



## 2015 Regional Enterprise Survey Sampling Methodology

### August, 2015

For the implementation of the 2015 Turkey Regional Enterprise Survey, the World Bank Group (WBG) developed a customized sampling methodology. This customized approach is based on the global methodology the WBG employs in *Enterprise Surveys* around the world. The global methodology is explained in detail on the *Enterprise Surveys* website.<sup>1</sup>

This note provides a description of the methodology adapted to this particular survey. The main differences compared to the standard methodology include:

- 1) The inclusion of micro establishments (with less than 5 employees);
- 2) A target sample size, at country level, larger than the typical Enterprise Survey;
- 3) Multiple geographic levels of coverage and representativeness, as requested by the Ministry of Development.

Details of the sampling methodology were agreed during a meeting in January 2015 and in subsequent correspondence between the World Bank and the Ministry of Development.

#### MAIN PRINCIPLES

The objectives of the sampling methodology are as follows:

- 1) To generate a sample of 6,000 establishments which is representative of the formal, non-agricultural, non-extractive private economy, in order to substantiate assertions about this part of the economy. The sampling frame for the selection of the sample is managed by the Turkish Statistical Institute (TurkStat). It includes around 2.7 million establishments, registered with the Turkish Commercial Registry, that operate in the following sectors (in NACE 2 codes): 10-33, 41-56, 58, 61, 62 and 79.
- 2) To generate sub-samples that would allow for the analysis of specific groups of establishments. The Ministry of Development specifically requested the World Bank to design a sample that would allow:
  - a. Analysis for each NUTS-2 region at the overall private sector level;
  - b. Analysis for selected industries at the incentive-region level (each incentive region consisting of a group of NUTS-2 regions);
  - c. Analysis by firm-size groups (micro, small, medium, and large) at the four incentive-region level.

---

<sup>1</sup> More details can be found by accessing the Sampling Methodology Note posted on the *Enterprise Surveys* website:

[http://www.enterprisesurveys.org/~media/GIAWB/EnterpriseSurveys/Documents/Methodology/Sampling\\_Note.pdf](http://www.enterprisesurveys.org/~media/GIAWB/EnterpriseSurveys/Documents/Methodology/Sampling_Note.pdf)

Analysis at any of the levels described above requires a minimum number of observations. The sampling strategy that achieves these objectives is stratified random sampling.<sup>2</sup> The levels of stratification are: regional location (26 NUTS-2 level regions), firm size (micro, small, medium, and large), and firm activity (8 industries).

For example, the sample would contain enough observations so that one can explore productivity differences for the same sector in two different incentive regions, i.e. textiles manufacturing in Istanbul (incentive region 1) could be compared with textile manufacturing in incentive region 2. Or, one could compare the level of overall labor productivity in Tekirdağ (NUTS-2 code TR21) with that of Balıkesir (TR22).

- 3) For each stratum, the minimum sample size was defined to guarantee at least a level of precision of 7.5%, with 90% confidence intervals, for the following variables:<sup>3</sup>
  - i. Estimates of population proportions (percentages); and
  - ii. Estimates of the mean of log of sales.

As explained below<sup>4</sup>, this requirement means that for each stratum, the sample size must be, in most cases, at least 120 establishments.

Given that both the Enterprise Survey and the Regional Investment Climate Survey include more than 100 indicators, the computation of the minimum sample size required is complicated since it depends on the variance of each indicator. However, many of the indicators computed from the survey are proportions, such as percentage of firms that engage in X activity or chose Y action. In this case the computation of the sample size is simplified by the fact that the variance of a proportion is bounded. Assuming the worst case of maximum variance (0.5) the minimum level of precision is still guaranteed.

Table 1 exhibits minimum sample sizes for different population sizes for estimates of proportions with 5% and 7.5% precision in 90% confidence intervals, assuming maximum variance.<sup>5</sup> With 5% precision, the minimum sample size tends to a sample size of 270, as population size increases with 7.5% precision the sample size tends to 120. Note that if the population size of an industry falls below 1,500, the

<sup>2</sup> See Cochran, William 1977 “Sampling Techniques”, John Wiley and Sons

<sup>3</sup> A 7.5% precision of an estimate in a 90% confidence interval means that the true value of the parameter in the population is within the 7.5% range of the parameter estimated within the sample, with a probability of 90%.

<sup>4</sup> More details can be found by accessing the Sampling Methodology note posted on the Enterprise Surveys website at:

[http://www.enterprisesurveys.org/~media/GIAWB/EnterpriseSurveys/Documents/Methodology/Sampling\\_Note.pdf](http://www.enterprisesurveys.org/~media/GIAWB/EnterpriseSurveys/Documents/Methodology/Sampling_Note.pdf)

<sup>5</sup>

$$n = \left[ \frac{1}{N} + \frac{N-1}{N} \frac{1}{PQ} \left( \frac{k}{z_{1-\alpha/2}} \right)^2 \right]^{-1} \quad \text{where } N = \text{population size, } P = \text{population proportion, } Q = 1 - P$$

k=desired level of precision,  $z_{1-\alpha/2}$  is the value of the normal standard coordinate for a desired level of confidence,  $1-\alpha$ .



required sample size for proportions may be reduced considerably (see figure 1 and 2). Although a 5% precision would be most desirable, a precision of 7.5% is in line with the budget constraints for the Enterprise Survey work around the world. Consequently, an operational 120 samples size per stratum was selected

In addition, to ensure that each of the 81 NUTS-3 cities were included and added to the representativeness of the data a minimum of 25 observations per NUTS-3 was required. An additional restriction was added to limit the micro sample to a maximum of half of the sample, at country level.

**Table 1-- Sample Sizes Required with 5% and 7.5% Precision and 90% Confidence**

Population size	Sample Size 5.0%	Sample Size 7.5%
50	42	36
100	73	55
200	115	75
300	143	86
400	162	93
500	176	97
600	187	100
700	195	103
800	202	105
900	208	106
1000	213	107
1250	223	110
1500	229	111
1750	234	113
2000	238	113
2500	244	115
3000	248	116
5000	257	117
10000	263	119
50000	269	120
100000	270	120

**FINAL SAMPLE DESIGN**

The final number of industrial sector strata and incentive-regions was the result of an optimization problem: allocating the sample within the desired constraints (including an overall sample size of 6,000) and with a distribution as close as possible to the original distribution in the population.

As a result, the final sample design features:

- **5 incentive regions.**

The incentive regions were defined based on four economic incentives regions in Turkey with the only exception being that one incentive region was split into two incentive regions due to the very high firm population concentration in Istanbul (around 24% of the Turkish firms are located in Istanbul). See Appendix 1 and 2 for details.

- **8 industrial sectors of stratification**

The following 6 sectors were chosen as sectors of stratification due to their contribution to value added and/or employment:

- 1) Food,
- 2) Textiles and Wearing apparel,
- 3) Fabricated metal, Machinery and Motor Vehicles,
- 4) Construction,
- 5) Wholesale and Retail;
- 6) Transport.

All other eligible sectors were also included in the following two residual strata:

- 7) Other manufacturing sectors;
- 8) Other services sectors.

See Appendix 3 for details.

- **26 NUTS 2 regions.**

- **4 establishment size groups:**

- 1) Micro establishments (less than 5 employees);
- 2) Small establishments (between 5 and 19 employees);
- 3) Medium establishments (between 20 and 99 employees);
- 4) Large establishments (over 100 employees).

- **81 NUTS 3 regions**

All 81 NUTS 3 regions have guaranteed coverage ranging from 25obs. in Tunceli to 960 obs. in Istanbul with an average of 74 obs.).

Note that the sample design allocates interviews in such a way that the sample is as close to proportional as possible subject to fulfilling the constraints imposed. The design is arrived at as a result of a linear-programming process that accommodates the following constraints:

1. Each NUTS2 region has a minimum of 120 target interviews. This constraint ensures the sample is representative at the NUTS2 level for each of the twenty six regions and meets precision targets outlined above.
2. Each NUTS3 province has a minimum of 25 target interviews. This assures that we generate data that is nominally inclusive of all 81 NUTS 3 provinces.
3. Each incentive-region has at least 120 target interviews in each of the eight sectors outlined above. This results in a minimum of 960 target interviews per incentive-region and assures that we achieve our minimum precision targets for each sector at each of the five incentive regions. This constraint ensures the sample is representative at the sector level at each of the five incentive regions and meets the precision targets outlined above.
4. Each incentive-region has at least 120 target interviews in each of the four size categories outlined above. This constraint ensures the sample is representative at the size category at each of the five incentive regions and meets the precision targets outlined above

5. The allocation of interviews in a given cell<sup>6</sup> cannot exceed the number of interviews that could be reasonably expected if all of the contacts available in the frame were used. For example, it would not be reasonable to target 12 interviews in a cell where only 15 contacts exist for attempted recruitment, due to the expected rates of nonresponse or problems in contacting the firms.
6. A maximum of 50% of the target interviews will be composed of micro firms. This restriction ensures that we interview enough larger firms to speak meaningfully about them.
7. A total of 6000 target interviews will be included. This is a result of a practical budget constraint.
8. The probability of selection in any cell cannot be zero. At least one target interview will be drawn for each cell in which contacts exist.
9. The solution will have integer values. As it is not possible to target a fraction of an interview, the solution must result in integer values for all cells.

It is often the case that no one single constraint explains why any individual cell or region gets a larger or smaller number of interviews than any other cell or region. Each of the constraints listed above shift the allocation further away from a perfectly proportional one. It is not possible to elaborate the specifics of the allocation of target interviews in any cell or even region as each is the result of the intersection and interaction of all the constraints listed above simultaneously across 2592 cells.<sup>7</sup> However, the algorithm used ensures that the final survey design fulfils all survey design goals and restrictions as outlined above. A few examples below demonstrate the complexity of such a task.

EXAMPLE 1: it was requested that a minimum sample size be allocated to each NUTS3 province. As the table shows, while the TR51, TR52, TR62, TR72 regions (let's call them group 1) have only 1, 2, 2 and 3 provinces respectively, the TRA2, TRB2, TRC2, TRC3 regions (let's call them group 2) have 4, 4, 2, and 4 provinces respectively. As a consequence, the average sample per province is in fact smaller for group 2 (81) than for group 1 (96).

EXAMPLE 2: The sample allocated to a specific region may also be driven by the relative importance of that region within a specific industrial sector, since another constraint is the total number of interviews by sector for each incentive-region (minimum 120). For example, region TRC2 represents 4% of the national food sector, and 4% of the national transport sector (similar relative importance of TR51).

EXAMPLE 3: Another binding constraint may be the target number of interviews by establishment size. One of the objectives is to have up to 50% of the sample composed by micro firms, and region TRC2 represents 4% of the total number of micro firms at the national level (comparable to TR52 or TR72).

---

<sup>6</sup> Here, a 'cell' refers to the intersection of each of the stratifying variables. For example, Large, Food Manufacturers, in Izmir is a considered a 'cell' while food manufacturers alone is not a cell.

<sup>7</sup> There are 2592 cells in a table that crosses 4 size categories with 8 sectors and 81 NTUS3 regions.  
 $4*8*81=2592$



It is useful to note that the resulting sample is quite different from a proportional sample. A non-proportional sample that exploits the stratification is more efficient in achieving a target precision level than a purely proportional sample is. In this case, the number of interviews required to achieve the desired 7.5% target with 90% confidence is dramatically smaller than a proportional sample would be. Table 2 demonstrates that a proportional sample design would require a minimum of 13,260 interviews in order to meet the 7.5% precision target with 90% confidence for each NUTS2 region. The actual number of interview required to fulfill the targets outlined by the constraints above may be significantly higher as this table only shows the constraint of a minimum of 120 interviews per NUTS2 region.








**Table 2-- Sample Sizes required (proportional sample vs proposed sample)**

CODE	NUTS 2 REGIONS	NUMBER OF ESTABLISHMENTS IN FRAME	INCENTIVE REGION	SAMPLE SIZE PROPOSED	SAMPLE SIZE USING PROPORTIONAL SAMPLING
TR10	(İstanbul)	618,328	1	960	2,975
TR31	(İzmir)	162,783	2	150	783
TR41	(Bursa, Eskişehir, Bilecik)	137,285	2	200	661
TR42	(Kocaeli, Sakarya, Düzce, Bolu, Yalova)	126,476	2	257	609
TR51	(Ankara)	168,755	2	158	812
TR61	(Antalya, Isparta, Burdur)	138,262	2	195	665
TR21	(Tekirdağ, Edirne, Kırklareli)	69,045	3	133	332
TR22	(Balıkesir, Çanakkale)	74,096	3	124	357
TR32	(Aydın, Denizli, Muğla)	140,341	3	216	675
TR33	(Manisa, Afyon, Kütahya, Uşak)	109,671	3	216	528
TR52	(Konya, Karaman)	83,767	3	154	403
TR62	(Adana, Mersin)	124,607	3	199	600
TR72	(Kayseri, Sivas, Yozgat)	61,102	3	153	294
TR81	(Zonguldak, Karabük, Bartın)	38,423	3	131	185
TR83	(Samsun, Tokat, Çorum, Amasya)	89,913	3	194	433
TR63	(Hatay, Kahramanmaraş, Osmaniye)	92,987	4	301	447
TR71	(Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir)	48,571	4	216	234
TR82	(Kastamonu, Çankırı, Sinop)	26,756	4	128	129
TR90	(Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane)	96,604	4	380	465
TRA1	(Erzurum, Erzincan, Bayburt)	28,436	4	123	137
TRB1	(Malatya, Elazığ, Bingöl, Tunceli)	46,918	4	208	226
TRC1	(Gaziantep, Adıyaman, Kilis)	78,199	4	244	376
TRA2	(Ağrı, Kars, Iğdır, Ardahan)	<b>24,937</b>	5	157	<b>120</b>
TRB2	(Van, Muş, Bitlis, Hakkari)	42,675	5	241	205
TRC2	(Şanlıurfa, Diyarbakır)	77,496	5	333	373
TRC3	(Mardin, Batman, Şırnak, Siirt)	49,030	5	229	236
TR	Türkiye	2,755,463		<b>6000</b>	<b>13260</b>

**Note:** TRA2 is the smallest region but 120 interviews have to be completed to achieve 7.5% precision target with 90% confidence interval. Total number of interviews required is more than doubled



APPENDIX 1. OVERALL SAMPLE DISTRIBUTION, PER INCENTIVE AND INCENTIVE REGION

INCENTIVE REGION	INCENTIVE REGION	COUNT	MAP COLOR	COUNT
<b>Region 1</b>	Region 1	960		1,920
	Region 2	960		
<b>Region 2</b>	Region 3	1,520		1,520
<b>Region 3</b>	Region 4	1,600		2,560
<b>Region 4</b>	Region 5	960		
	TOTAL	6,000		6,000

Map 1. Visual representation of incentive regions coverage







APPENDIX 2. INCENTIVE REGIONS AND INDUSTRIES SAMPLE DISTRIBUTION.

INCENTIVE REGION 1									
NUTS 2	Food	Textiles, wearing apparel	Fab metal, machinery, motor vehicles	Other manufacturing	Construction	Wholesale and retail	Transport	Other services	Total
TR10	120	120	120	120	120	120	120	120	960
<b>TOTAL</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>960</b>

INCENTIVE REGION 2									
NUTS 2	Food	Textiles, wearing apparel	Fab metal, machinery, motor vehicles	Other manufacturing	Construction	Wholesale and retail	Transport	Other services	Total
TR31	12	15	14	21	19	21	20	28	150
TR41	26	30	29	25	20	25	21	24	200
TR42	44	39	41	30	27	29	21	26	257
TR51	12	13	14	22	30	20	28	19	158
TR61	26	23	22	22	24	25	30	23	195
<b>TOTAL</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>960</b>



INCENTIVE REGION 3									
NUTS 2	Food	Textiles, wearing apparel	Fab metal, machinery, motor vehicles	Other manufacturing	Construction	Wholesale and retail	Transport	Other services	Total
TR21	19	19	18	17	20	12	13	15	133
TR22	15	11	14	15	16	20	14	19	124
TR32	21	26	22	24	29	34	26	34	216
TR33	28	31	26	26	26	33	22	24	216
TR52	15	13	20	19	15	32	25	15	154
TR62	17	18	17	21	19	54	28	25	199
TR72	22	17	21	21	18	24	14	16	153
TR81	16	18	16	13	14	24	16	14	131
TR83	27	27	26	24	23	27	22	18	194
<b>TOTAL</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>260</b>	<b>180</b>	<b>180</b>	<b>1520</b>



INCENTIVE REGION 4									
NUTS 2	Food	Textiles, wearing apparel	Fab metal, machinery, motor vehicles	Other manufacturing	Construction	Wholesale and retail	Transport	Other services	Total
TR63	27	30	32	31	26	72	48	35	301
TR71	32	27	36	25	26	33	16	21	216
TR82	20	20	18	16	16	18	9	11	128
TR90	39	37	37	39	42	78	64	44	380
TRA1	14	9	11	16	17	27	14	15	123
TRB1	25	22	22	22	28	46	20	23	208
TRC1	23	35	24	31	25	45	30	31	244
<b>TOTAL</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>180</b>	<b>319</b>	<b>201</b>	<b>180</b>	<b>1600</b>

INCENTIVE REGION 5									
NUTS 2	Food	Textiles, wearing apparel	Fab metal, machinery, motor vehicles	Other manufacturing	Construction	Wholesale and retail	Transport	Other services	Total
TRA2	28	18	22	23	18	13	13	22	157
TRB2	32	26	31	30	29	24	40	29	241
TRC2	29	38	34	32	44	62	51	43	333
TRC3	31	38	33	35	29	21	16	26	229
<b>TOTAL</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>120</b>	<b>960</b>