

Sample Design Summary: ESS Round 9

1.1 Target Population

Number of residents aged 15 or older in the country: 52,406,861

Source and reference date: Istat (<http://dati.istat.it/>) , 1 January 2018

1.2 Population Coverage

The frame for selecting persons is the Italian Public Register of Individuals (*Liste Anagrafiche comunali – LAC*), which includes all people resident in Italy. The LAC is acquired by ISTAT, the National Institute of Statistics, from the municipality, with a continuous update. The information contained in the list and related to each individual are municipality of residence (and the address), date of birth, and sex. Using the date of birth, it is possible to build the sampling frame for all people aged 15 year or more (a person is treated as 15 or older if she or he is 15 at the 1st of September 2018). The continuous updating of LAC made by ISTAT, through a centralized transmission system developed by ISTAT, guarantees a high quality of data and a low probability of not finding the sampled individuals.

2. Summary of the Sample Design

A two domain sampling design is used.

The first sampling domain consists of the biggest municipalities within Italy (around 14% of the Italian population). Here a one-stage sampling design is used, where persons are sampled by stratified simple random sample from the Public Register of Individuals (LAC) from each municipality in domain 1. The allocation of the sample size is proportional to target population within the strata.

The second sampling domain consists of all municipalities that are not included in the first domain. Here a two-stage sampling design is used. At the first stage municipalities are selected as Primary Sampling Units (PSUs) by stratified sampling. The stratification is done by crossing two variables, Geographical area (NUTS-1) and demographic size class of the resident population aged 15 and over (4 classes). The allocation of PSUs to strata is proportional to target population within the strata. Within strata, PSUs are selected with probability proportional to size of target population. At the second stage persons are selected from the sampled municipalities using a simple random sample.

3. Sample Design Details

3.1 Domain 1

First Sampling Stage

- unit:** Persons
- frame:** Public Register of Individuals
- size:** 770
- strata:** Municipalities: Roma, Milano, Napoli, Torino, Palermo, Genova, Bologna, Firenze, Bari
- allocation:** Proportional to the target population size within the strata
- algorithm:** Simple random sample within strata

3.2 Domain 2

First Sampling Stage

- unit:** Municipalities
- frame:** List of all 8,037 municipalities within the domain
- size:** 163
- strata:** Strata are defined by crossing two variables; Geographical area (NUTS-1 level, namely 5

geographic areas: North-West, North-East, Center, South, Islands) and size of the resident population aged 15 and over (4 classes: less than 2,000; 2,000 - 10,000; 10,001 - 50,000 persons; 50,001 and more), according to the standard classification used by Italian National Institute of Statistics (ISTAT) (see Table: Target Population Size by Strata in Domain 2)

allocation: Proportional to the target population size within the strata (see Table: Allocation of Municipality Sample in Domain 2)

algorithm: Systematic sampling with probability proportional to the target population size of the municipalities

Second Sampling Stage

unit: Persons

frame: Public Register of Individuals

size: 29 persons per PSU (4727 in total)

strata: NA

allocation: NA

algorithm: Simple random sample

Comment

If possible, the design in domain 2 will change to 175 PSUs with 27 persons in each, rather than 163 PSUs with 29 persons in each. This will slightly reduce deffc. This change depends on agreement with the survey agency.

4. Planning the Sample Size

History of Planned and Realised Values

ESS	d	$p.\bar{b}$	\bar{b}	$p.\rho$	ρ	$p.Deff$	Deff	$p.Deff_c$	Deff _c	$p.Deff_p$	Deff _p
1		4.0	10.98	0.03	0.108	1.20	1.93	1.09	1.67	1.10	1.16
2		14.6	14.7	0.02	??	1.540	1.529	1.273	1.274	1.21	1.20*
6	1	NA	NA	NA	NA	1.00	1.000	1.00	1.00	1.00	1.000
6	2	14	6.45	0.02	0.092	1.26	1.501	1.26	1.501	1.00	1.000
6	t	NA	NA	NA	NA	1.225	1.434	1.225	1.434	1.00	1.000
8	1	NA	NA	NA	NA	1.00	1.000	1.00	1.00	1.00	1.000
8	2	13.0	14.3	0.06	-	1.72	1.802	1.72	1.798*	1.00	1.002
8	t	NA	NA	NA	NA	1.619	1.687	1.619	1.685*	1.00	1.001

Planned and Benchmark Values of the Sample Sizes

ESS	d	$p.rr$	rr	$p.ri$	ri	$p.n_{gross}$	n_{gross}	$p.n_{net}$	n_{net}	$p.n_{eff}$	n_{eff}
1		0.68	0.437	0.02	0.08	3,000	3,000	2,000	1,207	1,667	626
2		0.62	0.602	0.052	0.028	2,588	2,613	1,521	1,529	988	1,000
6	1	-	-	NA	NA	329	327	-	128	NA	NA
6	2	-	-	NA	NA	2,451	2,451	-	832	NA	NA
6	t	0.60	0.36	0.101	0.043	2,780	2,778	1,890	960	1,543	669
8	1	0.45	0.398	0.02	0.029	769	770	340	298	NA	NA
8	2	0.45	0.506	0.02	0.027	4,727	4,727	2,084	2,328	NA	NA
8	t	0.45	0.491	0.02	0.027	5,496	5,497	2,424	2,626	1,497	1,557

Parameters of the Planned Gross Sample Size

Domain	Achieved interviews per cluster (\bar{b})	Intraclass Correlation Coefficient (ρ)	Design Effect due to Selection Probabilities ($Deff_p$)	Response Rate (rr)	Ineligible Rate (ri)	Effective Sample Size (n_{eff})	Domain weight (γ)
1	NA	NA	1.00	0.400	0.029		0.140
2	14.2	0.06	1.00	0.505	0.027		0.860
t	NA	NA	1.00	0.490	0.027	1500	NA

Design Effect: Domain 1

$$Deff_c = 1.000^*$$

$$Deff_p = 1.000^*$$

$$Deff = Deff_p \times Deff_c = 1.000^*$$

Design Effect: Domain 2

$$\begin{aligned} Deff_c &= 1 + (\bar{b} - 1) \times \rho \\ &= 1 + (14.2 - 1) \times 0.06 \\ &= 1.792^* \end{aligned}$$

$$Deff_p = 1.000^*$$

$$Deff = Deff_p \times Deff_c = 1.792^*$$

Design Effect: Overall

$$\begin{aligned} Deff &= (0.14 \times 1.000) + (0.86 \times 1.792) \\ &= 1.681^* \end{aligned}$$

**results have been rounded to 3 d.p.s*

Gross Sample Size

$$\begin{aligned}
 \text{Min. } n_{net} &= \text{Deff} \cdot n_{eff} \\
 &= 1.681 \times 1500 \\
 &= 2522^{**} \\
 \hline
 \text{Target } n_{net} &= 2620^{**} \\
 \hline
 n_{gross} &= \frac{n_{net}}{rr \times (1 - ri)} \\
 &= \frac{< 2620}{0.49 \times (1 - 0.027)} \\
 &= 5495^{**} \\
 \hline
 \end{aligned}$$

*** results have been rounded to 0 d.p.s.*

5. Sampling Design Data File (SDDF)

Variables to be included in the SDDF

Variable	Description
IDNO	Respondent identification number
CNTRY	IT
PROB1	Domain 1: Inclusion probability of persons within municipalities. Domain 2: Inclusion probability of municipalities
PROB2	Domain 1: Equal to one for all persons. Domain 2: Conditional inclusion probability of persons within selected municipalities
PSU	Domain 1: Respondent identification number. Domain 2: Municipality identification number
STRATEX	Domain 1: Municipality identification number. Domain 2: An identification variable of the cross-classification of NUTS2 area and demographic size class
DOMAIN	Sampling domain identifier
OUTCOME	Summary field outcome

BIRTH	Date of birth
GENDER	Gender
CITIZENSHIP	Citizenship
MALE	Number of men aged 15 or over in the municipality
FEMALE	Number of women aged 15 or over in the municipality
AGE_15_24	Number of people aged 15-24 in the municipality
AGE_25_34	Number of people aged 25-34 in the municipality
AGE_35_44	Number of people aged 35-44 in the municipality
AGE_45_54	Number of people aged 45-54 in the municipality
AGE_55_64	Number of people aged 55-64 in the municipality
AGE_65	Number of people aged 65 or over in the municipality
UNMARRIED	Number of unmarried people aged 15 or over in the municipality
MARRIED	Number of married people aged 15 or over in the municipality
DIVORCED	Number of divorced people aged 15 or over in the municipality
WIDOWED	Number of widowed people aged 15 or over in the municipality
DENSITY	Population density of the municipality
EMPLOYEE	Number of employees aged 15 or over in the stratum
SELF_EMPLOYED	Number of self-employed people aged 15 or over in the stratum
UNEMPLOYED	Number of unemployed people aged 15 or over in the stratum
INACTIVE	Number of economically inactive people aged 15 or over in the stratum
EDU_LEV_0	Number of people aged 15 or over not holding a primary level diploma
EDU_PRIMARY_1	Number of people aged 15 or over holding only a primary level diploma
EDU_LOW_SEC_2	Number of people aged 15 or over holding a lower secondary level diploma
EDU_UPP_SEC_3C	Number of people aged 15 or over holding a 2-3 years upper secondary level diploma
EDU_UPP_SEC_3AB	Number of people aged 15 or over holding a 4-5 years upper secondary level diploma
EDU_TER_5	Number of people aged 15 or over holding a tertiary level diploma
FOREIGNERS	Number of people aged 15 or over in the municipality with foreign nationality
ITALIANS	Number of people aged 15 or over in the municipality with Italian nationality

Probabilities of Selection

1. Sampling Domain

1. $PROB1_{i|h1} = \frac{n_{h1}}{N_{h1}}$

2. $PROB2_{i|h1} = 1$

2. Sampling Domain

1. $PROB1_{i|h2} = m_{h2} \frac{N_{ih2}}{N_{h2}}$

2. $PROB2_{i|h2} = \frac{q}{N_{ih2}}$

- $PROB1_{i|h1}$ = Value of *PROB1* for all persons in the *h*-stratum in domain 1
- $PROB2_{i|h1}$ = Value of *PROB2* for all persons in the *h*-stratum in domain 1
- $PROB1_{i|h2}$ = Value of *PROB1* for all persons in the *h*-stratum in domain 2
- $PROB2_{i|h2}$ = Value of *PROB2* for all persons in the *h*-stratum in domain 2
- n_{h1} is the number of persons selected in the *h*-th stratum in domain 1
- N_{h1} is the size of the target population in the *h*-th stratum in domain 1
- m_{h2} number of municipalities selected in the *h*-th stratum in domain 2
- N_{h2} the size of the target population in the *h*-th stratum in domain 2
- N_{ih2} the size of the target population in the *i*-th PSU in the *h*-th stratum in domain 2
- q is the number of persons selected in each PSU in domain 2 (expected to be either 29 or 27)

Appendix

Target Population Size by Strata in Domain 2

Nuts 1 Region	Pop. Size Class				All
	1-2000	2001-10000	10001-50000	50001 -	
<i>North-west</i>	1454330	4451321	4221823	1345144	11472618
<i>North-east</i>	564292	3375583	3189693	2580752	9710320
<i>Center</i>	400667	1862686	3631316	1756227	7650896
<i>South</i>	864772	2989865	5064033	2078636	10997306
<i>Islands</i>	330708	1453638	2086731	1363736	5234813
<i>All</i>	3614769	14133093	18193596	9124495	45065953

Allocation of Municipality Sample in Domain 2 (assuming $m_2=163$)

Nuts 1 Region	Pop. Size Class			
	1-2000	2001-10000	10001-50000	50001 -
<i>North-west</i>	5	16	15	5
<i>North-east</i>	2	12	12	9
<i>Center</i>	2	7	13	6
<i>South</i>	3	11	18	8
<i>Islands</i>	1	5	8	5

