HOUSEHOLD BUDGED SURVEY SAMPLING METHODOLOGY IN TÜRKİYE

Yılmaz ERŞAHİN
TURKSTAT

10 -12 April 2019 – Astana / KAZAKISTAN

Presentation Plan

- Introduction
- Survey Coverage
- Estimation Level and Sample Size
- Sampling Method
- Address Frame
- Stratification Criterion
- Sampling Distribution
- Weighting

Introduction

The purpose of this study: Through the data on the consumption expenditures obtained from the Household Budget Survey (HBS), the aim is to;

- Select and updating goods and weight of goods for the consumption price index
- Monitor changes over time in consumption patterns of households,
- Compile data that will help private final consumption expenditure estimations of national income calculations
- Obtain information for poverty line and other socio-economic analysis
- Compile data necessary for the studies on determination of minimum wage studies, etc.

Survey Coverage

Geographical Coverage:

All settlements in the territory of Republic of Turkey were covered in sample selection. Only small clusters (villages/ quaters) which have address less than 20 were not covered in the survey as they do not allow reaching the sufficient block size.

Survey Coverage (Continue)

Covered Population:

All population in private households who are living in the territory of Republic of Turkey.

Not Covered Population:

The population defined as the institutional population (residents of dormitories, hotels, schools, kinder-gardens, rest homes for elderly persons, special hospitals, military barracks and recreation quarters for officers)

Estimation Level and Sample Size

Sample size of the survey was calculated to produce yearly estimations on the base of ***total of Türkiye***.
And also by aggregating 3 years data NUTS Level 2 totals

The calculated sample size is yearly 15552 ***households***. (adress)

Aggregated sample size : 46656 household (adress)

Sampling Method: Multi stage stratified cluster sampling (2 stage)

Primary Sampling Unit (PSU): ***Cluster***

[Clustering was done, so that cluster size is approximately 200 addresses]

Clusters were selected in each design domain separately by using probability proportional to size (***PPS***)

Secondary Sampling Unit (SSU): ***Address (Household)***

Mountly 6 addresses were selected s***ystematically*** within each selected cluster. (According to sample design 36),

Address Frame

The frame used in sampling studies of survey is the **National Address Database** which is the base of 2007 Address Based Population Registry System.

The database is interactive and updated twice (January and August) in the year for sampling issues.

Stratification Criteria

"*Urban*" : Settlements with population 20.001 and over

"*Rural*" : Settlements with population 20.000 and below

NUTS : (nomenclature of territorial units for statistics) nomenclature was established to reduce development differences among territorial regions and serves as a reference for the collection, development and harmonization of European Union (EU) regional statistics and for socio-economic analyses of the regions. Since NUTS classification is used only by EU member states, version used in Turkey, as a candidate country, is named statistical regions (SR-İBBS)

ROTATION PATTERN

Sub Sample		2019				2020			
		Quarter				Quarter			
		I	II	III	IV	I	II	III	IV
15	A								
16	A	A							
17		A	B						
18			B	A					
19				A	B				
20					B	A			
21						A	B		
22							B	A	
23								A	B
24									B
25									

Sampling Distribution

Monthly Number of Sample Cluster and Household

	Urban	Rural	Total
Number of Sample Cluster	172	44	216
Number of Sample Household	1032	264	1296

Quarerly Number of Sample Cluster and Household

	Urban	Rural	Total
Number of Sample Cluster	172	44	216
Number of Sample Household	3096	792	3888

Weighting

Following steps will use for weighting of design:

- *Design weights calculation*,
- *Nonresponse adjustment*,
- *Integrative calibration*,
- *Trimming*,
- *Overall inflation factor*.

Weighting (Continue)

Design Weights

$$w_i = \frac{1}{p_i}\frac{1}{g_{ij}}$$

w_i : design weight for i^{th} cluster

p_i : selection probability of i^{th} cluster

g_{ij} : selection probability of j^{th} address in i^{th} cluster

Weighting (Continue)

Selection probabilities of ith cluster (PSU) are calculated below:

$$p_i = \frac{b * S_i}{\sum_{i=1}^{N} S_i}$$

p_i : selection probability of ith cluster

b : number of the selected clusters

S_i : total number of address in ith cluster

N : number of cluster in design domain

Weighting (Continue)

Selection probabilities of jth address in ith cluster (SSU) are calculated below:

$$g_{ij} = \frac{c}{S_i}$$

c : number of the selected address from ith cluster (6)
g_{ij} : selection probability of jth address in ith cluster
S_i : total number of address in ith cluster

Weighting (Continue)

Nonresponse Adjustment Factors

Household nonresponse adjustment is going to be made at household level in each cluster.

This factor is inversely proportional to response rate.

Response rate at household level for i^{th} cluster R_i is calculated below:

$$R_i = \frac{\text{number of interviewe d households in cluster i}}{\text{number of eligible households in cluster i}}$$

Weighting (Continue)

Nonresponse adjustment factor at household level for i^{th} cluster r_i is calculated below:

$$r_i = \frac{1}{R_i}$$

Weighting (Continue)
Non response results

	2017
Total	15552
Respond	12166
Non response	1745
Out of scope	1641
non response rate	0,125

=non_resp/(eligible)=non_resp/(non_resp+respons)

Weighting (Continue)

Integrative Calibration

Integrative calibration procedure, based on iterative proportional fitting, will be applied so that the weight given to each member within the household will be equal to the household weight.

Consistent results with projected population distribution will be achieved in terms of age/gender group , and NUTS2/rural-urban characteristics.

Weighting (Continue)

Trimming

Just after the calculation of non-response and calibration weights, any computed weight outside the following limits is recoded to the boundary of these limits:

$$\frac{1}{L} \leq \frac{w_h}{\overline{w}} \leq L$$

$\overline{w}$: mean value of household calibrated weight (at the same time the calibrated weight of individuals within the household)

L : upper limit to trim the weights

The value of L is going to be taken as 3 as being a reasonable value for determining the boundaries.

Weighting (Continue)

Overall Inflation Factor

At the last stage of weighting, the overall inflation factor is applied by simply dividing the mid-dated projected population of the fieldwork to the sampled population.

$$F = \frac{P}{\sum w_k}$$

F : overall inflation factor

P : total projection population (except institutional population)

w_k : estimation related to the k^{th} individual after nonresponse adjustment, calibration and trimming procedures

Final weights are calculated by multiplying all the weights of individuals by overall inflation factor.

Thank you…

Organisation of Islamic Cooperation
STATISTICAL, ECONOMIC AND SOCIAL RESEARCH AND TRAINING CENTRE FOR ISLAMIC COUNTRIES
"Cooperation for Development"

yilmazersahin@tuik.gov.tr