Tunisia: Rural Youth Survey

Basic Information Document (BID)

Background

Survey Instruments

Sample Design

The Household Survey has a sample size of 1,400 households the entire rural area of Tunisia, as defined by the Tunisian Statistical Office, *Institut National de la Statistique* (INS). For the purpose of sampling, administrative governorates were grouped into 3 Survey Regions. The data is representative on the level of these Survey Regions, which largely correspond to socio-economically and geographically distinct rural zones. The first survey region covers the Coast and includes coastal governorates in the North and East of the country. The second survey region covers South and includes the southern governorates. The third survey region is covers the rural Interior of Tunisia and includes the remote areas of central and western Tunisia, incl. the Algerian border.

The sample was drawn from the latest available census, the 2004 General Census of Population and Housing, provided by the INS. This census also provided the sampling frame for the corresponding Urban and Peri-Urban Youth Survey. For determining the number of households in rural areas, proportionality of the possible locations was used to ensure representativeness. Because of the overall research focus on youth, the sampling design ensures representativeness of youth population, which is defined by ages 15-29. The proportionality to youth population size is based on the disaggregation of Tunisia into Enumeration Areas (EA)[1]. Each EA contains about 100-120 households. In total 70 EAs were randomly selected, with 29 EAs along the Coast, 10 EAs in the South, and 31 EAs in the Interior survey regions. The relative distribution between the survey regions corresponds to their respective shares of youth population. From each of these 70 EAs, 20 households were randomly selected, leading to a total sample size of 1,400 households.

The random sampling of PSUs was performed by experts from the INS, who were also responsible for the sample frame. The drawing of 20 households from each PSU is processed on a systematic and clearly defined approach. A random-walk procedure was conducted for each of the PSUs of the sample, which included 2 separate starting points at opposing ends of the east-west dimension of each PSU, and moving towards the population center of the PSU to allow a full coverage of both centrally and remotely located households.

Implementation

---

[1] EAs are also referred to as *District de Recensement* by INS

Data Set

Appendix 2: Sampling plan

The first step is to determine the sample size. We have explored different sample sizes of youth (500-4500) and for different proportions (5% to 50%) showing the frequency of a given phenomenon regarding young Moroccans. Thus we have considered rare phenomena among youth (5%) and phenomena fairly represented (50%). The figure below shows the evolution of absolute errors on the estimated proportions of different levels (5% to 50%) by the level of sample size. We note that evolution is marked by a significant decrease in going from a size of 500 to 2500 youth, and then beyond this level of sample size, the drop starts declining.

Figure 1.Evolution of the absolute error on the estimated proportions in the sample size
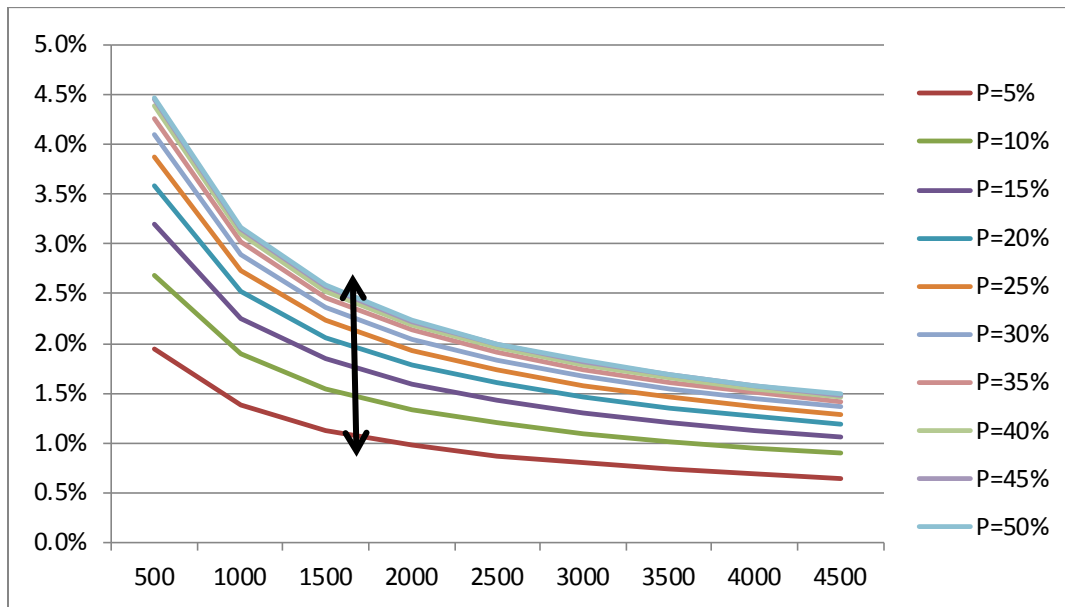
For budgetary and logistics considerations, 1,400 young people aged between 15 and 30 were selected as a sample size. For this sample size, see below the level of the absolute error. This table shows that the absolute error is only 0.8% for the estimation of a proportion of 5%, confidence interval [4.2 - 5.8%] with 95% chance that the true proportion is included in this interval. Also, the table shows that the absolute error reaches the level of 1.8% for the estimation of a proportion of about 50%, confidence interval [48.2 - 51.8%] with 95% chance that the true proportion is included in this interval. This information shows that the accuracy of indicators whatever their level is significant for a sample size of 3000 youth.

Table 1: *(Sample size = **3,000** young people aged between 15 and 30)*

Proportion (p)	Absolute error
P=5%	**0,8%**
P=10%	**1,1%**
P=15%	**1,3%**
P=20%	**1,5%**
P=25%	**1,6%**
P=30%	**1,7%**
P=35%	**1,7%**
P=40%	**1,8%**
P=45%	**1,8%**
P=50%	**1,8%**

Data from the General Census of Population and Habitat (RCPH 2004)indicated thatthere are 1.5 young people aged between 15 and 30 per Moroccan household, indicating that to reach 3,000 young people approximately 2000 households need to be interviewed.The 2000 households are selected from the sampling frame of the master sample prepared following the RGPH 2004.

It is important to note that the country was divided into several primary units (PU) each comprising about 600 households. These primary units were constituted according to a set of criteria reflecting the differences between urban and rural, boundaries of regions, districts, rural communities, cities and municipalities. The master sample represents 20% of the country, and it consists of 1848 primary units (PU 1124 urban and 724 rural PU). Two stratification criteria were considered for urban PU (city size and type of habitat) and one stratification criterion was considered for the rural PU (degree of isolation expressed by the distance from the paved road). Both distributions are given in the following tables.

Table2a.ized*Distribution of PU and urban households in the master sample according to the criteria: city size and type of habitat*

Size of the city	Type of habitat	Number of household	Number of Primary Units (PU)
Less than 100,000 inhabitants	Luxurious and modern habitat	9 732	31
	Old city (Medina)	36 610	46
	Illegal housing	25 868	31
100,000 to 500,000 inhabitants	Luxurious and modern habitat	29 635	31
	New medina	14 222	31
	Old medina	89 786	123
	Illegal housing	42 407	77
500,000 inhabitants and more	Luxurious and modern habitat	89 022	139
	New medina	28 494	62
	Old medina	266 376	446
	Illegal housing	68 634	107
TOTAL		**700 786**	**1 124**

Table 2b. *Distribution of PU and rural households in the master sample according to the criterion: level of isolation: distance from the paved road*

Distance to paved road	Number of household	Number of Primary Units (PU)
Less than 5 km	350 792	524
5 to less than 20 km	101 191	154
20 km and more	13 589	46
TOTAL	**465 572**	**724**

The PUs are also broken down into 12 sub-units of about 50 households. The sampling for this survey is based on the 125 primary units (76 urban and 49 rural PUs) from the 1848 PU of the sample master. Proportionality was considered to ensure the representation of both areas of residence (urban and rural). Then one secondary unit (SU) was selected per primary unit (PU) sample. Once the 125 SU are obtained, 16 households within each SU were selected, which made a total sample of 2000 households. Within each household, all young people aged between 15 and 29 were listed.

Table3a. *Distribution of urban SU of the sample according to the criteria: size of the city and type of habitat*

Size of the city	Type of habitat	Number of PU
Less than 100 000 inhabitants	Luxurious and modern habitat	2
	Old city (Medina)	3
	Illegal housing	2
100 000 to 500 000 inhabitants	Luxurious and modern habitat	2
	New medina	2
	Old medina	9
	Illegal housing	5
500 000 inhabitants and more	Luxurious and modern habitat	10
	New medina	4
	Old medina	30
	Illegal housing	7
TOTAL		**76**

Table 3b. *Distribution of rural SU of the sample according to degree of isolation: distance from the paved road*

Distance from the paved road	Number of SU
Less than 5 km	36
5 to less than 20 km	10
20 km and more	3
TOTAL	**49**

The probability of drawing a household (k) is given by:

$$P = P_{EM} \, x P_{hi} \, x P_{hij} \, x P_{hijk}$$

With

P_{EM} : Probability of drawing the master sample (20%)

P_{hi} : Probability of drawing PU (i) in the stratum (h) of the sample

P_{hij} : Probability of drawing SU (j) of PU (i) in the stratum (h) of the sample

P_{hijk} : Probability of drawing the household (k) of SU (j) of PU (j) from the stratum (h) of the sample

Where :

$$P_{hi} = n_h \frac{Q_{hi}}{Q_h} \; ; \; P_{hij} = \frac{1}{m_{hi}} \text{ et } P_{hijk} = \frac{16}{Q_{hij}}$$

Q_h : Number of households from stratum h (h=1, …, H) H=11 for urban H=3 for rural

Q_{hi} : Number of households in PU (i) from stratum h

n_h : Number of PU from stratum h

m_{hi} : Number of SU in PU (i) from stratum h

The extrapolation coefficient of household k is the inverse of the probability (P), $\dfrac{1}{P}$