



Survey Methodology

For Energy Consumption Survey in Transport Sector

**Abdulhakeem Eideh,
Al-Quds University, Palestine
msabdul@science.alquds.edu**

**Prepared for the UN Economic and Social Commission for Western Asia for the
Project with IDB and DIFD**

**Strengthening Statistical Capacity of Arab Countries in Producing Energy Statistics on
Consumption in Transport Sector**

Outline

I. Outline

II.Objectives of the Survey

III.Collection of statistical data

III.1.Sampling frame

III.2.Sampling Methods

III.3.Post-stratification

III.4.Quota sampling

IV.Sampling design for ECSTS

IV.1.Design for road transportation

V.Estimation under quota sample design

VI.Sampling design – Countries

VI.1.Egypt

VI.2.Jordan

VI.3.Palestine

VII.References

II. Objectives of the Survey

- **Provide statistical information in Egypt, Jordan and Palestine about**
 - **Energy consumption**
 - **Expenditures**
 - **Energy efficiency**
- **In all modes of transportations: road, railway, maritime, and air**
- **Link the resulted statistics to national policies on energy production and consumption based on collected statistics.**

III. Collection of Statistical Data

- National statistical offices, three kinds of statistics are published:
 - **Statistics based on sample surveys**
 - **Statistics based on censuses**
 - **Statistics based on administrative registers**
- (ESCWA, 2013): 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.)
- So, 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

III.1 Sampling Frames

III.2 Sampling Methods

- See
- <http://css.escwa.org.lb/SD/Docs/ECTS-PartI.pdf>

III.3. Post-stratification

- Post-stratification is the stratification made **after the selection of the sample with some variables that are suitable for stratification**, such as: Vehicle type, engine capacity, etc
- The stratum to which a unit belongs is not known until the data have been collected
- The stratum sizes may be obtainable fairly **from official statistics**, but the **units can classify into strata only after the sample data is known** (Kish, 1965)
- This technique is found to be most practical for surveys where vehicle consumption of gasoline may be expected to vary with age, type, engine capacity etc.

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
- 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

Why Poststratify?

- Poststratification is used to increase the efficiency of estimation
- Since “calibrating” to the known post-stratum population totals (obtainable fairly from official statistics) can produce better estimates than a simple random sample of the same population
- Thus poststratification is a weighting technique used for estimation, giving “adjusted” or “corrected” estimates
- It makes up for the fact that some part of the population is underrepresented in the sample

Why Poststratify? - Cont

- Underrepresentation can occur because of an unlucky sample, which is often the case when a stratifying variable obviously exists and simple random sampling is used
- **If a sample is badly balanced for some characteristic the post-stratified estimator automatically corrects for this**
- Poststratification is also used for adjusting for underestimation due to nonresponse in a sample
- To correct imbalances in several characteristics using poststratification, sample values need to be poststratified using more than one variable

Poststratification Weight

- When conducting a survey, having a representative sample of the population is of paramount importance
- But sometimes, despite your best efforts, you 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 your sample is distributed may differ from the way it is distributed in the population**

Poststratification Weight – Cont

- For example, your sample may consist of 60 percent vehicles with engine capacity greater than or equal 1600 cubic centimeter (CC), and vehicles with engine capacity less than 1600 CC make up only 40 percent of the population
- This introduces bias into any estimate you may obtain from your sample data because statistical procedures will give greater weight to those vehicles you oversampled
- You can correct for these biases mathematically with a poststratification survey weight

Calculation Poststratification Weight

- In order to calculate a poststratification weight, you need an auxiliary variable to which you can compare your sample data
- For example, if you were conducting a survey of energy consumption of gas in transport sector for private vehicles in a village, you would need census data that show the engine capacity characteristics of the population of all private vehicles in a village
- You could then compare your sample 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

Calculation Poststratification Weight - Cont

- It is called a poststratification weight because you can only compute it after you have collected all your data
- The stratification part comes from the fact that you use various known strata (such as engine capacity groups for private vehicles in a village) of the population to adjust your sample data to conform more to the population's parameters

Calculation Poststratification Weight - Cont

- **Example:**
- **We use the vehicle type = private as quota h**
- **The variable engine capacity as auxiliary variable for poststratification, since it is correlated with energy consumption of the vehicles**
- **You may use a spreadsheet or a calculator to compute the poststratification weights**

Calculation Poststratification Weight - Cont

Poststratification weighs

Stratum	Population Size	Population Proportion	Sample Size	Sampling Sample	Poststratification Weight
Engine Capacity Groups	Population Size, N_h	Population Proportion, P_h	Sample Size, n_h	Sampling Sample, p_h	$pw_h = P_h / p_h$
Less than or equal 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 2500	$N_2 = 300$	$P_2 = N_2 / N = 0.3$	$n_2 = 100$	$p_2 = n_2 / n = 0.5$	$pw_2 = P_2 / p_2 = 0.6$
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	

Calculation Poststratification Weight - Cont

- As we can see from the table, since $P_1 = N_1/N = 0.5$ is less than ,
- $p_1 = n_1/n = 0.25$ 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, we assume that $w_{hi} = 1$, (Battaglia, Michael, P. (2008)) for all h and i . In this example engine capacity group for private vehicles with less than or equal 1600 CC is up weighed since the weight for this category of private vehicles is $pw_1 = P_1/p_1 = 2$

Poststratified Estimation

- The poststratified estimate of the population mean of stratum h and overall population mean are:

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

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

Poststratified Estimation – Cont

- The poststratified estimate of the population total of stratum h and overall population total are:

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

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

III.4 Quota Sampling

- See
- <http://css.escwa.org.lb/SD/Docs/ECTS-PartI.pdf>

IV. Sampling design for ECSTS

- **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)**
- **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)**

IV. Sampling design for ECSTS - Cont

- In ECSTS project, known sampling 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, age of vehicles, etc

IV.1. Sampling 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**

IV.1. Design for Road Transportation – Cont

- **Sampling frame:** Frequency distributions for 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 governorates (geographic region)
- **Sampling design:** Stratified quota sampling
- **Stratified variables:** governorates and vehicles type

IV.1. Design for Road Transportation – Cont

- For each governorate, divide the target population into several sub-populations based on certain characteristics of vehicles that are correlated with the survey control variables such as: type, age, engine capacity, type of fuel, etc.
- In this case, the survey variable is energy consumption
- Since quota sampling is **representative of vehicle type** in each governorate, then the **population structure is close to the sample structure with respect to this variable**

IV.1. Design for Road Transportation – Cont

- **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**

IV.1. Design for Road Transportation – Cont

- **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, vehicles stations, 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**

IV.1. Design for Road Transportation – Cont

- This methodology (quota sampling) is often justified as a way to avoid:
 - (a) 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
 - (b) It also avoids non-response since the interviewer continues questioning the vehicles' owners until she/he reaches the total number of the sample size

V. Estimation under quota sample design

- See
- <http://css.escwa.org.lb/SD/Docs/ECTS-PartI.pdf>

VI. Sampling design – Countries

- **Road transport sector:**
- Includes all engine-based vehicles and trailers designed for transportation of goods and passengers

VI. Sampling design – Countries - Cont

VI.1. Egypt

- **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:** quota sampling
- The allocations are performed independently for each **control variable:** province, governorate, and vehicle type, using proportional allocation method

VI. Sampling design – Countries - Cont

VI.2. Jordan

- **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, etc.
- The sampling frame **does not contain personal information** that enables direct contact with the owner or driver of the vehicles
- **Sampling design:** stratified quota sampling method
- The sample is allocated for each governorate for each vehicle type **using proportional allocation approach**

VI. Sampling design – Countries - Cont

VI.3. Palestine

- **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.
- **Sampling design:** stratified quota sampling with two stages:
 - **Stage 1:** enumerate all motor vehicle inspection centers in Palestine
 - **Stage 2:** sample selection of vehicles distributed across governorates, vehicles' types, models, type of engine, and engine capacity
- The sample was distributed to the quota of the sampling design, so that it has a sufficient number of vehicles

VII. References

- Australian Bureau of Statistics. Statistical Language. (2013). www.abs.gov.au/ausstats/abs@.nsf/web+pages/Citing+ABS+Sources. June 5, 2013.
- Bethlehem, J. (2009). *Applied Survey Methods: A Statistical Perspective*. John Wiley & Sons, New York.
- Cochran, W.G. (1977). *Sampling Techniques*. John Wiley & Sons, New York.
- Deville, J.C. (1991). A theory of quota surveys. *Survey Methodology*, 17, 163–181.
- Doherty, M. (1994) Probability versus Non-Probability Sampling in Sample Surveys. *The New Zealand Statistics Review*, March 1994 issue, pp 21-28.
- Holt, D. and Smith, T.M.F. (1979). Poststratification. *Journal of the Royal Statistical Society*, A142, 33-46.
- King, B. (1985). Surveys combining probability and quota methods of sampling. *Journal of the American Statistical Association*, 80 (392), 890-896.
- Lessler, J.T., and Kalsbeek, W.D. (1992). *Nonsampling error in surveys*. Wiley.
- Lohr, S.L. (2009). *Sampling: design and analysis*, Second Edition. London: Brooks/Cole.
- 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
- 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
- Moser, C.A. (1952). Quota Sampling. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 115, No. 3 (1952), pp.411-423.
- 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.
- Särndal, C.E, Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.
- Stephenson C.B. (1979). Probability Sampling with Quotas: An Experiment. *The Public Opinion Quarterly*, Vol. 43, No. 4 (Winter, 1979), pp. 477-496.
- Valliant, R., Dever, J.A., and Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York.
- [Wallgren](#), A. and [Wallgren](#), B. (2007). *Register-based Statistics: Administrative Data for Statistical Purposes*. John Wiley & Sons, Inc., New York.
- Others