the frame are the units to which the probability sampling scheme is applied.

Auxiliary information or auxiliary variable (or frame variable): According to Sarndal et al. (1992), auxiliary variable is any variable about which information is available prior to sampling (known for each of the population elements), measure of size, demographic information, which is used for: (a) Special sampling techniques, such as stratification and probability-proportional-

to-size selections sample; and (b) Special estimation techniques, such as poststratification, ratio or regression estimation.

The sampling frame provides the means of identifying and contacting the units of the target population. There are two main categories of frames: list and area frames.

Direct element sampling frame

Sarndal et al. (1992) described the direct element sampling frame or list frame as:

1. The units in the frame are identified through an identifier $k = 1,\ldots,N_F$, where N_F is the number of sampling units.
2. All units are available if selected in a sample. Example: address, telephone number, location on a map, or other device for making contact is specified in the frame or can be made available.
3. The frame is organized in a systematic fashion, for example, the units are ordered by geography or by size.
4. The frame contains a vector of additional information for each unit; such information may be used for efficiency improvement such as stratification or to construct estimators that involve auxiliary information.
5. When estimation is required for domains (sub-populations), the frame specifies the domain to which each unit belongs.
6. Every element in the population of interest is present only once in the frame.
7. Any element that does not figure in the population of interest is not added to the frame.
8. Every element in the population of interest which is present in the frame implies the frame gives access to the whole population of interest.

In practice, a frame is a file with an element identifier k running from 1 to N_F (Sarndal et al., 1992). It may contain other information. It can be stated that everything available in the frame about the k[th] element as a vector $\mathbf{x}_k = \left(x_{1k},\ldots,x_{jk},\ldots,x_{qk}\right)$, where x_{jk} is the value of the j^{th} variable for the k^{th} element; $j = 1,\ldots,q$ and $k = 1,\ldots,N_F$. The value x_{jk} may be quantitative (for example, salary for individual k) or qualitative (for example, sex for individual k). The frame can be seen as a matrix with N_F rows and q columns as follows:

Table 2.1. Frame Representation

Identifier	Auxiliary Variables (q) – Known Vector
1	$\mathbf{x}_1 = \left(x_{11},\ldots,x_{j1},\ldots,x_{q1}\right)$
2	$\mathbf{x}_2 = \left(x_{12},\ldots,x_{j2},\ldots,x_{q2}\right)$
⋮	⋮
k	$\mathbf{x}_k = \left(x_{1k},\ldots,x_{jk},\ldots,x_{qk}\right)$
⋮	⋮
N_F	$\mathbf{x}_{N_F} = \left(x_{1N_F},\ldots,x_{jN_F},\ldots,x_{qN_F}\right)$

An **area frame** is a unique type of list frame where all the units of the frame are represented as geographical areas. The survey population is located within these geographic areas. Area frames may be used when an adequate list frame is unavailable, in which case the area frame can be used as a basis for creating a list frame. Area frames are usually made up of a hierarchy of geographical units. Frame units at one level can be subdivided to form the units at the next level. Large geographic areas like provinces may be composed of districts or municipalities with each of these further divided into smaller areas, such as city blocks. In the smallest sampled geographical areas, the population may be listed in order to sample units within this area (Simard and Franklin, 2005).

2.5. Survey Planning

Sarndal et al. (1992) identifies some important aspects of survey planning. They are as follows:
(1) Specifying the objective of the survey.
(2) Translation of the subject-matter problem into a survey problem.
(3) Specification of target population.
(4) Specification of auxiliary variables (known variables), study variables, population parameters to be estimated.
(5) Construction of sampling frame.
(6) Specification of budget, staff, data processing and other equipment.
(7) Specifications of time schedule and accuracy of estimates.
(8) Specifications of data collection method, including questionnaire construction.
(9) Specification of sampling design, sample selection mechanism, and sample size determination.
(10) Specifications of data processing methods including editing and imputation.
(11) Specifications of formulas for point estimator and measure of precision (variance estimation).
(12) Training of personnel, organization of field work.
(13) Allocation of resources to different survey operations.
(14) Allocation of resources to control and evaluation.

2.6. Censuses

"A **census** is a study of every unit, everyone or everything, in a population. It is known as a complete enumeration, which means a complete count" (Australian Bureau of Statistics, Statistical Language, 2013). In other words, complete enumeration or census is a special type of surveys where the whole population is observed.

In order to carry out a census, a total enumeration of the population has to be identified. The **population** is defined as "any complete group with at least one characteristic in common. Populations are not just people. Populations may consist of, but are not limited to, people, animals, businesses, buildings, motor vehicles, farms, objects or events". (Australian Bureau of Statistics, Statistical Language, 2013).

Although a census is a complete enumeration of the population, it has advantages and disadvantages. It "provides the true measure of the population", without having to address the

sampling errors; its data can be benchmarked for future studies, and "provides detailed information about sub-groups within the population". On the other hand, a census is costly, takes a long period of time to implement and may face great problems in enumerating all the units of the population within a selective timeframe (Australian Bureau of Statistics, Statistical Language, 2013).

3-Sampling methods

In survey research there are mainly two different methods of sampling: first, random or probability-based sampling schemes or survey design methods, and non-random or non-probability-based sampling.

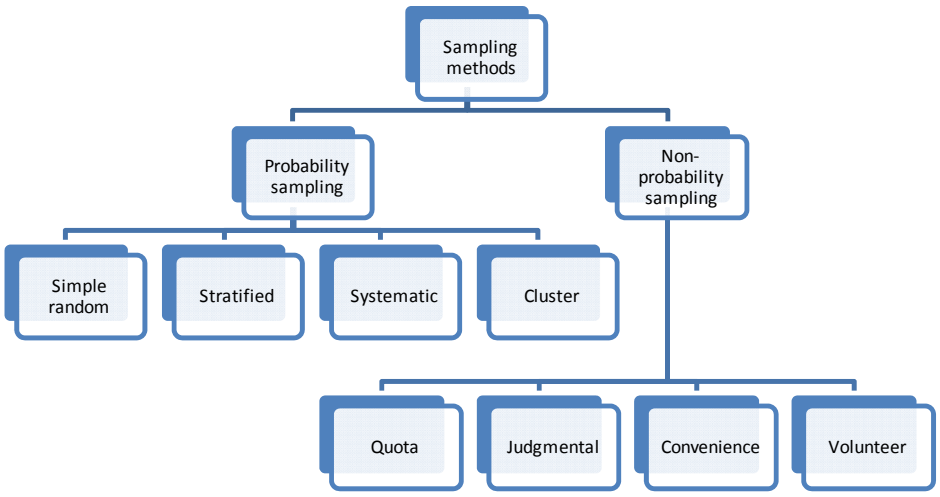


Figure 3.1. Probability and non-probability sampling methods

According to Cochran (1977), Sarndal et al. (1992) and Lohr (2009), let $U = \{1,..., N\}$ denote a finite population consisting of N sampling units, where N known and the sampling units are distinguishable individually. Let y be the **survey** or **target** or **study** variable of interest and let y_i be the value or characteristic of y for the ith population unit. In the **randomization theory** or **design-based approach** to sampling, the $y_i's$ are considered to fixed quantities but unknown numbers and are not realizations of random variables, and any probabilities used arise from the probabilities of selecting units to be in the sample. Below are a few general equations of finite population mean, total population, population variance, adjusted population variance, and coefficient of variation.

The **finite population parameter** or **finite population quantity** is any real valued function of the population values $y_1,..., y_N$. For example

Finite population mean or **population mean:**

$$\overline{Y}_U = \overline{Y} = N^{-1}\sum_{i=1}^{N} y_i$$

Total population:

$$Y_U = Y = \sum_{i=1}^{N} y_i = N\overline{Y}$$

Population variance:

$$\sigma^2 = \sigma_U^2 = N^{-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2$$

Adjusted population variance:

$$S^2 = S_U^2 = \left(N-1\right)^{-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2 = \frac{N\sigma^2}{N-1}$$

Population coefficient of variation or relative standard deviation:

$$CV_y = S/\overline{Y}$$

3.1. Probability sampling

In random sampling or probability-based sampling schemes, all items have some chance of selection that can be calculated.

According to Sarndal et al. (1992), **probability sampling** is an approach to sample selection that satisfies certain conditions, which, for the case of selecting elements directly from the population, are described as follows:

(a) Define the set of samples, $\Im = \{s_1,..., s_M\}$, that are possible to obtain with the sampling procedure.

(b) Sampling design: A known probability of selection $p(s)$ is associated with each $s \in \Im$.

(c) Inclusion probability: The sampling procedure gives every element in the population a nonzero probability of selection.

(d) Selection of one sample by a random mechanism (carried out by an easily implemented algorithm to eliminate the selection biases) such that each $s \in \Im$ receives exactly the probability $p(s)$.

Cochran (1977) identifies five common random sampling techniques: **simple random sampling** (each item in a population has equal chance of inclusion in the sample), **systematic sampling** (sometimes called interval sampling, means setting an interval for each selection, i.e. selecting every fifth sampling unit), **stratified sampling** (population is divided into groups called strata. A sample is then drawn from within these strata), **cluster sampling** (divides the population into groups, or clusters. A number of clusters are selected randomly to represent the population, and then all units within selected clusters are included in the sample. No units from non-selected

clusters are included in the sample) and **multi-stage sampling** (like cluster sampling, but involves selecting a sample within each chosen cluster, rather than including all units in the cluster. Thus, multi-stage sampling involves selecting a sample in at least two stages).

3.1.1. Simple random sampling (with and without replacement)

Simple random sampling without replacement (**SRSWOR**) is a method of selecting a sample s of n distinct units from a population $U = \{1,..., N\}$ of N sampling units (size of the population assumed known) such that every one of the distinct samples has the same chance of being drawn, or all possible distinct samples are equally likely to be chosen (Cochran, 1977). Once the number has been chosen, it is removed from the remaining list of samples. Unlike that, the simple random sampling with replacement (**SRSWR**) dictates to keep the chosen number, hence allowing for a repetitive selection of the same number again. In SRSWOR, the probability of selecting any individual sample s of n distinct units is:

$$p(s) = \Pr(s) = \frac{1}{{}_N C_n} = \frac{n!(N-n)!}{N!}$$

Where $n! = n \times n-1 \times \cdots \times 2 \times 1$.
The probability that unit i in s is n/N. That is: $\Pr(i \in s) = n/N$ for all $i = 1,..., N$.

Simple random sampling is the basic selection method, and all other random sampling techniques can be viewed as an extension or adaptation of this method (Banning, 2012).

Estimation of the Population Mean:
The sample mean is represented by:

$$\bar{y} = \bar{y}_s = \frac{1}{n}\sum_{i=1}^{n} y_i$$

and is an unbiased estimator of the population mean. That is, $E(\bar{y}) = \bar{Y}$.

The variance of $\bar{y}$ (measure of sampling error) is given by:

$$V(\bar{y}) = (1-f)\frac{S^2}{n}$$

Where $f = n/N$ is the sampling fraction.

It should be emphasized that $V(\bar{y})$ is a measure of the variability among estimates $\bar{y}_k$, over all possible samples $k = 1,\ldots,{}_N C_n$.

The factor $(1-f) = (N-n)/N$ for $V(\bar{y}_s) = S_U^2(1-f)/n$ is called the **finite population correction**. Intuitively, with small populations, the greater f, the more information about the population and the smaller the $V(\bar{y}_s)$. If $N = 5$ and all 5 observations are sampled, then there is

only one possible sample of size 5 without replacement, with $\bar{y} = \bar{Y}$, then $V(\bar{y}) = S^2(1-f)/n = S^2(1-1)/N = 0$. There is no sampling variability.

An unbiased estimator of $V(\bar{y})$ is:

$$\hat{V}(\bar{y}) = (1-f)\frac{s^2}{n}$$

Where $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the variance of the sample values.

The estimate of standard error (se) of $\bar{y}$ is:

$$se(\bar{y}) = \sqrt{\hat{V}(\bar{y})} = \frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$$

The estimated coefficient of variation of an estimate $\bar{y}$ - gives a measure of relative variability of an estimate, is:

$$\hat{CV}(\bar{y}) = \frac{\sqrt{\hat{V}(\bar{y})}}{\bar{y}} = \frac{se(\bar{y})}{\bar{y}} = cv(\bar{y})$$

Under simple random sampling with replacement (**SRSWR**), $\bar{y}$ is an unbiased estimator of $\bar{Y}$ with variance:

$$V(\bar{y}) = \frac{(N-1)S^2}{nN} = \frac{\sigma^2}{n}$$

Confidence Interval: For the point estimation, there is a single numerical value of the unknown finite population parameter $\bar{Y}$ without any idea of how confident it might be, based on a sample data, that the estimate is anywhere close to the true value of the finite population parameter. The question that arises is: is it possible to make any probability statement about the unknown finite population parameter in the finite population from which the SRSWOR is drawn? This question is very well answered by Hajek (1960). An approximate $100(1-\alpha)\%$ confidence interval for the finite population mean $\bar{Y}$ is: $\left(\bar{y} - z_{\alpha/2}se(\bar{y}), \bar{y} + z_{\alpha/2}se(\bar{y})\right)$.

Where $z_{\alpha/2}$ is the upper $\alpha/2$ point on the standard normal distribution. If $n < 50$, then replace $z_{\alpha/2}$ by $t_{\alpha/2}$ - the upper $\alpha/2$ point on the Student's t distribution with $n-1$ degrees of freedom.

An approximate $100(1-\alpha)\%$ error margin of estimating the finite population mean $\bar{Y}$ by the sample mean $\bar{y}$ is:

$$d = z_{\alpha/2} se(\bar{y})$$

Estimation of the **total population** $Y = \sum_{i=1}^{N} y_i = N\bar{Y}$:

Let

$$\hat{Y} = N\bar{y}_s = \sum_{i=1}^{n} \frac{N}{n} y_i = \sum_{i=1}^{n} w_i y_i$$

be an estimate of the finite population total , where $w_i = N/n = 1/\Pr(i \in s)$ is the **sampling or inclusion weight** for sampling unit.

Statistical properties of $\hat{Y}$:

(a) $\hat{Y} = N\bar{y}$ is an unbiased estimator of $Y = \sum_{i=1}^{N} y_i$.

(b) An unbiased estimator of $V(\hat{Y})$ is:

$$\hat{V}(\hat{Y}) = N^2(1-f)\frac{s^2}{n}$$

(c) The estimate standard error of $\hat{Y}$ is:

$$se(\hat{Y}) = \sqrt{\hat{V}(\hat{Y})} = \frac{Ns}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$$

(d) The estimated coefficient of variation of $\hat{Y}$ is:

$$cv(\hat{Y}) = \frac{N \times se(\bar{y})}{N\bar{y}} = cv(\bar{y})$$

(e) An approximate $100(1-\alpha)\%$ confidence interval for the finite population total Y is:

$$\left(N\bar{y} - z_{\alpha/2} N \times se(\bar{y}), N\bar{y} + z_{\alpha/2} N \times se(\bar{y})\right)$$

Estimation of the sample size:

Anyone designing SRSWOR must decide what amount of sampling error in the estimates is tolerable and must balance the precision of the estimate with the cost of the survey. Even though many survey variables may be measured, an investigator can often focus on one or two survey variables that are of primary interest in the survey and use these for estimating a sample size. In the planning of a sample survey, a stage is always reached at which a decision must be made about the size of the sample. This depends on the precision of the estimate to be specified. The statement of precision desired may be made by giving the amount of sampling error that will be tolerated in the sample estimate (Cochran, 1977).

Sample size in estimating finite population mean or population total:

According to Cochran (1977), suppose it is specified that the **absolute error** in estimating the finite population mean $\bar{Y}$ by the sample mean $\bar{y}$ is less than d with confidence coefficient $100(1-\alpha)\%$, then the precision is expressed as:

$$P\left(\left|\bar{y}-\bar{Y}\right| \le d\right) = 1-\alpha$$

The investigator must decide on reasonable values of the risk α and the margin error or sampling error d in many surveys. For many surveys, $\alpha = 0.05$. Assuming that $\bar{y} \sim NORMAL\left(E(\bar{y}), V(\bar{y})\right)$, then the value of n satisfies:

$$d = z_{\alpha/2}\sqrt{V(\bar{y})} = z\sqrt{V(\bar{y})},\ z = z_{\alpha/2}$$

This equation connects n with the desired degree of precision.

Solving for n, we get:

$$n = \frac{n_0}{1+n_0/N}$$

Where $n_0 = (zS/d)^2$.

Note that n_0 does not depend on the population size N and $n_0 = \left(z^2S^2\right)/d^2$ is the sample size for random sample from a population with large size. That is, if N is large, a first approximation of n is:

$$n_0 = \left(z^2S^2\right)/d^2 = S^2/V$$

Where V is the desired variance of the sample mean $\bar{y}$.

According to Cochran (1977), most commonly, the **relative error** r is desired to be controlled. Suppose it is specified that the relative error r in estimating the finite population mean $\bar{Y}$ by the sample mean $\bar{y}$ is less than r with $100(1-\alpha)\%$ confidence coefficient, then the precision is expressed as:

$$P\left(\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right| \le r\right) = 1-\alpha$$

Then

$$n = \frac{n_0}{1+(n_0/N)}$$

Where $n_0 = z^2CV_y^2/r^2$.

So that, n_0 is a function of population coefficient of variation, $CV_y = S/\overline{Y}$. This is often more stable and easier to guess in advance than the population variance S^2 (Lohr, 2009).

Specifying S^2:

Typically the population variance S^2 is unknown. In practice there are four methods of estimating S^2 for sample size determination (Cochran, 1977):

(a) Use of the results of a **pilot study** – a small sample taken to provide information about S^2.

(b) Use of previous studies or data in literature of the same or similar populations. Lohr (1999, p 41) pointed out that "you are rarely the first person in the world to study anything related to your investigation".

(c) By guesswork about the structure of the population, assisted by some mathematical model.

(d) Use of double-sampling scheme.

Specifying CV_y:

According to Valliant et al. (2013), to put the method described above into practice, a value for the target coefficient of variation, $CV0$, must be set. To some extent, the value is arbitrary although rules of thumb have been developed over the years. A CV of an estimate of 50% would imply that a normal-approximation confidence interval formed by adding and subtracting two standard errors of an estimate would cover zero. Such an estimate obviously is highly imprecise. The US National Center for Health Statistics flags any estimate it publishes that has a CV of 30% or more and labels it as "unreliable". Often, an estimate with a CV of 10% or less is considered "reliable", but the purposes to which the estimate will be put must be considered.

3.1.2. Systematic random sampling

This method uses a systematic selection of the elements within the sample; for example, sampling every 15th vehicle in a total of 5,000 vehicles. Sometimes **systematic sampling** is used as proxy for SRSWOR, when no list or addresses of the population exists. Such cases occur in the surveys which are being implemented as part of the project on energy consumption in the transport sector. There is a frame of vehicles contains auxiliary information, like vehicle type, age, engine capacity, municipality, etc., but the problem is there is no address for vetches. The apparent advantages of systematic random sampling over SRSWOR are:

(a) It is easier to draw a sample and easier to apply without mistake.

(b) Intuitively, systematic sampling seems likely to be more precise than SRSWOR.

a. Linear systematic random sampling

According to Sarndal et al. (1992), consider a finite population of size N, the units of which are identified by the labels $1,\ldots,k,\ldots,N$. Let a be the fixed **sampling interval** and let n be the integer part of N/a. Then $N = na + c$, where the integer c satisfies $0 \le c < a$. Two cases are available for consideration:

Case 1. If $c = 0$, a sample of size n will be drawn.

Case 2. If $c > 0$, the sample size is going to be either n or $n+1$, so the sample size is random and not fixed.

The selection is as follows:
(1) Select with equal probability $1/a$ a random integer, say r, called **random start**, between 1 and a (inclusive).
(2) The selected sample is composed of:

$$s = s_r = \{k : k = r + (j-1)a \le N;\ j = 1,\ldots,n_s\},\ r = 1,\ldots,a$$

Where the sample size n_s is either $n+1$ (when $r \le c$) or n (when $c < r \le a$).

The set of all possible (non-overlapping) distinct systematic samples a are equally likely to be chosen. So that the probability of selecting any individual systematic sample s of n distinct units is: $\Pr(S = s) = p(s) = 1/a$.

Note the systematic samples $s_1,\ldots,s_a$ represent a partition of $U = \{1,\ldots,N\}$.

For example, suppose that $N = 11$ and $a = 3$, then the integer part of $N/a = 11/3$ is 3. Therefore, $N = 11 = (3)(3) + 2$, and $c = 2 > 0$. Hence, the sample size is going to be either 3 or 4. The following table contains the systematic samples.

Table 3.1.Systematic sampling

	Systematic sample		
	s_1	s_2	s_3
y values	1	2	3
	4	5	6
	7	8	9
	10	11	

The first two systematic samples have $n = 4$ and the last one has $n = 3$.
Now, if $N = 12$ and $a = 3$, then the integer part of $N/a = 12/3 = 4$. Therefore, $N = 12 = (3)(4) + 0$, and $c = 0$. Hence the sample size is 4. Table 3.2 gives the possible systematic samples.

Table 3.2.Systematic sampling

	Systematic sample		
	s_1	s_2	s_3
y values	1	2	3
	4	5	6
	7	8	9
	10	11	12

The three systematic samples have size $n = 4$.

A point estimator of the population mean is the systematic sample mean $\bar{y}_{sy}$.

Sarndal et al. (1992) states that one of the disadvantages of systematic random sampling is that it is not possible to construct an unbiased estimator of $V(\bar{y}_{sy})$. But by treating the systematic sample as SRSWOR, it is possible to construct some biased, but useful variance estimator. That is:

$$\hat{V}(\bar{y}_{sy}) = \frac{1-f}{n} s^2 = \left(\frac{1}{n} - \frac{1}{N}\right) s^2$$

Where $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_{sy})^2$.

b. Circular systematic random sampling

According to Sarndal et al. (1992), assume N is not a multiple integer of n, that is, $N = na + c$, $c > 0$. In circular systematic random sampling, the sequential list of the population labels or units $U = \{1,2,\ldots,N\}$ is first prepared on the circle in clockwise direction. The main steps involved in selecting a sample using circular systematic random sampling scheme are as follows (Sarndal et al., 1992):

(a) Select a random number from 1 to N and name it as **random start**, say r .

(b) Choose some integer value of $a = N/n$ or round to nearest integer and name it as **skip** or **span.**

(c) Select all units in the sample with serial numbers:

$$r + ja \quad \text{if} \quad r + ja \le N$$
$$r + ja - N \quad \text{if} \quad r + ja > N, \; j = 0,1,\ldots,n-1$$

For example to select a sample of size 3 from a population of size 10, the process is as follows:

(i) Random start: Select a random start from 1 to 10, say $r = 3$.

(a) Skip or span: The nearest integer of $a = 10/3$ is $a = 3$.

(b) Compute $r + ja = 3 + 3j$; $j = 0,1,2$. Since $r + ja = 3 + 3j < 10$, therefore the sample would be with serial numbers: 3, 6, 9.

(ii) Random start: Select a random start from 1 to 10, say $r = 7$.

(a) Skip or span: The nearest integer of $a = 10/3$ is $a = 3$.

(b) Compute $r + ja = 7 + 3j$; $j = 0,1,2$. Since $7 + 2(3) = 13 > 10$, therefore the sample would be with serial numbers: 3, 7, 10.

Note that different random starts give different samples.

3.1.3. Stratified random sampling

According to Cochran (1977), Lohr, (2009), and Kish (1965), the finite population in **stratified sampling** is divided into non-overlapping subpopulations called **strata**. The strata constitute the whole population so that each sampling unit belongs to exactly one **stratum**. From each strata probability sample (a sample for which every element in the sampling frame has a positive chance or probability of being part of the sample) is selected. The selections in the different strata are independent. Then extract the information obtained from the different strata to obtain overall finite population estimates.

Choosing stratification variables:
Choosing the stratification variables has at least five advantages, according to Lohr (1999):
1. Avoid selecting a sample that is poorly distributed across the population.
2. A guaranteeing measure of certain sample sizes in groups that will be studied separately (i.e. domains).
3. Administrative convenience (e.g. a mail survey might be used for units in some strata but personal interviews for the remaining strata).
4. Cost management (e.g. data collection in some strata might be more expensive than in other strata).
5. Improving the sample efficiency for full population estimates by grouping units together that has similar mean and variance properties.

According to Valliant et al. (2013), stratification by size with an efficient allocation is an example of point 5 above. This method uses a size variable that is correlated with measured unit in the survey. In the ECSTS, the vehicle type, weight, engine capacity, and/ or age, are correlated with fuel consumption.

According to Valliant et al. (2013) the **principles of stratification** are as follows:
(a) The strata should be non-overlapping and should comprise the whole population.
(b) The stratification of population should be done in such a way that strata are homogenous within themselves with respect to the characteristics of interest and heterogeneous between strata.
(c) In many practical situations when it is difficult to stratify with respect to the selected characteristics, administrative convenience may be considered as the basis for stratification.
(d) If the limit of precision for certain subpopulation is given, it will be better to treat each subpopulation as a stratum.

To fully capitalize on the potential of the stratified sampling technique, the sampler must first resolve a number of technical questions. The objective is to select an efficient, yet practical, stratified sample. To do so, samplers need to answer some important questions to assist them in deciding the sampling method:

a- **The construction of strata:**

- If there is a choice, which stratification variable or variables should be used? **Stratification variable** is the characteristic or characteristics used for subdividing the population into strata. For example, would an age and sex stratification be preferred to stratification by occupational groups?
- How should strata be demarcated? If the stratification uses age groups, what age intervals should be used to set up the strata?
- How many strata should there be? If age is a stratification variable, how many age groups should there be?

b- **Choice of sampling and estimation methods within strata:**

- Are the sampling design and size specified in each stratum? Often the sample type of sampling design is applied in all of the strata.
- Is there an estimator specified for each stratum? Often this choice is also made uniformly for all strata.
- What effect would post-stratification (choosing the strata after seeing the data) have on the estimates?

Stratified sampling for estimating population mean:

In **stratified random sampling,** the finite population $U = \{1,\ldots,k,\ldots,N\}$ of N sampling units is divided into *H* non-overlapping strata or subpopulations, denoted by $U_1,\ldots,U_h,\ldots,U_H$ with N_h sampling units in the h^{th} stratum, where $N_1 + \ldots + N_H = N$ known in advance, where $U_h = \{k : k \text{ belongs to stratum } h\}$, $h = 1,\ldots,H$ so that $U = U_1 \cup \ldots \cup U_H$. An SRSWOR of size n_h, denoted by s_h, is independently taken from stratum U_h, for all $h = 1,\ldots,H$. The resulting total sample is $s = s_1 \cup \ldots \cup s_H$. Let

y_{hj} = the j^{th} value of the study variable in the h^{th} stratum, where $j = 1,\ldots,N_h$ and $h = 1,\ldots,H$,

$W_h = N_h/N$ = relative size of stratum h or proportion of population units falling in stratum h.

$\bar{Y}_h = N_h^{-1}\sum_{j=1}^{N_h} y_{hj} = N_h^{-1}Y_h$ = population mean in stratum h or true mean of stratum h,

$\bar{Y} = N^{-1}Y = N^{-1}\sum_{h=1}^{H}\sum_{j=1}^{N_h} y_{hj} = \sum_{h=1}^{H} W_h\bar{Y}_h$ = overall population mean,

$S_h^2 = (N_h - 1)^{-1}\sum_{j=1}^{N_h}\left(y_{hj} - \bar{Y}_h\right)^2$ = population variance of stratum h.

A commonly employed estimator of the population mean is the **stratified sample mean**, $\bar{y}_{st}$, defined by:

$$\bar{y}_{st} = \sum_{h=1}^{H}\frac{N_h}{N}\bar{y}_h = \sum_{h=1}^{H} W_h\bar{y}_h$$

Where:

n_h = sample size of stratum h.

$n = n_1 + ... + n_H = \sum_{h=1}^{H} n_h$ = overall sample size.

$\bar{y}_h = n_h^{-1} \sum_{j=1}^{n_h} y_{hj} = n_h^{-1} y_h$ = sample mean in stratum h.

Note that the stratified sample mean, $\bar{y}_{st} = \sum_{h=1}^{H} W_h \bar{y}_h$, is not in general the same as the (overall) sample mean, $\bar{y} = \sum_{h=1}^{H} \frac{n_h}{n} \bar{y}_h$, unless $n_h / n = N_h / N$, or equivalently $n_h = n(N_h / N)$ for all $h = 1, \ldots, H$. When this occurs, it is named **proportional sampling allocation.**

Statistical properties of stratified sampling mean:

(a) The stratified sampling mean $\bar{y}_{st}$ is an unbiased estimator of the overall population mean $\bar{Y}$.

(b) The variance of $\bar{y}_{st}$ is:

$$V(\bar{y}_{st}) = \sum_{h=1}^{H} W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$

(c) The estimated standard error of $\bar{y}_{st}$ is:

$$se(\bar{y}_{st}) = \sqrt{\hat{V}(\bar{y}_{st})} = \sqrt{\sum_{h=1}^{H} W_h^2 (1 - f_h) \frac{s_h^2}{n_h}}$$

Where $s_h^2 = (n_h - 1)^{-1} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$ is the sample variance of stratum h, and $f_h = n_h / N_h$ is the sampling fraction of stratum h.

The stratified sampling estimator $\hat{Y}_{st} = N\bar{y}_{st}$ of the overall total population $Y = N\bar{Y}$ can be written as:

$$\hat{Y}_{st} = N\bar{y}_{st} = \sum_{h=1}^{H} \sum_{j=1}^{n_h} w_{hj} y_{hj}$$

Where $w_{hj} = N_h / n_h$ is the **sampling weight.**

For example, a city transportation department is conducting a survey to determine the gasoline usages of its residents. Stratified random sampling is used, and the four city wards are treated as the strata. The amount of gasoline purchased in the last week is recorded for each household sampled. The strata sizes and the summary information obtained from the sample are given in the following table:

Table 3.3.Strata size and summary sample information

Stratum h	N_h	n_h	$\bar{y}_h$	s_h^2
1	3750	50	12.6	2.8
2	3272	45	14.5	2.9
3	1387	30	18.6	4.8
4	2475	30	13.8	3.2

Then the estimated mean of weekly gasoline usage per household, for the city population is the stratified sample mean:

$$\bar{y}_{st} = \sum_{h=1}^{4} \frac{N_h}{N} \bar{y}_h = \frac{3750}{10894}12.6 + \frac{3272}{10894}14.5 + \frac{1387}{10894}18.6 + \frac{2475}{10894}13.8$$
$$= 14.21$$

And the unbiased estimated variance of $\bar{y}_{st}$ is:

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{4} W_h^2 (1 - f_h) \frac{s_h^2}{n_h} = \left(\frac{3750}{10894}\right)^2 \left(1 - \frac{50}{3750}\right)2.8 + \left(\frac{3272}{10894}\right)^2 \left(1 - \frac{45}{3272}\right)2.9$$
$$+ \left(\frac{1387}{10894}\right)^2 \left(1 - \frac{30}{1387}\right)4.8 + \left(\frac{2475}{10894}\right)^2 \left(1 - \frac{30}{2475}\right)3.2$$
$$= 0.02$$

a- Post-stratification sampling

Post-stratification is the stratification made after the selection of the sample. With some variables that are suitable for stratification, the stratum to which a unit belongs is not known until the data have been collected. Personal characteristics such as age, sex, race, and educational level are common examples. The stratum sizes N_h may be obtainable fairly from official statistics, but the units can classify into strata only after the sample data is known (Kish, 1965).

In **post-stratified sampling scheme,** a sample of n units is first selected, **without stratification**, from the population of N units using any sampling design. The population is stratified into H strata on the basis of some auxiliary information available. In post-stratified sampling, the values of N_h where $h = 1, \ldots, H$ and $\sum_{h=1}^{H} N_h = N$ may or may not be known. Each sample unit selected with the chosen sampling design is then post-stratified or placed in stratum h based on the **auxiliary information associated with each sampled unit** such that $\sum_{h=1}^{H} n_h = n$. Thus the difference between stratified and post-stratified sampling schemes is a fixed number of strata sample sizes of n_h in stratified sampling, whereas in post-stratified sampling it is a random variable.

Kish (1965) pointed out the requirements for post-stratification:
(a) Information on the proportions W_h of the population in the several strata.
(b) Information for classifying the sample cases into same strata.

Post-stratification does not require that every member of the population be classified and sorted into its stratum before selection, as proportionate selection does. These considerations address the following question: When should post-stratification be used instead of proportionate selection?

According to Kish (1965): (a) The stratifying variable may be unavailable for classifying and sorting each population element. For example, in a sample of a factory, suppose that union membership is confidential and not available for all workers; suppose also that the respondent reveal their membership, and that the union will give the data on the total number of members. Then union membership can be used for post-stratification, but not for stratification in the selection. Unavailable may also simply mean that it is too expensive to stratify the entire population. When the total vote for each party is available after an election, this variable can be used for post-stratification, but not for selection. (b) The stratifying variable, although available, may not be used. Perhaps at the time of selection the sampler overlooked it, or for other reason failed to use it. Perhaps there were too many variables available, and some others variables were chosen instead.

Holt and Smith (1979) showed that post-stratification is potentially more efficient than stratification. This technique is found to be most practical for surveys where individual responses may be expected to vary with age, sex, occupation, education, state, country, race etc. Usually none of these variables is available for stratification at the **individual level** prior to sampling. Such a situation is called **conditional post-stratification.** In some situations, censuses may provide information on all of these variables at the aggregate level. In other situations, aggregated level of information may not be available; it is then referred to as **unconditional post-stratification**, which is of interest in practice.

Conditional post-stratification:
Conditional post-stratification can be defined as a sampling technique in which SRSWOR is selected without stratification. Once the strata sample size $n_h \geq 2$ is known, $h = 1, \ldots, H$ such that $\sum_{h=1}^{H} n_h = n$, the traditional stratified estimator is used. If there is any $n_h < 2$ then the h^{th} stratum is merged with another t^{th} stratum such that the composition is homogeneous.

Let $\bar{y}_{pst}$ be the post-stratified sample mean defined by:

$$\bar{y}_{pst} = \sum_{h=1}^{H} W_h \bar{y}_h, W_h = N_h / N .$$

Statistical properties of $\bar{y}_{pst}$, as an estimator of the population mean $\overline{Y}$:

(a) The post-stratified sample mean, is an unbiased estimator of the population mean $\overline{Y}$.

(b) Espejo and Pineda (1997) proposed an estimator to estimate $V(\bar{y}_{pst})$ as:

$$\hat{V}_c(\bar{y}_{pst}) = \frac{N-1}{N-n}\hat{V}(\bar{y}) - \sum_{h=1}^{H}\frac{W_h}{n_h}\sum_{j=1}^{n_h} y_{hj}^2 + \bar{y}_{pst}^2$$

Where

$$\hat{V}(\bar{y}) = \sum_{h=1}^{H}\frac{W_h}{n_h}\sum_{j=1}^{n_h} y_{hj}^2 - \bar{y}_{pst}^2 + \sum_{h=1}^{H} W_h^2\frac{s_h^2}{n_h}$$

It is to be noted here that the conditional formula of variance under post-stratified sampling schemes is useful when constructing the confidence intervals or testing hypotheses using survey results.

Unconditional post-stratification:

The unconditional post-stratification is a mechanism for the analysis of the exact variance for the post-stratification estimators when the stratum sample sizes $n_h \geq 2$ is known, $h = 1,\ldots,H$, $\sum_{h=1}^{H} n_h = n$ being the sample size from a finite population. The h^{th} stratum is collapsed with another t^{th} stratum with homogeneous composition to the h^{th} stratum. Thus the n_h is random before selecting the sample. Let $\bar{y}_{pst}$ be the post-stratified sample mean.

The **statistical properties** of $\bar{y}_{pst}$, as an estimator of the population mean $\overline{Y}$, are as follows:

(a) The post-stratified sample mean, defined as:

$$\bar{y}_{pst} = \sum_{h=1}^{H} W_h\bar{y}_h, W_h = N_h/N,$$

is an unbiased estimator of the population mean $\overline{Y}$.

(b) An estimator of $V(\bar{y}_{pst})$ is given by:

$$\hat{V}_{uc}(\bar{y}_{pst}) = \frac{1-f}{n}\sum_{h=1}^{H} W_h s_h^2 + \frac{1}{n^2}\sum_{h=1}^{H}(1-W_h)s_h^2$$

Note that $\hat{V}_{uc}(\bar{y}_{pst})$ is the sum of two components: the first component is the variance of the stratified sample mean $\bar{y}_{st}$ under stratified random sampling with proportional allocation. The second component is the additional variance due to post-stratification and it reflects the fact that the n_h in the resulting sample are random variables. For large n the second component is smaller compared to the first component. Thus for large samples:

$$\hat{V}_{uc}\left(\bar{y}_{pst}\right) \approx \frac{1-f}{n}\sum_{h=1}^{H} W_h s_h^2$$

The post-stratification technique results in estimates which are nearly as good as the proportional allocation. It is to be noted here that unconditional formula of variance is useful when planning for a survey (before the sample is selected), or when evaluating one methodology against the other.

3.1.4. Single stage sampling - unequal probability of selection

Sampling from a finite population often highly correlates the values of some auxiliary variable x with the study variable y , and are available for all sampling units of the population. The variable x may be taken as a **measure of size** of a unit. For example, in ECSTS, the data on the age, type, weight, region of vehicles may be available from vehicles registration databases. In such cases, instead of sampling the units with equal probabilities with or without replacement one may sample the units with **probability proportional to size-measure x (PPS) with or without replacement.**

Since a vehicle with higher values of **weight** is expected to contribute more to the energy consumption than those with smaller weights, one may expect that a selection procedure which gives higher selection probabilities to bigger units than to smaller units should be more efficient than a simple random sample. Therefore PPS sampling with replacement (PPSWR or PPS) and PPS sampling without replacement (PPSWOR or πPS) are considered.

a- Probabilities proportional to size with replacement sampling:

The general form of sampling with replacement has the following features. Suppose that $p_1,\ldots,p_N$ are given positive numbers satisfying $\sum_{i=1}^{N} p_i = 1$ and $\Pr\left(i \in s\right) = p_i, i = 1,\ldots,N$. The selected sampling unit, i_1, is replaced. The same set of probabilities is used to select the second sampling unit**,** i_2 ,… and the n^{th} sampling unit i_n . The n draws are independent.

Hansen and Hurwitz (HH) (1943) estimator, which estimates the total population, denoted by $\hat{Y}_{pps}$ expressed as $Y = \sum_{i=1}^{N} Y_i$ is given by:

$$\hat{Y}_{pps} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{p_i}$$

Statistical properties of $\hat{Y}_{pps}$:

The HH estimator $\hat{Y}_{pps}$ is unbiased for Y , with:

$$\hat{V}\left(\hat{Y}_{pps}\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{p_i} - \hat{Y}_{pps}\right)^2$$

Note that, $\hat{V}\left(\hat{Y}_{pps}\right)$ is exactly zero if p_i is exactly proportional to y_i. In practice it is not possible to choose p_i proportional to y_i, because y_i, $i = 1,\ldots,N$ are not ordinarily known up to a constant proportionality. However, if p_i can be chosen to be nearly proportional to y_i, then $\hat{V}\left(\hat{Y}_{pps}\right)$ would still be quite small. This is why the PPS estimator has such that the ratio y_i/x_i is roughly constant for all $i = 1,\ldots,N$. So $p_i \propto x_i$ is taken, where x_i is an auxiliary variable called **measure of size**, roughly proportional to y_i.

This procedure yields a small variance and is a particular application of PPSWR sampling. If there is a choice between different size measures, then the measure of size most roughly proportional to the study variable is likely to be the best.

b. Probabilities proportional to size without replacement sampling:

Generally, sampling with replacement is less precise than sampling without replacement, an advantage of selection with replacement is that the formulas for the true and estimated variances of the estimators are simple. Nevertheless, in large surveys, much research has been done on unequal probability sampling without replacement.

The most popular estimator of the total population under probabilities proportional to size without replacement is the Narain (1951) and Horvitz-Thompson (HT) (1952) estimator, known as the HT estimator.

The **Horvitz-Thompson estimator** of the total population $Y = \sum_{i=1}^{N} y_i$ is a linear estimator of the sample observations; $y_1,\ldots,y_n$, which is defined by:

$$\hat{Y}_{HT} = \sum_{i=1}^{n} w_i y_i = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i\in s} \frac{y_i}{\pi_i}$$

Where $\pi_i = \Pr(\text{unit } i \text{ in sample})$ and $w_i = 1/\pi_i = 1/np_i$ is the sampling weight of the i^{th} sampling unit for all $i = 1,\cdots,N$.

Statistical properties of $\hat{Y}_{HT}$:

The HT estimator $\hat{Y}_{HT}$ is an unbiased estimator of Y, with:

$$V\left(\hat{Y}_{HT}\right) = V_{HT} = \sum_{j=1}^{N}\sum_{i=1}^{N}\left(\pi_{ij} - \pi_i\pi_j\right)\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}$$

Yates and Grundy (1953) and Sen (1953) give an alternative expression for $V\left(\hat{Y}_{HT}\right)$.

$$V\left(\hat{Y}_{HT}\right) = V_{SYG} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2\left(\pi_i\pi_j - \pi_{ij}\right)$$

Where $\pi_{ij} = \Pr(\text{units } i \text{ and } j \text{ in sample})$.

Note that $V(\hat{Y}_{HT}) = V_{SYG}$ is exactly zero if $\pi_i = n\, y_i / Y$. In practice it is not possible to choose π_i proportional to y_i, because y_i, $i = 1, \ldots, N$ is not ordinarily known before sampling. However, if the $\pi_i = n\, x_i / X$ can be chosen so that y_i is nearly proportional to x_i, then $V(\hat{Y}_{HT}) = V_{SYG}$ would still be quite small. So $\pi_i \propto x_i$ is taken, where x_i is an auxiliary variable called measure of size, roughly proportional to y_i. This procedure yields a small variance and is a particular application of PPSWOR sampling.

Variance estimation:
According to Sarndal et al. (1992), and Wolter (2007), the technique consists of taking a variance estimator approximate to sampling with replacement and using it for sampling without replacement. For example, under PPSWOR sampling design:

$$\hat{V}_0 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{p_i} - \hat{Y}_{HT} \right)^2$$

is an estimator of $V(\hat{Y}_{HT})$. The tradeoff for simplification is that $\hat{V}_0$ is no longer unbiased for V. In many cases, however, the bias is positive, so that $\hat{V}_0$ will be "on the safe side".

3.2. Nonprobability sampling:

According to the Australian Bureau of Statistics, non-random sampling is a sampling technique based on specific characteristics chosen non-randomly to select the actual sample. Four common non-random sampling techniques are: **quota sampling** (type of stratified sampling in which selection within the strata is non-random), **convenience sampling** (does not produce a representative sample of the population because people or items are only selected for a sample if they can be accessed easily and conveniently, **volunteer sampling** (a common method of volunteer sampling is phone-in sampling, used mainly by television and radio stations to gauge public opinion on current affairs issues such as preferred political party, capital punishment, etc. People are asked by telephone to vote on a particular issue within a certain time, with no limit to the number of people who can call in, and **expert or judgmental sampling** (relies upon "experts" to choose the sample elements. Supporters claim that the method avoids the potential, when random techniques are used, of selecting a "bad" or odd sample, such as one in which all the sample elements unluckily fall in, say, the south region.

3.2.1. Quota sampling

The application of quota sampling implies at the first step the construction of a matrix containing all possible combinations of theoretically relevant respondent characteristics being of interest for the individual study. As a second step the target-number of respondents for each combination of characteristics is defined (Hoffmeyer-Zlotnik and Krebs, 1996).

The population is divided into different subpopulations just as in stratified random sampling, but with one important difference: the probability sampling is not used to choose individuals in the subpopulation for the sample. In extreme versions of quota sampling, choice of units in the sample is entirely at the discretion of the interviewer, so that a sample of convenience is chosen within each subpopulation.

In quota sampling, specified numbers (quotas) of particular types of population units are required in the final sample. For example, to obtain a quota sample with $n = 100$ vehicles, one might specify that the sample contains 60 vehicles of type 1, and 40 vehicles of type 2, but give no further instructions about how these quotas are to be filled. Thus, quota sampling is not a form of probability sampling; the probabilities of inclusion of each individual in the sample are unknown. It is often used when probability sampling is impractical, overly costly, or considered unnecessary, or when the persons designing the sample does not have additional knowledge (Lohr, 2009).

Thus, from the point of view of the theoretical assumptions, the design of a quota sample is very similar to a simple random sample. But while in random sampling the rules for contacting a specific person are specified, this is not the case in quota sampling. Here the interviewer has a big amount of freedom to choose a specific person for an interview as long as she/he sticks to the combination of characteristics a priori defined by the "quota-matrix" (Noelle-Neumann, 1963).

Thus, the researcher usually knows little about the strategy applied by interviewers to complete the interviews with the target number of respondents possessing the prescribed quota characteristics. The interviewer is free in selecting a specific person as respondent. The researcher mainly expects that the number of interviews with respondents possessing the target characteristics corresponds to the target number and that the interviews are completed within a pre-defined sample point. In the ideal case, each interviewer should be restricted to only one sample point.

To select a sample in a scientifically justified way, two techniques are required: a sampling design based on probability sampling and a sampling frame. It must be possible with the aid of the frame to:
(a) Identify and select a sample in a way that respects a given probability sampling design, and
(b) Establish contact with selected elements – by telephone, visit at home, mailed questionnaire, etc. In order to select a statistically valid sample that will produce accurate statistics about the energy consumption of vehicles in road transportation sector, each vehicle must have a positive probability or chance (equal or unequal) of being selected in the sample.

The big drawback of quota sampling is that it is unknown if the units chosen for the sample exhibit selection bias. If the selection of units is totally up to the interviewer, she/he is likely to choose the most accessible members of the population. The most accessible members of a population are likely to differ in a systematic way from less accessible members. Thus, unlike in stratified random sampling, one cannot say that the estimator of the total population from quota sampling is unbiased over repeated sampling—one of the usual criteria of goodness in probability samples.

In fact, in quota samples, the sampling error over repeated samples cannot be measured and there is no way of estimating the bias from the sample data. Since the selection of units is up to the individual interviewer, one cannot expect that repeating the sample will give similar results. Thus, anyone drawing inferences from a quota sample must necessarily take a **model-based approach** (Lohr, 2009).

Under model-based sampling, a model is used to define the distribution of the target population (Stephenson, 1979) with respect to the variables of interest. For example, in the ECSTS, the variable of interest is energy consumption in transport sector, in particular road transport sector.

The model is usually defined by quotas for subgroups or cells based on the cross-classification of known information relevant to the outcome of interest. Examples of quota cells include geographic region by vehicles type and age of vehicles. Moser and Stuart (1953) point out that the quotas can be either "independent", which means that the quotas are based on the marginal distribution, or "interrelated", which means that the quota requirements are made for each cross-classified subgroup. In either case no frame is explicitly required; however, knowledge of the population of interest is required for proper specification of the sampling distribution (Deville, 1991; Moser, 1952).

If the model assumptions hold, there is no bias in the estimates produced (Deville, 1991). Moreover, the model-based sampling paradigm does not require known selection probabilities or even random sampling. Once the quotas are defined, essentially any sampling method can be used to identify and select sample members for each quota cell (Moser, 1952). Thus, the requirements for data collection differ greatly between the two paradigms (quota sampling and probability sampling). Under the traditional paradigm (probability-based method), rigid controls of field procedures are specified so that the sampling instructions are properly executed and any interviewer effects on responses are minimized.

In carrying out the sampling instructions, interviewers must complete data collection on the entire sample, regardless of the achieved response rate, and conduct callbacks sufficient to reduce the proportion of non-respondents and minimize the impact of non-response on the survey results (King, 1985). Conversely, the model-based sampling paradigm (e.g. quota sampling) allows data collection to stop in a particular quota cell once the quota is met. In addition, interviewers are allowed great flexibility in how they collect the data. Callbacks and other attempts to re-contact non-respondents are not required, so long as the quota requirements are achieved (Moser, 1952).

Quota sampling has been repeatedly criticized. However, its advantages go beyond these critics. According to Moser (1952):

(a) Sampling errors are of comparatively small importance than the very considerable and intractable non-sampling errors which arise in the collection of the data.

(b) The main argument for quota sampling is that it is very cheap. This cheapness is due largely to the much lower travelling costs and to the lack of call-backs.

(c) Quota sampling is much easier from an administrative viewpoint. There is no need to go through the tedious stage of drawing the sample; no problems of non-contacts, call-backs, or substitute lists; no apparent problem of refusals; no need, unless this is specially required, to do much evening interviewing or to send field-workers to out-of-the-way areas. Responsibility for the sample is largely transferred to the individual interviewer, and the office burden is consequently lightened. Interviewers with experience of random and quota sampling usually prefer the more elastic, less controlled and less tiring quota method.

(d) If the field-work on a survey has to be completed within a very short period of time, one day-quota sampling may be the only feasible method.

(e) Quota samplers generally believe that instructions to and constraints on interviewers are sufficient to guard against the main dangers of bias, but they mostly agree that this is a matter of belief rather than fact. Quota sampling is also defended on the grounds that, although the sample may be biased with regard to certain characteristics, it may be quite satisfactory for others.

(f) It is independent of the existence of lists. As long as good lists-suitable for sampling individuals and households-are available, this is not a point of substance. But if, for instance, the National Register were to be withdrawn or to lose its present accuracy, there would then remain only the alternatives of either using one of the other, and less satisfactory, lists; of using area sampling; or of using quota sampling-if this technique has been shown to be of any reliability.

Quota sampling practice (Moser, 1952):

- **Quota controls:** The standard control (stratification variables) used in energy consumption survey may be: region, age, type of vehicle, and power engine, etc., depending on the data available at the population level.
- **Quota scheme:** There are two alternatives; the controls may be inter-related (cross-classification of vehicle type and age categories) or they may be set independently (vehicle type only).
- **Filling of the quota:**
  - **Location of interviews:** interviews may be made-unless instructions to the contrary are given- at gas stations, vehicle licensing offices, vehicles stations, in offices, factories, parks, public places or in the street.
  - **Spread of interviews:** It is a common criticism of quota sampling that it is unlikely to achieve an adequate geographical spread within the areas-districts, towns, etc., selected for sampling. This point, furthermore, is a source of worry to some market research practitioners, and various attempts are made to secure an adequate spread.
  - **Timing of interviews:** No definite rules are imposed for the timing of interviews.

- **Field organization:**
  - **Interviewers:** Interviewers are preferred to be selected from the local community, i.e., people residing in the town or in the vicinity of the town of work. In other cases, interviewers are selected on purpose to be not from the local community. Eventually, the purpose of the survey imposes whether selected interviewers can be local or not.

- **Training and Supervision:** Training mainly consists of supervised interviewing. It is generally felt that one of the main values of supervision is the psychological support it gives to the field-staff. Interviewers are greatly helped in their work by the knowledge that supervisors provide, to solve any difficult problems and so on.
- It would be useful to evaluate the importance of training and of experience of interviewers regarding the speed and accuracy of quota completion and the representativeness of the final sample.

- **Checks:** A point of great importance is the validity of honesty and accuracy of field-workers. Every organization makes attempts at checks as much in the hope of deterring the cheat as of detecting it. Some checking is possible if the respondents can be traced, and all the organizations now ask for the names and addresses of interviewees (except, in the case of one organization, on political surveys). Good checks are especially vital on two questions: whether the interview actually took place, and whether any distortion of classification data took place in order to fill the quotas. Checks on the quality of the interview are, of course, needed in all types of sampling, and are not a special concern of this research.
- **Actual Field Practice:** Generally, interviewers are free in deciding their work methodologies, i.e. how they plan their quota in the first place, daily planning, quota division by area and the decision behind it, etc. The only appropriate manner to facilitate the interviewer's job and be of desired quality is for field supervisors to train and convey the required/ preferable field-work method.

The main impression is on how much individual judgment freedom is granted for interviewers after allowing for the various controls and instructions. By and large, each interviewer can choose whether to cover a small or a large area; whether they interview in the street, in factories or in the gas station; in the morning, afternoon or evening. The quota cells are generally fairly wide, so that interviewers experience little difficulty early on, although the cases at the end may be quite hard to find. There are two different approaches to the task:

First approach: Some interviewers like to spend the first day or two of their assignment looking for the types of persons whom they know is difficult to find, and hope to spend the rest of their time filling in the easier cells in a more leisurely way.

Second approach: Choose the easy quota cells during the first few working days, and then have plenty of time to the difficult quota cells. On this approach, investigators tend to start with some street interviewing. During the early stage the investigator is also on the look-out for persons to put into the difficult cells. It has not been easy to find out which the generally difficult cells are, and no figures are available to show how long it takes investigators to complete various parts of an assignment.

3.3. Estimators

Post-stratification assigns weights to sample units by using an auxiliary variable *x*. Basically this variable provides ancillary information about the target population that can be easily retrieved mostly from existing data sources like registers. The ratio and regression estimators are estimators coming under this category and have been developed on the assumption that there is a linear correlation between the survey or study variable and the auxiliary variable.

According to Cochran (1977), the **ratio** and **regression estimators** are the most widely used methods of estimating finite population parameters, such as population mean and population total in survey sampling. These two techniques are based on auxiliary information or **auxiliary variable or subsidiary variable** – any variable about which information is available prior to sampling. The value of the auxiliary variable x, is known for each unit of the population elements so that the values $x_1,\ldots,x_N$ are available prior to sampling. The goal of using estimators is to increase the accuracy of the survey results.

Cochran (1940) was the first to show the contribution of known auxiliary information in improving the efficiency of the estimator of the population mean and total population in survey sampling.

3.3.1. Ratio estimators

When we trying to decide what kind of estimator to use, the first step is to graph sample values $\{(x_i, y_i), i \in s\}$. If the relationship between y_i and x_i is a straight line through the origin and the variance of y_i about this line is proportional to x_i, then the ratio estimator is the best (Cochran, 1977).

a. Simple random sampling

According to Cochran (1977) and Lohr (2009), $U = \{1,..., N\}$ denotes a finite population consisting of N units. Let y be the target or study variable of interest and let y_i be the value of y for the i^{th} population unit. At this stage the values y_i are assumed to be fixed unknown quantities. Suppose that an estimate is needed for the population mean $\bar{Y}$ or total population Y of y. A probability sample s is drawn from U according to a specified sampling design such as SRSWR, SRSWOR, systematic sampling, unequal probability of selection sampling, and stratified sampling. The sample size is denoted by n. Let $x_1,..., x_N$, $i \in U$ be the values of an auxiliary variable x, correlated with y. Assume that the population total $X = \sum_{i=1}^{N} x_i$ or population mean $\bar{X} = X/N$ of the auxiliary variable x is known. Let (x_i, y_i), $i \in s$ be the available sample values of (x, y).

Let: $S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})$ is the **population covariance** between x and y,

$\rho = \frac{S_{xy}}{S_x S_y}$, which is the **population Pearson correlation coefficient** of the sample between x and y,

$\bar{x}$ is the sample mean of the auxiliary variable,

$\bar{y}$ is the sample mean of the survey variable y,

$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ is the **sample covariance** between x and y, and

$\hat{\rho} = \frac{s_{xy}}{s_x s_y}$ is the **Pearson correlation coefficient** between x and y.

Ratio estimator of $Y = \sum_{i=1}^{N} y_i$:

The population total $Y = \sum_{i=1}^{N} y_i$ of the survey variable y can be written as:

$$Y = X\frac{Y}{X} = RX$$

Where $R = Y/X$ is the ratio of the population totals Y and X .

If the population total X of auxiliary variable x is known, and R is estimated by $\hat{R} = \bar{y}/\bar{x} = y/x$, then the **ratio estimator** of the population total Y is defined by:

$$\hat{Y}_R = \hat{R}X = N\bar{y}\frac{\overline{X}}{\bar{x}}$$

Similarly, the **ratio estimator** of the population mean $\overline{Y}$ is defined by:

$$\hat{\overline{Y}}_R = \hat{R}\overline{X} = \frac{\bar{y}}{\bar{x}}\overline{X} = \bar{y}\frac{\overline{X}}{\bar{x}}$$

Where $\hat{R} = \bar{y}/\bar{x}$ is the estimator of the ratio $R = \overline{Y}/\overline{X}$.

Note that, if x_i is the value of y_i at some previous time the ratio method uses the sample to estimate the relative change $R = \overline{Y}/\overline{X} = Y/X$ that has occurred since that time. The estimated relative change $\hat{R} = \bar{y}/\bar{x}$ is multiplied by the known population total X on the previous occasion to provide an estimate of the current total population Y .

Statistical Properties of $\hat{Y}_R$ and $\hat{\overline{Y}}_R$:

With large n, $\hat{Y}_R$ and $\hat{\overline{Y}}_R$ are approximately unbiased with:

$$V\left(\hat{Y}_R\right) = N^2\frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - Rx_i\right)^2 = N^2\frac{1-f}{n}\left(S_y^2 + R^2S_x^2 - 2RS_{xy}\right)$$

$$\hat{V}\left(\hat{Y}_R\right) = N^2\frac{1-f}{n}\left(s_y^2 + \hat{R}^2s_x^2 - 2\hat{R}\hat{\rho}s_xs_y\right)$$

$$V\left(\hat{\overline{Y}}_R\right) = \frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - Rx_i\right)^2 = \frac{1-f}{n}\left(S_y^2 + R^2S_x^2 - 2RS_{xy}\right)$$

$$\hat{V}\left(\hat{\bar{Y}}_R\right)=\frac{1-f}{n}\left(s_y^2+\hat{R}^2 s_x^2-2\hat{R}\hat{\rho}s_x s_y\right)$$

If $x_i = A$ for all $i=1,\ldots,N$, then $\hat{\bar{Y}}_R=\hat{\bar{Y}}=\bar{y}$ and $V\left(\hat{\bar{Y}}_R\right)=V_{Ran}\left(\hat{\bar{Y}}\right)$.

Efficiency comparison: The ratio estimator of Y is more efficient than the SRSWOR based estimator of Y, that is $V\left(\hat{Y}_R\right)\le V_{Ran}\left(\hat{Y}\right)$, if and only if $\rho=\frac{S_{xy}}{S_x S_y}\ge\frac{C_x}{2C_y}$.

b. Stratified random sampling

In stratified random sampling, the finite population of N sampling units is divided into H strata with N_h sampling units in the h^{th} stratum, where $h=1,\ldots,H$ and $N_1+\ldots+N_H=N$. An SRSWOR is taken independently of size n_h from stratum h, such that $n_1+\ldots+n_H=n$. Let the j^{th} sample unit of the study variable and auxiliary variable in the h^{th} stratum be denoted by y_{hj} and x_{hj}, respectively. The use of auxiliary information under stratified random sampling gives two ratio estimators namely: **separate ratio estimator** and **combined ratio estimator** (Cochran, 1977 and Lohr, 2009).

1. Separate ratio estimator

According to Cochran (1977) and Lohr (2009), the estimation of the total population is done by using the ratio estimator per stratum and adding the total of all strata. Assume that the separate strata totals X_h are known.

The **separate ratio estimator** $\hat{\bar{Y}}_{RS}$ of the population mean $\bar{Y}=\sum_{h=1}^{H}W_h\bar{Y}_h$ is given by:

$$\hat{\bar{Y}}_{RS}=\sum_{h=1}^{H}W_h\hat{\bar{Y}}_{Rh}=\sum_{h=1}^{H}W_h\frac{\bar{X}_h}{\bar{x}_h}\bar{y}_h$$

The **separate ratio estimator** $\hat{Y}_{RS}$ of the population total $\bar{Y}=\sum_{h=1}^{H}N_h\bar{Y}_h$ is given by:

$$\hat{Y}_{RS}=N\hat{\bar{Y}}_{RS}=\sum_{h=1}^{H}N_h\frac{\bar{y}_h}{\bar{x}_h}\bar{X}_h$$

Statistical Properties of $\hat{\bar{Y}}_{RS}$:

If independent SRSWOR is selected in each stratum and $n_1,\ldots,n_H$ are large, then:

(a) $\hat{\bar{Y}}_{RS}$ is approximately unbiased estimator of $\bar{Y}$.

(b) The variance of the separate ratio estimator $\hat{\bar{Y}}_{RS}$ of the population mean $\bar{Y}$ is given by:

$$V\left(\hat{\bar{Y}}_{RS}\right)=\sum_{h=1}^{H}W_h^2\frac{1-f_h}{n_h}\left(S_{yh}^2+R_h^2S_{xh}^2-2R_h\rho_h S_{xh}S_{yh}\right)$$

Where $R_h=Y_h/X_h$ and $\rho_h=S_{xyh}/S_{xh}S_{yh}$ - the correlation coefficient in stratum h.

(c) $\hat{V}\left(\hat{\bar{Y}}_{RS}\right)=\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}\left(s_{yh}^2+\hat{R}_h^2 s_{xh}^2-2\hat{R}_h s_{xyh}\right)$.

2. Combined ratio estimator

According to Cochran (1977), one starts with estimating the total population or population means of each variable using the SRSWOR estimator, and then takes the ratio estimator of the total population or population mean. Assume that population total X is known.

In each stratum, the SRSWOR estimators of the population means $\bar{Y}$ and $\bar{X}$ are the stratified sample means of the variables y and x are given by:

$$\hat{\bar{Y}}=\bar{y}_{st}=\sum_{h=1}^{H}W_h\bar{y}_h\ ,\ \hat{\bar{X}}=\bar{x}_{st}=\sum_{h=1}^{H}W_h\bar{x}_h$$

Similarly, for the population totals:

$$\hat{Y}=N\bar{y}_{st}=\sum_{h=1}^{H}N_h\bar{y}_h\ ,\ \hat{X}=N\bar{x}_{st}=\sum_{h=1}^{H}N_h\bar{x}_h$$

The **combined ratio estimator** $\hat{\bar{Y}}_{RC}$ of the population mean $\bar{Y}$ is given by:

$$\hat{\bar{Y}}_{RC}=\frac{\hat{\bar{Y}}_{st}}{\hat{\bar{X}}_{st}}\bar{X}=\frac{\sum_{h=1}^{H}W_h\bar{y}_h}{\sum_{h=1}^{H}W_h\bar{x}_h}\bar{X}=\hat{R}_C\bar{X}$$

Where: $\hat{R}_C=\frac{\bar{y}_{st}}{\bar{x}_{st}}$.

The **combined ratio estimator** $\hat{Y}_{RS}$ of the population total Y is given by:

$$\hat{Y}_{RC}=\frac{\hat{Y}_{st}}{\hat{X}_{st}}X=\frac{\sum_{h=1}^{H}N_h\bar{y}_h}{\sum_{h=1}^{H}N_h\bar{x}_h}X$$

Statistical Properties of $\hat{\bar{Y}}_{RC}$: If the total sample size n is large, then

(a) $\hat{\bar{Y}}_{RC}$ is approximately unbiased estimator of the population mean $\bar{Y}$.

(b) The variance of the combined ratio estimator $\hat{\bar{Y}}_{RC}$ of the population mean $\bar{Y}$ is given by:

$$V\left(\hat{\bar{Y}}_{RC}\right)=\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}\left(S_{yh}^2+R^2S_{xh}^2-2R\rho_h S_{xh}S_{yh}\right)$$

(c) $\hat{V}\left(\hat{\bar{Y}}_{RC}\right)=\sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h}\left(s_{yh}^2+\hat{R}^2 s_{xh}^2-2\hat{R}s_{xyh}\right).$

The combined ratio estimator is advisable to be used when there is a small sample in each stratum. If the sample is large enough for the variance formula to be applied to each stratum, it is advisable to use the separate ratio estimate (Cochran, 1977).

3.3.2. Regression Estimators

The **regression estimator** is used when the relationship between y_i and x_i is a straight line that does not go through the origin. In such cases, it is suggested to use an estimator based on simple linear regression model of y_i on x_i rather than the ratio of the two variables (Cochran, 1977).

a. Simple random sampling

The **linear regression estimator** of the population total $\bar{Y}$ is defined by:

$$\hat{\bar{Y}}_{\text{Reg}}=\bar{y}_{\text{Reg}}=\bar{y}+b\left(\bar{X}-\bar{x}\right)=\hat{\bar{Y}}+b\left(\bar{X}-\hat{\bar{X}}\right)$$

Where b is an estimator of the change in the study variable y per unit change in the auxiliary variable x, which is given by:

$$b=\hat{B}=\frac{s_{xy}}{s_x^2}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

Similarly, the **linear regression estimator** of the population total Y is defined by:

$$\hat{Y}_{\text{Reg}}=N\hat{\bar{Y}}_{\text{Reg}}=N\bar{y}+b\left(X-N\bar{x}\right)=\hat{Y}+b\left(X-\hat{X}\right)$$

The rationale of the linear regression estimator of the population mean is that if the sample mean $\bar{x}$ of the auxiliary variable x is below the population mean $\bar{X}$, it is expected also that $\bar{y}$ is below average by an amount $b\left(\bar{X}-\bar{x}\right)$ because of the simple linear regression of y on x.
Note that:

(a) If $b=0$, then $\hat{\bar{Y}}_{\text{Reg}}=\bar{y}$.

(b) If $b=\bar{y}/\bar{x}$, then $\hat{\bar{Y}}_{\text{Reg}}=\bar{y}+\frac{\bar{y}}{\bar{x}}\left(\bar{X}-\bar{x}\right)=\frac{\bar{y}}{\bar{x}}\bar{X}=\hat{\bar{Y}}_{\text{R}}$.

(c) If $b=1$, then $\hat{\bar{Y}}_{\text{Reg}}=\bar{y}_{\text{Reg}}=\bar{y}+\left(\bar{X}-\bar{x}\right)=\bar{X}+\left(\bar{y}-\bar{x}\right).$

When b is unknown, the **Statistical Properties of the regression estimator** for large sample size n is:

(a) $\hat{\bar{Y}}_{\text{Reg}}$ is asymptotically unbiased estimator of $\bar{Y}$.

(b) The variance of $\hat{\bar{Y}}_{\text{Reg}}$ is:

$$V\left(\hat{\bar{Y}}_{\text{Reg}}\right)=\frac{1-f}{n}S_y^2\left(1-\rho^2\right)$$

(c) $\hat{V}\left(\hat{\bar{Y}}_{\text{Reg}}\right)=\frac{1-f}{n}s_y^2\left(1-\hat{\rho}^2\right)$.

This shows that, for large sample size, the linear regression estimator $\hat{\bar{Y}}_{\text{Reg}}=\bar{y}_{\text{Reg}}$ is always more efficient than the sample mean per unit $\hat{\bar{Y}}=\bar{y}$ if $\hat{\rho}\neq 0$.

b. Stratified random sampling

The finite population of N sampling units is divided into H strata with N_h sampling units in the h^{th} stratum, where $h=1,\ldots,H$ and $N_1+\ldots+N_H=N$. SRSWOR is independently taken of size n_h from stratum h, such that $n_1+\ldots+n_H=n$. Let the j^{th} sample unit of the study variable and auxiliary variable in the h^{th} stratum be denoted by y_{hj} and x_{hj}, respectively. The use of auxiliary information under stratified random sampling gives two regression estimators namely: **separate regression estimator** and **combined regression estimator**.

1. Separate linear regression estimator

According to Cochran (1977), one starts with estimating the total population or population means of each variable using the linear regression estimator, and then takes the weighted sum of the linear regression estimators of the total population or population mean. Assume that the separate strata totals X_h are known.

The **separate linear regression estimator** $\hat{\bar{Y}}_{\text{RegS}}$ of the population mean $\bar{Y}=\sum_{h=1}^{H}W_h\bar{Y}_h$ is given by:

$$\hat{\bar{Y}}_{\text{RegS}}=\sum_{h=1}^{H}W_h\hat{\bar{Y}}_{\text{Reg}h}=\sum_{h=1}^{H}W_h\left\{\bar{y}_h+b_h\left(\bar{X}_h-\bar{x}_h\right)\right\}$$

Where b_h is the within-stratum least square estimator of the true or population regression coefficient $B_h=S_{xyh}/S_{xh}^2$, and given by:

$$b_h=\frac{\sum_{i=1}^{n}\left(x_{hj}-\bar{x}_h\right)\left(y_{hj}-\bar{y}_h\right)}{\sum_{i=1}^{n}\left(x_{hj}-\bar{x}_h\right)^2}=\frac{s_{xyh}}{s_{xh}^2}$$

If independent SRSWOR is selected in each stratum and $n_1,\ldots,n_H$ are large, then:

(a) The variance of the separate linear regression estimator $\hat{\bar{Y}}_{\text{RegS}}$ of the population mean $\bar{Y}$ is given by:

$$V\left(\hat{\bar{Y}}_{\text{RegS}}\right)=\sum_{h=1}^{H}W_h^2\frac{1-f_h}{n_h}\left(S_{yh}^2+b_h^2S_{xh}^2-2b_hS_{xyh}\right)$$

(c) $\hat{V}\left(\hat{\bar{Y}}_{\text{RegS}}\right)=\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}\left(s_{yh}^2+b_h^2 s_{xh}^2-2b_h s_{xyh}\right)$

2. Combined linear regression estimator

According to Cochran (1977), one starts with estimating the total population or population means of each variable using the SRSWOR estimator, and then takes the linear regression estimator of the population mean. Assume that population total X is known. Under SRSWOR in each stratum, the estimators of the population means $\bar{Y}$ and $\bar{X}$ are the stratified sample means of the variables y and x, and are given by:

$$\hat{\bar{Y}}=\bar{y}_{st}=\sum_{h=1}^{H}W_h\bar{y}_h\,,\ \hat{\bar{X}}=\bar{x}_{st}=\sum_{h=1}^{H}W_h\bar{x}_h$$

The **combined linear regression estimator** $\hat{\bar{Y}}_{\text{Reg}C}$ of the population mean $\bar{Y}$ is given by:

$$\hat{\bar{Y}}_{\text{Reg}C}=\bar{y}_{st}+b\left(\bar{X}-\bar{x}_{st}\right)$$

Where b is the weighted mean of the stratum sample regression coefficients b_h and is given by:

$$b=\frac{\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}s_{xyh}}{\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}s_{xh}^2}$$

If independent SRSWOR is selected in each stratum and $n_1,\ldots,n_H$ are large, then:

(a) The variance of the combined linear regression estimator $\hat{\bar{Y}}_{\text{Reg}C}$ of the population mean $\bar{Y}$ is given by:

$$V\left(\hat{\bar{Y}}_{\text{Reg}C}\right)=\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}\left(S_{yh}^2+b^2S_{xh}^2-2bS_{xyh}\right)$$

(c) $\hat{V}\left(\hat{\bar{Y}}_{\text{Reg}C}\right)=\sum_{h=1}^{H}W_h^2\,\frac{1-f_h}{n_h}\left(s_{yh}^2+b^2s_{xh}^2-2bs_{xyh}\right)$.

PART II

4. Preferred design for energy consumption surveys

Two elements are required to select a sample in a scientifically justified way: a sampling design based on probability sampling and a sampling frame (Bethlehem, 2009). A sampling frame is any material or device used to obtain observational access to the finite population of interest. It must be possible with the aid of the frame to: (a) identify and select a sample in a way that respects a given probability sampling design, and (b) establish contact with selected elements – by telephone, visit at home, mailed questionnaire, etc. In order to select a statistically valid sample that will produce accurate statistics about the energy consumption of vehicles in road transportation sector, each vehicle must have a positive probability or chance (equal or unequal) of being selected in the sample.

There are two sampling methods to choose from when implementing a survey, depending on the availability of the frame. The probability-based sampling method is used when the sampling frame is available, a list of the items or elements to survey. The selective probability methods are explained in section 2 of this report. Once the frame is not available, a model-based method, for example quota sampling is implemented. Although the frame is not available, knowledge of the population of interest is required for proper specification of the sampling distribution (Deville, 1991; Moser, 1952). Quota sampling is quite similar to stratified sampling (probability sampling method), where elements of judgments are used to select the sample (Doherty, 1994).

Under model-based sampling, a model is used to define the distribution of the target population (Stephenson, 1979) with respect to the variables of interest, for example, in the ECSTS, the variable of interest is energy consumption in transport sector, in particular road transport sector. The model is usually defined by quotas of subgroups or cells based on the cross-classification of known information relevant to the outcome of interest. Examples of quota cells include geographic region by vehicles type and age of vehicle.

In addition, quota sampling is very flexible with non-responses; when a non-response is obtained, interviewers move to the next sample until the quota is met. Interviewers are allowed great flexibility in how they collect the data. Callbacks and other attempts to re-contact non-respondents are not required, so long as the quota requirements are achieved (Moser, 1952).While comparing probability-based methods and quota sampling, the latter's financial cost is very minimal compared to probability-based, which can play a very important role in financially limited surveys (Doherty, 1994).

In ECSTS project, known frames are not available, and it is very expensive and time-consuming to construct a frame that lists all vehicles with respective contact information. Given this, the quota sampling method has been selected in the 3 selected countries to collect information on the energy consumption in the transport survey.

In the ECSTS, the construction of quotas within governorate or municipality is adopted based on the type of vehicles and age of vehicles. Quotas are made for each cross-classified subgroup of vehicle type and age of vehicles within governorates or municipalities. A frame or another external source of information can be used for this purpose. Because a predefined model is being

used to determine the sampling distribution of respondents, there are no coverage requirements for the sample. If the model assumptions hold, there is no bias in the estimates produced (Deville, 1991).

4.1. Sampling design

This section explains briefly the sampling design needed while conducting a survey for energy consumption in the transport sector. The section only covers the design for road transportation. Surveys for maritime, air and railways transportations are usually based on complete enumeration of the sub-sectors. This is due to their reliance on national administrative registries.

- **Sector:** Road transportation.
- **Target population:** All registered vehicles owned by individuals, enterprises and governments.
- **Sampling frame:** Auxiliary variables: province, municipality, engine capacity, vehicle type, vehicle age, etc. No access established for contact with vehicles.
- **Population** is composed of a number of provinces (geographic region).
- **Sampling design:** Stratified two stage sampling (random and non-random).
- **Stratified variable:** Provinces

First stage: Select a sample of municipalities or governorates from each province with probability proportional to the number of vehicles of the municipalities.

Second stage: For each municipality or governorate, divide the target population into several sub-populations based on certain characteristics of vehicles that are correlated with the survey control variables. In this case, the variable is energy consumption. Since energy consumption is related to vehicles type (individual cars, motorcycles, light trucks…) and age (e.g. less than or equal to 10 years and greater than 10 years), therefore the control variables are: vehicle type and vehicle age (if available in the sampling frame).

Since quota sampling is **representative of** vehicle type and vehicle age (the population structure is close to the sample structure), investigators draw a quota sample such that **the proportion of sample units from subpopulation is close to that in the population**. In other words, the size of each quota sample is proportional to the size of the subpopulation.

Once the sample size and quota have been determined, there is a need to start filling in each quota (purposively or systematically). In order to fill the quotas, gas stations, vehicle licensing offices, motor vehicle inspection centers, etc., can be targeted in order to reach the required number of completed questionnaires. The selection of how to fill in the quota is left within the

hands of the investigators, and is based on financial and other managerial factors. A combination of methods can be used to fill the quotas, for example, in case of registered vehicles, the chosen sample can be extracted from vehicle licensing offices and motor vehicle inspection centers combined; whereas in the case of informal vehicles, those that are not officially registered, gas stations can be used instead to fill the required quota.

4.2. Use of International recommendations for energy statistics

As pointed by the International Recommendations for Energy Statistics (IRES) (United Nations, 2011) "In the context of preparing for national energy consumption survey, the IRES are considered as an essential tool to use for conducting national surveys on energy production and consumption. IRES cover various issues relevant to the collection, compilation and dissemination of energy statistics, among other issues related to classifications and definitions. The approach used to develop IRES relies on consistency between concepts and classifications used and other fields of economic statistics (such as ISIC, CPC and HS). It provides a flexible framework for data collection, compilation, analysis and dissemination for national official statistics in a timely, internationally comparable and reliable manner."

IRES contain main recommendations and encouragements (United Nations, 2011), some of which are summarized below:

- Ensuring that energy statistics are developed with the highest quality possible. To ensure that, countries are encouraged to progress from collecting selected data items used primarily for internal purposes by various national establishments, to establish an integrated system of multipurpose energy statistics as part of their official statistics;
- Disseminate energy statistics to the general public, as it is considered as public good.
- Scope of energy statistics focused on basic energy statistics and energy balances.
- Collaboration between energy data collection and other data collection activities undertaken within the country, to avoid duplication of the work and ensure overall coherence.
- The use of a unified energy unit from the International System of Units as joule, although other energy units are possible to use.
- The use of reference list of data items for selecting the data items, based on each country's specificity and context. The selected data items should represent the national context and allow for a national assessment in terms of energy.
- The cooperation between various national energy stakeholders to design and implement energy surveys due to the financial and human resources special allocation.
- Countries to disseminate yearly national energy balance, with a clearly defined reference period.
- Due to the importance of energy statistics for policy makers, it is advisable to use disaggregated data for the final energy consumption, to cover the smallest collected unit possible (e.g. iron and steel, transport equipment, etc.).
- Transport should be disaggregated by mode of transport: Domestic aviation, Road, Rail, Domestic navigation, Pipeline transport, and Transport none elsewhere specified.

- Countries are encouraged to: develop their own national energy data quality assurance programmes; document these programmes; develop measures of data quality; and make these available to users.
- National dissemination policy should be user oriented, targeting all users and providing quality information in a clear manner to the general public.

4.3. Weighting and drawing inferences from quota samples

According to Battaglia (2008), one issue that arises with all probability samples and for many non-probability samples is the estimation procedures, specifically those used to draw inferences from the sample to the population.

Many surveys produce estimates that are proportions or percentages (e.g., the percentage of commuters who use public transportation), and weighting methods used to assign a final weight to each completed interview are generally given considerable thought and planning.

For probability sampling, the first step in the weight calculation process is the development of a base sampling weight. The base sampling weight equals the reciprocal of the selection probability of a sampling unit. The calculation of the base sampling weight is then often followed by weighting adjustments related to non-response and non-coverage. Finally, post-stratification or raking is used to adjust the final weights so that the sample is in alignment with the population for key demographic and socioeconomic characteristics (Battaglia, 2008).

In non-probability sampling, the calculation of a base sampling weight is meaningless since there are no known probabilities of selection. One could essentially view **each sampling unit as having a base sampling weight of one.** In this case, the sample mean for quota h is:

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} w_{hi} y_{hi} ,$$

Where $w_{hi} = 1$ for all i and h.

And the post-stratified mean is:

$$\hat{\bar{Y}} = \bar{y}_{st} = \sum_{h=1}^{H} W_h \bar{y}_h$$

Where $W_h = N_h / N$

Sometimes non-response and non-coverage weights are developed for non-probability samples. However, the most common technique is to use a weighting procedure such as post-stratification or raking to align the non-probability sample with the population that one would ideally like to draw inferences from. The post-stratification variables are generally limited to demographic and socioeconomic characteristics.

One limitation of this approach is that the variables available for weighting may not include key characteristics related to the non-probability sampling mechanism that was employed to select the sampling units (Battaglia, 2008). The results of weighting non-probability samples have been

mixed in the situation when benchmarks are available for a key survey outcome measure (e.g., the outcome of an election).

4.3.1. Post-stratification Weight

When conducting a survey, having a representative sample of the population is of paramount importance. But sometimes, despite best efforts, one may accidentally (or sometimes intentionally) oversample some kinds of vehicles and undersample others. In other words, the way a certain characteristic (such as vehicle age, vehicle engine capacity, etc.) of the sample is distributed may differ from its original distribution in the population. For example, the sample may consist of 60 percent vehicles with engine capacity greater than or equal to 1600 cubic centimeters (CC), and vehicles with engine capacity less than 1600 CC make up only 40 percent of the population. This introduces bias into any result estimation from the sampled data because statistical procedures will give greater weight to those vehicles that were oversampled. This can be corrected mathematically with a post-stratification survey weight.

4.3.2. Calculation post-stratification weight

In order to calculate a post-stratification weight, one needs an auxiliary variable to which sampled data can be compared to. For example, if one was conducting a survey on energy consumption of gas in the transport sector for private vehicles in a village, one would need census data that show the engine capacity characteristics of the population of all private vehicles in a village. The sample is then compared to the auxiliary data file, to make sure that the distribution of engine capacity characteristics of private vehicles is similar to that of the auxiliary data.

It is called a **post-stratification weight** because one can only compute it **after** data is collected. The stratification part comes from the fact that one uses various known strata (such as engine capacity groups for private vehicles in a village) of the population to adjust the sampled data to conform more to the population's parameters.

Table 4.1 shows an example on how to use post-stratification weight on the vehicle type private as quota *h* and the variable engine capacity as auxiliary variable, since it is correlated with energy consumption of the vehicles.

Table 4.1 Post-stratification weights

Engine Capacity Groups (Stratum)	Population Size N_h	Population Proportion P_h	Sample Size n_h	Sampling Sample p_h	Post-stratification Weight $pw_h = P_h / p_h$
Less than or equal to 1600	$N_1 = 500$	$P_1 = N_1/N = 0.5$	$n_1 = 50$	$p_1 = n_1/n = 0.25$	$pw_1 = P_1/p_1 = 2$
Between 1601 and	$N_2 = 300$	$P_2 = N_2/N = 0.3$	$n_2 = 100$	$p_2 = n_2/n = 0.5$	$pw_2 = P_2/p_2 = 0.6$

2500					
More than 2500	$N_3 = 200$	$P_3 = N_3/N = 0.2$	$n_3 = 50$	$p_3 = n_3/n = 0.25$	$pw_3 = P_3/p_3 = 0.8$
Total	$N = 1000$	1.00	$n = 200$	1.00	

Since $P_1 = N_1/N = 0.5$ is less than $p_1 = n_1/n = 0.25$ (table 4.1), therefore engine capacity group for private vehicles with less than or equal 1600 CC is **under represented** in the sample, while the other groups are **over represented** in the sample.

Under quota sampling, the rule of $w_{hi} = 1$ is assumed for all *h* and *i*. In this example, engine capacity group for private vehicles with less than or equal to 1600 CC is **up weighted** since the weight for this category of private vehicles is $pw_1 = P_1/p_1 = 2$. In addition to that, the other categories are **down weighted**, because the weights for these groups are $pw_2 = P_2/p_2 = 0.6$ and $pw_3 = P_3/p_3 = 0.8$.

So, the post-stratified estimate of the population mean of stratum *h* and overall population mean are:

$$\hat{\bar{Y}}_h = \bar{y}_{hpst} = pw_h\left(\frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}\right)$$

And $\hat{\bar{Y}} = \bar{y}_{pst} = \sum_{h=1}^{H} W_h \bar{y}_{hpst}$ respectively.

In addition to that, the post-stratified estimate of the population total of stratum *h* and overall total population are:

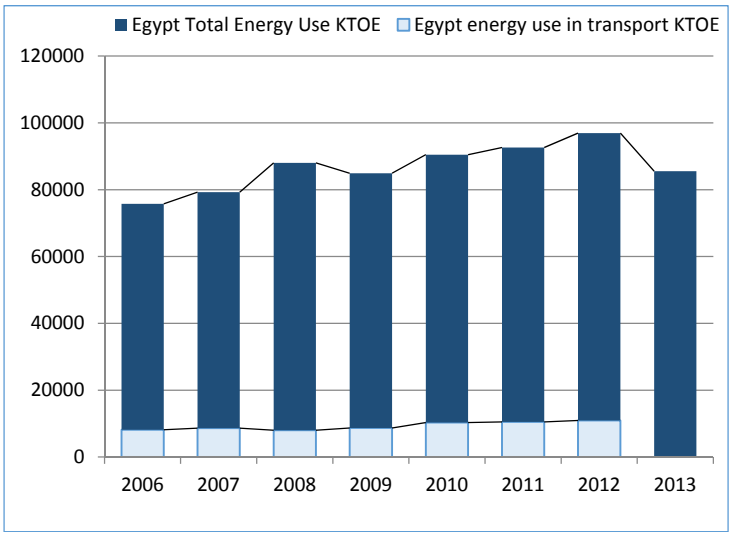
$$\hat{Y}_h = N_h \bar{y}_{hpst}$$

And $\hat{Y} = N\bar{y}_{pst}$ respectively.

5. Transport Energy Consumption survey – Egypt

The total end use of energy in Egypt amounted to 8,523 ktoe in 2014 as reported by the Central Agency for Public Mobilization and Statistics (CAPMAS) in the balance for Egypt. The energy use in the transport sector represented about 30 per cent of total petroleum products consumption. This annual rate of 5.2 per cent since 2006 grew faster than the 4.2 per cent increase of the total energy use, as shown in figure 5.1.

Figure 5.1 Egypt energy use in transport of total use 2006-2013



Sources: Statistical Yearbook, CAPMAS, 2013
ESCWA Statistical Abstract, 2014

Energy products consumed by the transport sector in Egypt comprise gasoline, diesel and natural gas. In addition, lube oils are used as lubricant for vehicles engines while fuel oil is used for road paving activity.

CAPMAS is the official governmental institution in Egypt conducting the national survey on energy consumption for the entire transport sector, namely, road, maritime, air and railways transport. This section of outlines the survey methodology used to conduct the energy consumption survey in the entire transport sector (CAPMAS, 2015). All questionnaires abide by the confidentiality of sharing information under the national statistics law and respondents were informed about this before starting to answer any question, along with the purpose of the questionnaire and its results usages.

5.1. Road transport sector

Egypt has a road network of a total length of 121.4 thousand km of which 108.8 thousand km are paved roads (90%) and 12.6 thousand km are unpaved roads (10%). The total number of vehicles in Egypt increased from 1.1 million in 1990 to 6.8 million vehicles in 2014 with an average annual growth rate of 8.8 per cent during that period. The private cars represent 51 per cent of total vehicles fleet compared to 5 per cent for taxis, 14 per cent for trucks, 2 per cent for buses, 28 per cent for motorcycles and Rickshaw (tuk tuk) and other types of vehicles. More than 51 per cent of the total vehicles fleet is operating in Cairo, Giza, and Alexandria provinces (CAPMAS, 2014). Of total vehicles fleet volume, 26 per cent of the fleet is of age of more than 27 years old and 25 per cent of age between 17 to 26 years, which result in inefficient fuel consumption and higher rates of pollutants emissions (Korkor, 2014).

Road transportation sector includes all engine-based vehicles and trailers designed for transportation of goods and passengers. It is necessary to distinguish between two types of ownership of vehicles, namely: vehicles owned by individuals and enterprises (including vehicles owned by enterprises, schools, government, tourists/ rentals, etc.).

5.1.1 Vehicles owned by individuals

Target population: all registered vehicles owned by individuals.

Sampling frame: the list of all registered vehicles owned by individuals, which contains the auxiliary variables: province, governorate, vehicle type, etc. The sampling frame does not contain personal information that enables direct contact with the owner or driver of the vehicles.

Sampling design: stratified two-stage sampling based on probabilistic and quota sampling methods. The **first stage sampling** is probabilistic and purposively sampling methods**.**

The subsequent allocations are performed independently for each control variable: province, governorate, and vehicle type. The stratification of the population by provinces is given in Table 5.1.

Table 5.1. Distribution of Car Fleet by Provinces in Egypt, 2014

Province	Governorate								Total
Greater Cairo	Cairo	Giza	Qalubia						
	2,073,596	847,860	308,373						3,229,829
North Ward	Domiat	Daqahlia	Sharqia	Karf Elshaiehk	Garbia	Minofia	Alexandria	Bahira	
	119,409	372,015	288,384	117,705	312,760	309,256	637,614	176,064	2,333,207
South Ward	Bany Souf	Fayoum	Minia	Assut	Suhaga	Qana	Aswan	Luxor	
	159,423	162,654	165,301	141,150	135,597	107,862	62,640	75,906	1,010,533
Qana	Port Saeed	Sweis	Elesmaelia						
	99,885	95,262	102,909						298,056
Borders	Red Sea	Awadi Gadid	Matrouh	North Saina	South Saina				
	51,817	30,027	38,802	19,770	25,913				166,329
Total									**7,037,954**

Sample size: based on budget considerations, the department of sampling and methodology and survey managers of the project decided that the total number of vehicles to be considered for the sample is 9,000 vehicles, which represent 0.001278781 of the total number of vehicles in the sampling frame. The 9,000 vehicles distributed over the provinces based on the proportional allocation method with the following calculations:

Province size sample= total sample size*(total province car fleet/ total car fleet of the country)

Using the given formula, the below results are obtained for determining the sample size for each province in Egypt.

Table 5.2. Sample Sizes of Provinces

Province	Greater Cairo	North Ward	South Ward	Qana	Borders	Total
Sample Size n_h	4,131	3,000	1,280	379	210	9,000

For example**,** the sample size of Greater Cairo is calculated as:

9,000*(3,229,829/7,037,954) = 4,131.

The sample size of governorates in some provinces (selection of number of governorates in each province) might be larger than others due to the high correlation between the selection criteria set by the survey (control variables) and the availability of these criteria in these provinces (for instance North and South Wards). The selection of the governorates from each province is performed by using the probability proportional to size method (PPS) to select the number of vehicles of the Governorate using systematic PPS method (Cochran, 1977). For other provinces, governorates are selected on purpose. The selection decision is based on opinions of which governorates are typical or representative in some sense or context. Table 5.3 summarizes the results obtained from stage one sampling.

Table 5.3. Sample of Governorates by Provinces

Province	**Governorate**			**Total**
Greater Cairo	Cairo	Giza	Qalubia	3
North Ward	Sharqia	Garbia		2
South Ward	Minia	Suhaga		2
Qana	Elesmaelia			1
Borders	Matrouh			1
Total	5	3	1	9

By using the proportional allocation method, the sample size of each province is allocated over the selected governorates, using the following calculations:

Sample size of governorate=sample size of province*(total governorate car fleet/total province car fleet)

For example, the sample size for Cairo is 4,131*(2,073,596/3,229,829) =2,641. For all the Egyptian provinces, the results are given in Table 5.4.

Table 5.4. Sample Allocation by Provinces for Selected Governorates

Province	**Governorate**			**Total**
Greater Cairo	Cairo	Giza	Qalubia	3
	2,641	1,090	400	4,131
North Ward	Sharqia	Garbia		2
	1,428	1,572		
South Ward	Minia	Suhaga		2
	700	580		1,280
Qana	Elesmaelia			1
	379			379
Borders	Matrouh			1
	210			210
Total				9,000

The **second stage sampling** is non-probability sample-quota sample. This stage includes considering a sample that is dependent on the vehicle type. A quota sample is representative with respect to vehicle type if the population distribution (population car fleet structure) is close to the sample distribution (sample structure). So, the investigator need to draw a quota sample such that the proportion of sample units from subpopulation is close to that in the population, that is, the size of each quota sample is proportional to the size of the subpopulation. The population frequencies of vehicle type and governorate are presented in Table 5.5.

This methodology (quota sampling) is often justified as a way to avoid the costly and time-consuming expense of listing all the vehicles in the sample governorates- as a prior stage before selecting the vehicles to be interviewed. It also avoids non-response since the interviewer continues questioning the vehicles' owners until she/he reaches the total number of the sample size.

Based on Tables 5.4 and 5.5, the sample size is allocated for each governorate for each vehicle type using proportional allocation approach. The resulting numbers are summarized in Table 5.6.

Table 5.5. Population Distribution of Number of Vehicles by Vehicle Type and Provinces in Egypt, 2014

	Vehicles			Buses					Truck and trail			Other Vehicles			Total
Governorate	**Private**	**Taxi**	**Total**	**Private**	**Tourism**	**Long Distance**	**Schools**	**Total**	**Freight**	**Trail**	**Total**	**Motor cycle**	**Rickshaw**	**Total**	
Cairo	1,479,082	76,513	1,555,595	10,472	10,344	4,565	5,898	31,279	152,583	9,438	162,021	267,703	0	267,703	2,016,598
Port Saeed	56,360	13005	69,36	299	109	440	20	868	8,682	1,538	10,220	17,275	0	17,275	97,728
Swis	44,935	4,414	49,349	336	17	669	24	1,046	7,578	640	8,218	29,023	0	29,023	87,636
Alexandria	444,242	23,203	467,445	3,194	411	9,841	1,037	14,483	85,443	10,045	95,488	49,986	0	49,986	627,402
Domiat	46,715	5,154	51,869	195	2	174	67	438	19,507	2,654	22,161	41,956	21	41,977	116,445
Daqahlia	126,813	15,223	142,036	1027	149	2505	303	3,984	72,874	10,894	83,768	123,013	10,438	133,451	363,239
Sharqia	86,247	13,970	100,217	3,722	441	2,355	494	7,012	70,499	3,838	74,337	86,603	7,241	93,844	275,410
Garbia	119,620	12,552	132,172	427	37	2,241	25	2,730	57,872	3,256	61,128	107,085	4,501	111,586	307,616
Qalubia	99,662	13,461	113,123	1,983	15	4,106	765	6,869	31,694	2,177	33,871	148,602	805	149,407	303,270
Karf Elshaiehk	34,764	6,512	41,276	66	55	121	6	248	33,239	1,863	35,102	30,641	3,538	34,179	110,805
Minofia	52,945	13,379	66,324	405	6	2,123	74	2,608	36,396	3,126	39,522	191,562	2,978	194,540	302,994
Bahira	49,665	12,806	62,471	460	2	390	44	896	63,550	3,102	66,652	30,794	5,614	36,408	166,427
Elesmaelia	51,190	5,039	56,229	391	0	170	148	709	14,696	759	15,455	27,182	162	27,344	99,737
Giza	480,736	23,162	503,898	11,266	1,120	7,988	1731	22,105	100,467	5,297	105,764	196,666	7,514	204,180	835,947
Bany Souf	35,417	8,606	44,023	249	4	182	28	463	26,198	702	26,900	85,100	25	85,125	156,511
Fayoum	28,736	11,894	40,630	190	2	311	46	549	18,938	1,709	20,647	94,749	153	94,902	156,728
Minia	38,690	8,850	47,540	352	7	340	127	826	39,563	1,226	40,789	66,106	3,690	69,796	158,951
Assut	50,745	8,546	59,291	216	4	124	59	403	26,519	1,791	28,310	45,684	1,654	47,338	135,342
Suhaga	36,131	11,576	47,707	102	3	86	125	316	36,343	1,501	37,844	45,002	993	45,995	131,862
Qana	26,707	7,955	34,662	233	15	78	15	341	19,659	806	20,465	48,778	325	49,103	104,571
Aswan	18,960	5,357	24,317	417	350	94	41	902	15,086	327	15,413	18,331	408	18,739	59,371
Luxor	11,347	3,880	15,227	94	1,517	239	0	1,850	7,987	68	8,055	48,292	475	48,767	73,899
Red Sea	22,705	3,905	26,610	401	2,537	145	33	3,116	9,241	717	9,958	9,163	0	9,163	48,847
Awadi Gadid	5,826	1,299	7,125	44	36	95	0	175	3,788	186	3,974	16,982	0	16,982	28,256
Matrouh	9,372	4,264	13,636	75	14	11	23	123	13,471	659	14,130	3,428	0	3,428	31,317
North Saina	4,920	2,435	7355	48	3	18	3	72	7,414	168	7,582	3,481	0	3,481	18,490
South Saina	7,894	1,767	9,661	315	1,440	48	5	1,808	6,051	170	6,221	4,418	0	4,418	22,108
Total	**3,470,426**	**318,727**	**3,789,153**	**36,979**	**18,640**	**39,459**	**11,141**	**106,219**	**985,518**	**68,657**	**1,054,175**	**1,837,605**	**50,535**	**1,888,140**	**6,837,687**
Percentage	**51%**	**5%**	**55%**	**1%**	**0%**	**1%**	**0%**	**2%**	**14%**	**1%**	**15%**	**27%**	**1%**	**28%**	**100%**

Table 5.6. Sample Size of Vehicle Type by Governorate, Quota Sample

Governorates	Vehicles			Buses					Truck and Trail			Other Vehicles			*Total*
	Private	Taxi	Total	Private	Tourism	Long Distance	Schools	Total	Freight	Trail	Total	Motor cycle	Rickshaw	Total	
Cairo	2,097	109	2,206	17	17	7	10	51	198	12	210	147	0	147	2,641
Giza	749	36	785	16	1	10	2	29	131	7	138	128	10	138	1,090
Qalubia	217	30	247	3	0	5	1	9	42	3	45	98	1	99	400
Sharqia	635	104	739	26	3	16	3	48	358	19	377	226	38	264	1,428
Garbia	856	91	947	3	0	14	0	17	293	17	310	275	23	298	1,572
Minia	288	66	354	3	0	2	0	5	173	5	178	147	16	163	700
Suhaga	234	75	309	1	0	1	0	2	158	7	165	100	4	104	580
Elesmaelia	241	24	265	2	0	1	0	3	55	3	58	52	1	53	379
Matrouh	82	38	120	1	0	0	0	1	75	4	79	10	0	10	210
Total	**5,399**	**573**	**5,972**	**72**	**21**	**56**	**16**	**165**	**1,483**	**77**	**1,560**	**1210**	**93**	**1,303**	**9,000**

The classification or stratification for the governorates excluded public services automobiles of 119,783, diplomatic licenses automobiles of 6,443, costumes licenses automobiles of 38,672 and public transit buses of 16,434 buses. It also excluded other vehicles for non-transport related activities such as agriculture trucks of 16,893 and heavy equipments of 2,042 which will bring the total of excluded vehicles to about 200,567.

After constructing the quotas and determining the sample size, investigators would look at different options to reach the quota needed for the sample size. There are no specific rules as to how the quotas should be filled in quota sampling.

A questionnaire has been developed to collect specific type of information from the respondents related only to respondents owning private cars. Although the questionnaire is divided into three main sections, the same questionnaire is used to collect information on vehicles that transport passengers and freight. The first section targets general information about the respondents, such as geographical location, gender, age and other contact details that can be used to contact respondents in cases of incompleteness of questionnaire or additional clarifications. The second section collects information on the vehicle, among which are the general information on the vehicle pertaining to its type, engine size, registration number and maintenance operations. The third section is divided through two sub-sections; the energy consumption patterns of the vehicles is the first sub-section where information collected is related to the type of consumed fuel, quantity of fuel consumed and cost in local currency; while energy efficiency questions represents the second sub-section of which amount of fuel consumed per kilometer, average distance traveled per one liter of fuel…

5.1.2. Vehicles owned by Enterprises

This section covers all vehicles owned and operated by enterprises, government, schools, etc.

Target population: all registered enterprises.

Sampling frame: auxiliary variables: total number of vehicles for each enterprise, etc.

Sampling design: probability sample: stratified probability proportional to size sampling, where the size variable is the total number of vehicles for each enterprise.

Table 5.7. Population distribution of Business units, 2014

Stratum	Stratum Size
Points belonging to the administrative system of the state	614
Private companies for public transport of passengers	46

From each stratum, survey managers select a 5% of the stratum size, see table 5.8.

Table 5.8. Sample Size of Business units

Stratum	Stratum Size
Points belonging to the administrative system of the state	30
Private companies for public transport of passengers	05

The questionnaire prepared for the vehicles owned by enterprises is very similar to the questionnaire of the privately owned vehicles (section 5.1.1) in matter of divisional structure and information required. The major difference between the questionnaires pertains to the general information collected from the type of the institution owing the vehicle (governmental, private or public).

5.2. Maritime Transport Sector

Target population: all business units operation in the Nile River.

Sampling frame: no sampling frame is available, but the frequency distribution of business units with type of operator is selected**.**

Sampling design: non-probability quota sample, where the control variable is the type of business unit.

The maritime transport sector is divided into two sub-sectors: sea transport and river based transport. The sea transport questionnaire focuses on the type of ships, traveled distance and type, amount and cost of fuel consumed.

The data collected from river based transport is slightly different, where individual transports and enterprises transport data is collected separately through two questionnaires. The two questionnaires target the collection of data for both passenger and freight transport in Egyptian rivers. Both questionnaires are divided through three sections: general information (contact information, educational level and geographical location), ship information (manufacturing year, size, engine size and maintenance) and energy consumption and efficiency (type of consumed energy, quantity and cost, number of monthly working days, distance traveled and efficiency through liters per kilometer).

5.3. Railway transport sector

The railway sector in Egypt is divided through trains and subways. Since the operating enterprises of these two sub-sectors are limited, CAPMAS has decided to perform a full enumeration of this sector based on administrative records.

The railway transport questionnaire, in addition to the descriptive data that intend to collect personal information of respondents, targets energy consumption data for long distance trains,

short distance and freight trains. The main indicators include the number of trains, usages by residents, their economic added value, type of energy used and quantities.

The subway questionnaire collects statistical information on the three functional lanes of the Egyptian subway. The type of data collected pertains to the daily number of trains and trips, distance traveled and maintenance, and energy consumed per hour and cost of the energy consumed in local currency.

5.4. Air transport sector

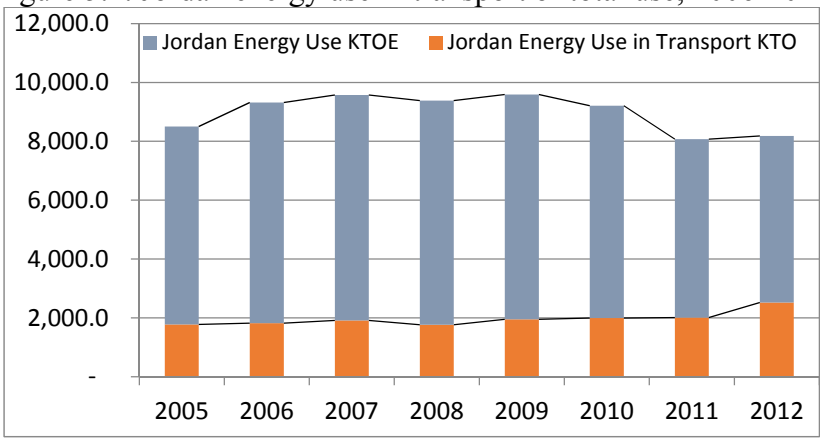
Similar to the maritime and railways sectors, the air transport sector has been fully enumerated based on administrative records due to the relatively small number of the operating enterprises.

The air transport sector questionnaire is divided into two sections. The general information section includes information about the company operating the airplanes; while the second section targets information related to the type and number of the aircrafts, quantity and cost of fuel consumed and maintenance cost in local currency. The questionnaire indicates whether the aircrafts are used for passengers or freight transport purposes.

6. Transport Energy Consumption survey- Jordan

The total end use of energy amounted to 5,095 ktoe as reported by the Department of Statistics (DOS) in the balance for Jordan in 2014. The energy use in the transport sector represented about 46 per cent of total energy end use (Fig. 5.2). In 2012, the energy consumed in transport augmented by 25 per cent compared to 4 per cent annual growth rate from 2006 to 2013. Energy products consumed by the transport sector in Jordan comprise Gasoline, Fuel oil, Diesel, Jet fuel.

Figure 5.2. Jordan energy use in transport of total use, 2006-2012



DOS is the official government institution in Jordan to conduct the national survey on energy consumption for the entire transport sector, namely, road transport, maritime, air and railways transport. All questionnaires abide by the confidentiality of sharing information under the national statistics law and respondents were informed about this before starting to answer any question.

6.1. Road transport sector

Road transportation sector includes all engine-based vehicles and trailers designed for transportation of freight and passengers.

Target population: all registered vehicles in Jordan owned by individuals and enterprises.

Sampling frame: the list of all registered vehicles owned by individuals, which contains the auxiliary variables: governorate, and vehicle type. The sampling frame does not contain personal information that enables direct contact with the owner or driver of the vehicles. The population frequencies of vehicle type and governorate are presented in Table 6.1.

Sampling design: stratified quota sampling method.

Sample size: the total sample size amounted to 10,400 vehicles, with about 7,400 vehicles selected from private and public vehicles and around 3,000 vehicles selected from government, judicial council and Aqaba Authority vehicles.

The sample size is allocated for each governorate for each vehicle type using proportional allocation approach (table 6.2).

The following vehicles are excluded from the sample:

1. Governmental vehicles which amount to 16,847.
2. Judicial council vehicles which amount to 8.
3. Aqaba Authority vehicles which amount to 372.

The above exceptions were made based on the following criteria:

1. Government vehicles sampling will amount to 20 per cent of all government vehicles, which is somehow higher than the sampling for the entire road sector.
2. Judicial council vehicles were excluded due to full enumeration.
3. Aqaba Authority vehicles were excluded due to full enumeration.

The data was collected from licensing centers in all governorates of the Kingdom; and in the absence of vehicles to cover all categories, data was collected from the vehicles parking stations.

Table 6.1. Population Distribution of Number of Vehicles by Type and Governorate in Jordan, 2014

Vehicle Type*	Governorate												Total
	Capital	Blqa	Zarqa	Madaba	Irbid	Mafraq	Jeresh	Ajlon	Qaraq	Tafila	Maan	Aqaba	
Small Private- Passengers	790,256	16,099	15,196	5,615	75,020	6,121	5,876	2,064	5,350	1,329	1,921	3,869	928,716
Small Public - Passengers	17,437	907	1,472	363	2,436	321	132	76	117	107	228	614	24,210
Intermediate Private - Passengers	8,807	1,230	568	205	848	115	267	72	201	67	57	195	12,632
Intermediate Public - Passengers	1,612	440	553	149	1,046	280	129	57	380	109	104	73	4,932
Private Bus	680	31	38	6	70	6	22	0	28	6	2	42	931
Public Bus	2,211	39	161	3	128	75	46	15	34	27	21	12	2,772
Freight	91,504	6,179	5,264	6,461	11,538	7,634	3,175	469	2,422	888	4,756	1,172	141,462
Big shipping	35,671	97	242	353	6,571	926	40	4	522	136	656	680	45,898
Joint transfer	91,443	2,887	2,668	1,610	10,081	4,652	1,734	891	1,390	449	1,279	882	119,966
Motorcycles	6,585	14	25	10	84	26	32	2	11		6	103	6,898
Construction vehicles	9,443	370	388	76	767	140	129	73	296	114	265	348	12,409
Agricultural vehicles	3,814	857	133	588	2,925	618	348	116	638	88	207	102	10,434
Total	1,059,463	29,150	26,708	15,439	111,514	20,914	11,930	3,839	11,389	3,320	9,502	8,092	1,303,168

* The vehicles type in Jordan is divided as follows:

a- Passenger cars include small and intermediate cars that can be used as private vehicles or for public transport (taxis). These include the above nomenclature: small private- passengers, small public- passengers, intermediate private- passengers and intermediate public- passengers.

b- Buses include two categories: public and private buses.

c- Freight transport includes freight and big shipping.

d- Joint transfer refers to the vehicles with double usages purposes, i.e. passengers and freight simultaneously.

e- Construction vehicles are those whose sole purpose is to be used in construction area for transport of personnel or for transport of construction goods.

f- Agricultural vehicles are those whose sole purpose is used in transporting goods or serving as equipment for agricultural purposes.

Table 6.2. Sample Distribution of Number of Vehicles by Vehicle Type and Governorate in Jordan, 2014

Vehicle Type	Governorate												Total
	Capital	Blqa	Zarqa	Madaba	Irbid	Mafraq	Jeresh	Ajlon	Qaraq	Tafila	Maan	Aqaba	
Small Private-Passengers	900	400	400	300	500	300	300	180	300	150	150	150	4,030
Small Public - Passengers	200	50	60	30	80	30	20	15	20	20	30	60	615
Intermediate Private - Passengers	110	50	30	15	60	7	20	10	20	10	6	15	353
Intermediate Public - Passengers	20	40	50	15	20	20	10	10	40	20	20	7	272
Private Bus	70	3	3	3	3	3	3	0	3	3	2	4	100
Public Bus	50	3	10	3	3	4	4	5	3	5	3	3	96
Freight	100	50	50	50	50	50	30	20	30	20	50	25	525
Big shipping	50	10	15	20	100	50	3	4	25	10	25	25	337
Joint transfer	100	50	50	25	50	50	25	25	25	20	25	30	475
Motorcycle	65	3	3	3	3	3	3	2	3	0	3	10	101
Construction vehicle	95	15	20	5	10	10	10	10	10	10	20	35	250
Agricultural vehicle	40	20	6	20	40	20	20	15	20	10	20	10	241
Total	1,800	694	697	489	919	547	448	296	499	278	354	374	7,395

The Jordanian road transport questionnaire is divided into three main categories (DOS, 2015): the general information section that includes contact information of respondents, geographical details and gender while the second category included technical information about the vehicle, i.e. registration plate number, type of vehicle (passenger and/ or freight), engine size, fuel consumption types, and type of maintenance. The third category tackled the energy efficiency of the vehicle, based on the monthly consumption of fuel, distance travelled, and percentage of vehicle usage within and outside of the city, and impact of maintenance on the fuel efficiency of the vehicle.

6.2. Maritime transport sector

The enterprises of the maritime transport sector, which accounts to 10 enterprises, are fully enumerated based on administrative registers. The full enumeration of this transport sub-sector is related to the limited number of enterprises, which facilitates the analysis and publishing of more precise results on the total consumption of fuel in this sub-sector, and annuls the error due to sampling. DOS has collected information from the registries of these companies, pertaining to the type and amounts of energy consumed by the enterprises.

6.3. Railway transport sector

The railway transport sector in Jordan accounts to two enterprises, which are fully enumerated based in administrative registers. The full enumeration of this transport sub-sector is related to the limited number of enterprises, which facilitates the analysis and publishing of more precise results on the total consumption of fuel in this sub-sector, and annuls the error due to sampling. DOS has collected information from the registries of these companies, pertaining to the type and amounts of energy consumed by the enterprises.

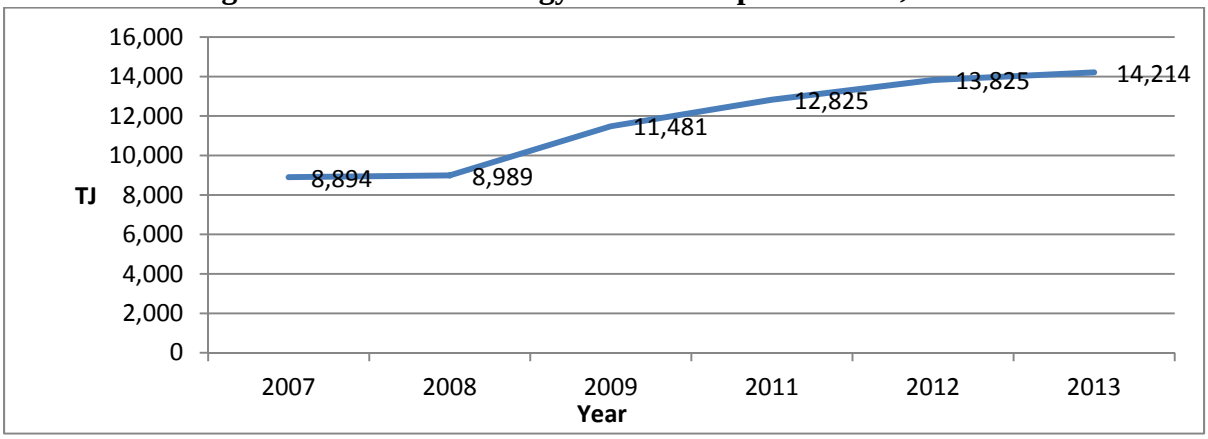
6.4. Air transport sector

The entire air transport sector amounts to 10 enterprises in Jordan. All the enterprises are enumerated in this survey based on administrative registers. The full enumeration of this transport sub-sector is related to the limited number of enterprises, which facilitates the analysis and publishing of more precise results on the total consumption of fuel in this sub-sector, and annuls the error due to sampling. DOS has collected information from the registries of these companies, pertaining to the type and amounts of energy consumed by the enterprises.

7. Transport Energy Consumption survey- Palestine

As a result of the conducted survey, PCBS reported the preliminary national results of the end use energy consumed in the transport sector, which is reported at 493 ktoe in 2014, compared to 455 ktoe in 2013. The energy use in the transport sector represented about 36 per cent of total energy end use in 2013. The trend of increase is shown in figure 5.3. In 2012, the energy used in the transport increased 25 per cent, compared to 4 per cent of annual increase between 2006 and 2012.

Figure 5.3. Palestine energy use in transport of total, 2007-2013



The Palestinian Central Bureau of Statistics (PCBS) is the official institution to conduct the national survey on energy consumption for the road transport sector in Palestine. Road transportation sector includes all engine-based vehicles and trailers designed for transportation of goods and passengers. All questionnaires abide by the confidentiality of sharing information under the national statistics law and respondents were informed about this before starting to answer any question.

Target population: all registered vehicles in Palestine in the year 2014.

Sampling frame: frequency distribution of all registered vehicles in Palestine in the year 2014, distributed according to governorate and types of vehicles, obtained from the Ministry of Transportation.

Sample size: the sample size is 6,974 vehicles.

Sampling design: stratified quota sampling divided through two stages. The first stage enumerates all motor vehicle inspection centers in Palestine, while the second stage consists of sample selection of vehicles distributed across governorates, vehicles' types, models, type of engine, and engine capacity. Several strata were conceived during the design phase of the survey to respond to specific criteria for vehicles and to be able to sample all types of operating vehicles

in Palestine. The selected strata were Governorates, types of vehicles, types of fuel consumed, vehicles manufacturing year, and engine size.

The survey was implemented in motor vehicles inspection centers across Palestine. However, in the absence of vehicles to cover all categories, data was collected from the parking vehicles and gas stations. The survey included the West Bank and Gaza strip.

The questionnaire design was divided through three main categories (PCBS, 2015). The first category collects contact information of the driver, including the registration plate number of the vehicle. The second category tackled information related to the type of the vehicle, manufacturing year, maximum load capacity, engine size and maintenance activities. The third category included questions related to the type of fuel consumed by the vehicle, energy efficiency indicators, monthly cost of fuel and cost of the annual maintenance activities and insurance policies.

References

1. Australian Bureau of Statistics. Statistical Language. (2013). *www.abs.gov.au/ausstats/abs@.nsf/web+pages/Citing+ABS+Sources*. June 5, 2013.
2. Banning R., Camstra A., and Knottrenus, P. (2012). Sampling theory – sampling design and estimation method. *Statistics Netherlands*.
3. Battaglia, Michael, P. (2008). Nonprobability sampling. *Encyclopedia of Research Methods*. SAGE Publications. 8 Nov.2011.
4. Bethlehem, J. (2009). Applied Survey Methods: A Statistical Perspective. *John Wiley & Sons*, New York.
5. Canadian Vehicle Survey: Annual. (2009). http://www.statcan.gc.ca/pub/53-223-x/53-223-x2009000-eng.htm.
6. Central Agency for Public Mobilization and Statistics Egypt (CAPMAS). (2015). Questionnaires on energy consumption in the transport sector. http://www.escwa.un.org/esab/surveyqAr.asp
7. Cochran, W.G. (1977). Sampling Techniques. *John Wiley & Sons*, New York.
8. Department of Statistics Jordan (DOS). (2015). Questionnaire on energy consumption in the transport sector. http://www.escwa.un.org/esab/surveyqAr.asp
9. Deville, J.C. (1991). A theory of quota surveys. *Survey Methodology*, 17, 163–181.
10. Doherty, M. (1994) Probability versus Non-Probability Sampling in Sample Surveys. *The New Zealand Statistics Review*, March 1994 issue, pp 21-28.
11. Espejo M.R. and Pineda M.D. (1997). On Variance Estimation for Poststratification: A Review. *Metron*, 209-220.
12. Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics* 35, 1491-1523.
13. Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics* 14, 333-362.
14. Hoffmeyer-Zlotnik, J.H.P. and Krebs, D. (1996). Different Methods of Survey Sampling in Germany. *Developments in Data Analysis. A. Ferligoj and A. Kramberger* (Editors) Metodološki zvezki, 12, Ljubljana: FDV.
15. Holt, D. and Smith, T.M.F. (1979). Poststratification. *Journal of the Royal Statistical Society*, A142, 33-46.
16. Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association*, 47, 663-685.
17. King, B. (1985). Surveys combining probability and quota methods of sampling. *Journal of the American Statistical Association*, 80 (392), 890-896.
18. Kish, L. (1965). *Survey sampling*. New York: John Wiley
19. Kokor, H. 2014. Policy reforms to promote energy efficiency in the transportation sector. Paper in the United Nations Development Account. Promoting energy efficiency investments for climate change mitigation and sustainable development. Case Study Egypt.
20. Lessler, J.T., and Kalsbeek, W.D. (1992). Nonsampling error in surveys. Wiley.
21. Lohr, S.L. (1999). *Sampling: design and analysis*, First Edition. London: Brooks/Cole.
22. Lohr, S.L. (2009). *Sampling: design and analysis*, Second Edition. London: Brooks/Cole.

23. Missaoui, W, and Bou Rahla, R. (2013). Training manual on methodologies for data collection on energy use by the transport sector and case studies from the Arab region. UNITED NATIONS, ESCWA.
24. Moser C.A, and Stuart A. (1953). An Experiment Study of Quota Sampling. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 116, No. 4 (1953), pp. 349-405
25. Moser, C.A. (1952). Quota Sampling. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 115, No. 3 (1952), pp.411-423.
26. Narain, R. D. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics* 3, 169–174.
27. Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-606.
28. Noelle-Neumann E. (1963). *Umfragen in der Massengesellschaft* . Reinbek. Rowohlt .
29. Palestinian Central Bureau of Statistics Palestine (PCBS). (2015). Questionnaire on energy consumption in the transport sector survey. http://www.escwa.un.org/esab/surveyqAr.asp
30. Särndal, C.E, Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.
31. Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics* 5, 119-127.
32. Simard M, and Franklin S. (2005). *Sample Design Guidelines*. Generations and Gender Programme. Survey Instruments. United Nations.
33. Stephenson C.B. (1979). Probability Sampling with Quotas: An Experiment. *The Public Opinion Quarterly*, Vol. 43, No. 4 (Winter, 1979), pp. 477-496.
34. UN DATA, Glossary. (2008). https://data.un.org/Glossary.aspx
35. United Nations (2011). International Recommendations for Energy Statistics. http://unstats.un.org/UNSD/energy/ires/default.htm
36. Valliant, R., Dever, J.A., and Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York.
37. Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). *Finite population sampling and inference*. New York: John Wiley.
38. Wallgren, A. and Wallgren, B. (2007). *Register-based Statistics: Administrative Data for Statistical Purposes.* John Wiley & Sons, Inc., New York.
39. Wolter K.M. (2007). *Introduction to Variance Estimation*, 2nd. Springer, New York.
40. Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society* B, 15, 253-261.