Special topics and recent developments in sampling theory and practice

Accuracy in Official Statistics

Alessio Guandalini <alessio.guandalini@istat.it>

ISTAT

Quality measures of data in Statistical Agency

Eurostat's definition of quality:

- Relevance
- Accuracy and reliability
- Timeliness and punctuality
- Coherence and comparability
- Accessibility and clarity

Quality measures of data in Statistical Agency

Relevance

European Statistics must meet the needs of users.

Accuracy and reliability

European Statistics must accurately and reliably portray reality.

Timeliness and punctuality

European Statistics must be disseminated in a timely and punctual manner.

Coherence and comparability

European Statistics should be consistent internally, over time and comparable between regions and countries; it should be possible to combine and make joint use of related data from different sources.

Accessibility and clarity

European Statistics should be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

Planning a survey

Objective of a survey

Investigate some characteristics of a population
(unemployed rate, household average income, turnover of companies, etc.)

- **Target population** $U = \{1, \ldots, k, \ldots, N\}$
  (individuals inhabitant, families inhabitant with at least a foreign member, active companies, etc.)
- **Variables of interest** $y_1, \ldots, y_t$
  (occupational status, income, number of employees, annual turnover, etc.)
- **Parameters** $f(y_1), \ldots, f(y_t)$
  (unemployment rate, mean household income, total of employees, etc.)
- **Domain of estimates**
  (National level, Geographic areas, Regions, Provincies)
- **Requirement of precision**
  (coefficient of variation, confidence intervals)

Planning a survey

Data collection methods

- PAPI (paper and pencil personal interview)
- CATI (computer assisted telephone interview)
- CAPI (computer assisted personal interview)
- CASI (computer assisted self interview)
- CAWI (computer assisted web interview)
- Mixed-mode

Planning a survey

The survey frame

Tool used to obtain access to the population

- **list frame**
  a list of names and addresses that provides direct access to 'individuals' (e.g., a list of hospitals, a list of restaurants, a list of students at a university)
- **area frames**
  lists of geographic areas that provide indirect access to individuals (e.g., the neighbourhoods in a city)

Planning a survey

Types of survey

SAMPLE SURVEY

+ reduces cost
+ reduces time
+ promptness
+ more detailed questions
+ small respondent burden

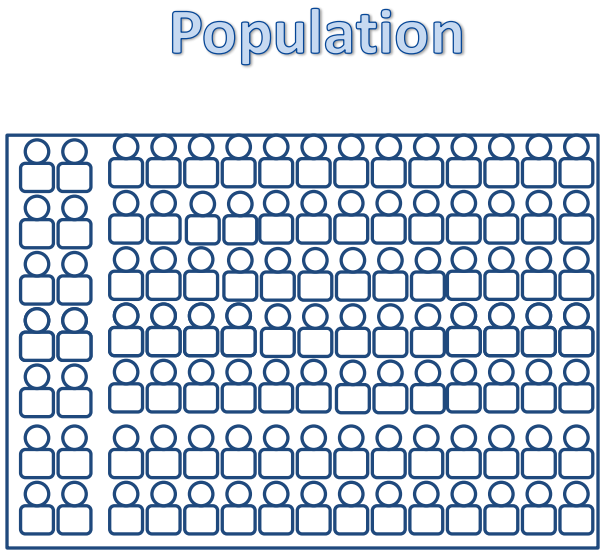
- sampling error
- non-sampling error

CENSUS

- large cost
- long duration
- less detailed questions
- high respondent burden
- non-sampling error

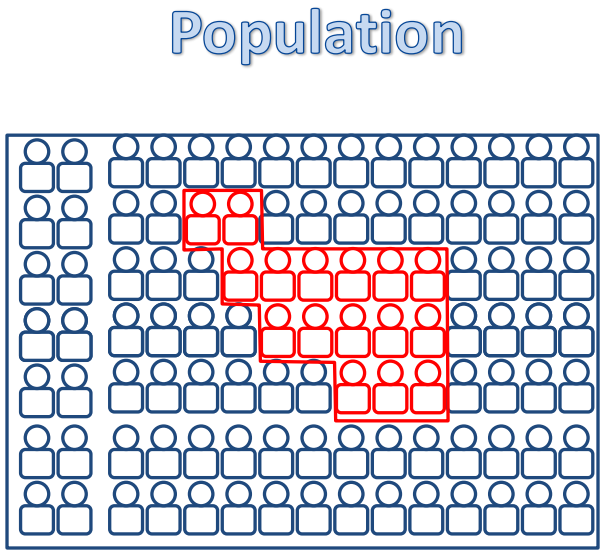
Planning a survey

Inference from sample survey data



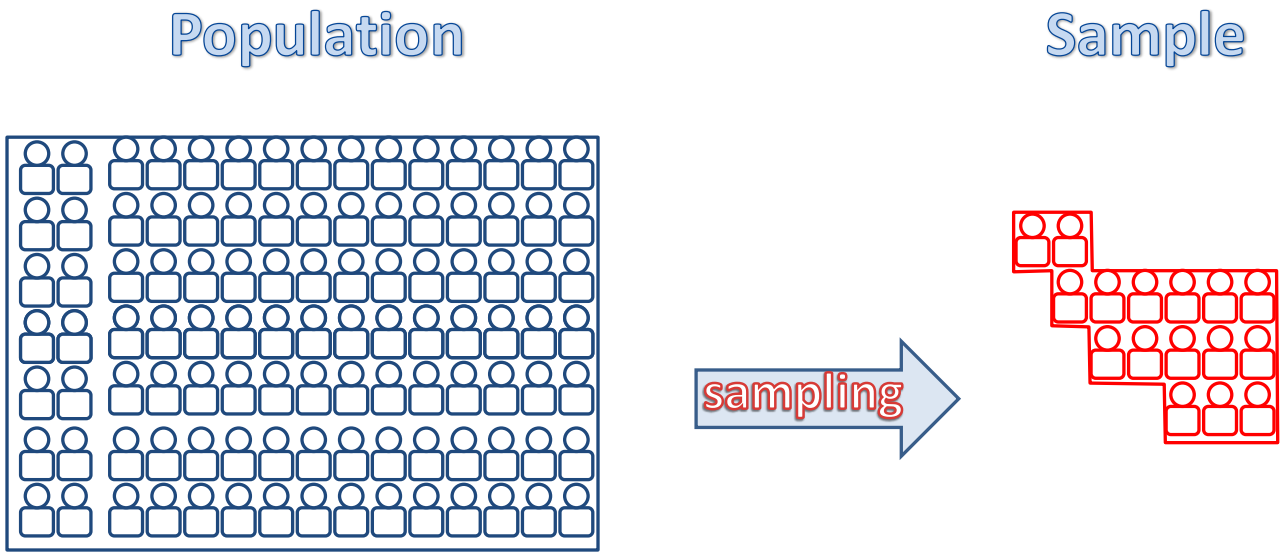
Planning a survey

Inference from sample survey data

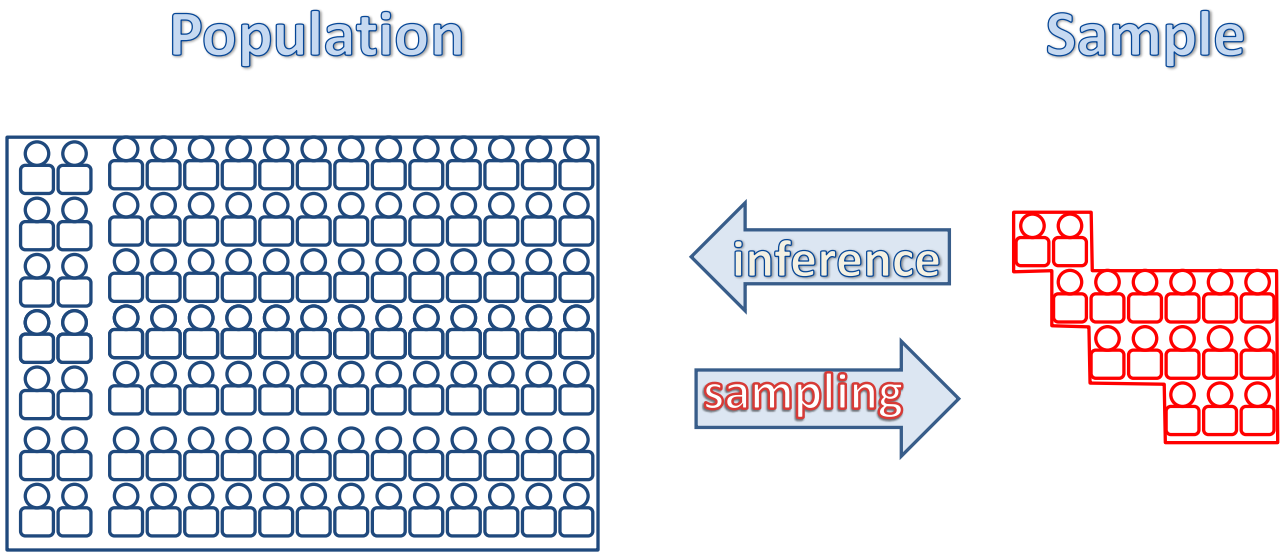


Planning a survey

Inference from sample survey data



Inference from sample survey data



Probability and non-probability sampling

There are two types of sampling methods:

- **probability sampling**
  every unit has a non zero chance of being selected, and that chance can be quantified, allowing clear and valid assumptions as basis for estimation
- **non-probability sampling**
  every item in the population does not have an equal chance of being selected. In this course we will not deal with non probability sampling, which is generally not recomended for official statistics

Sampling and non-sampling errors

Error sources in survey

Coverage error	is caused by the omission of some population units from the sampling frame
Sampling error	is error caused by the omission of some sampling frame units from the sample
Non-response error	is error caused by the omission of some sample units from the data
Interviewer error	is caused by differences in the ways that interviewers administer a survey
Respondent error	is caused by differences in the ways than sampling units respond to a survey (incomplete or inaccurate answers)
Coder error	is caused in defferences in the ways that the coders assign numeric codes to text answers
Instrument error	is caused by the wording of questions or other aspects of data collection instrument design

Sampling and non-sampling errors

$$\text{survey error} = \text{sampling error} + \text{non-sampling error}$$

Sampling error

Difference between estimate and population parameter as a result of sampling

Non-sampling error

Three sources:

- errors of non-observation
- errors of observation
- model assumption errors

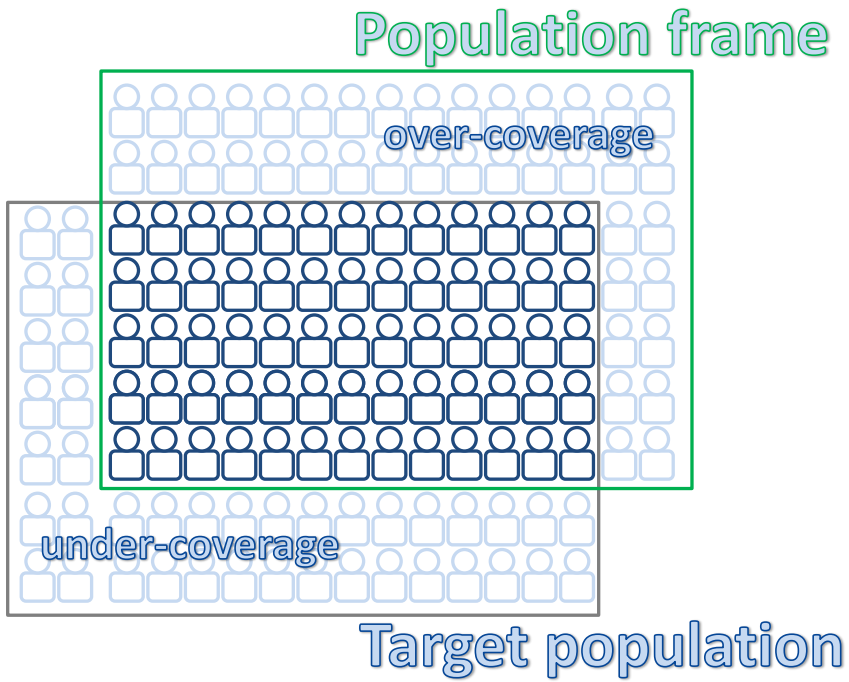
Sampling and non-sampling errors

Non-sampling errors

- Errors of non-observation
  - coverage error
  - non-response
    - complete
    - partial
- Errors of observation
  - measurement error
  - processing error
- Model assumption error

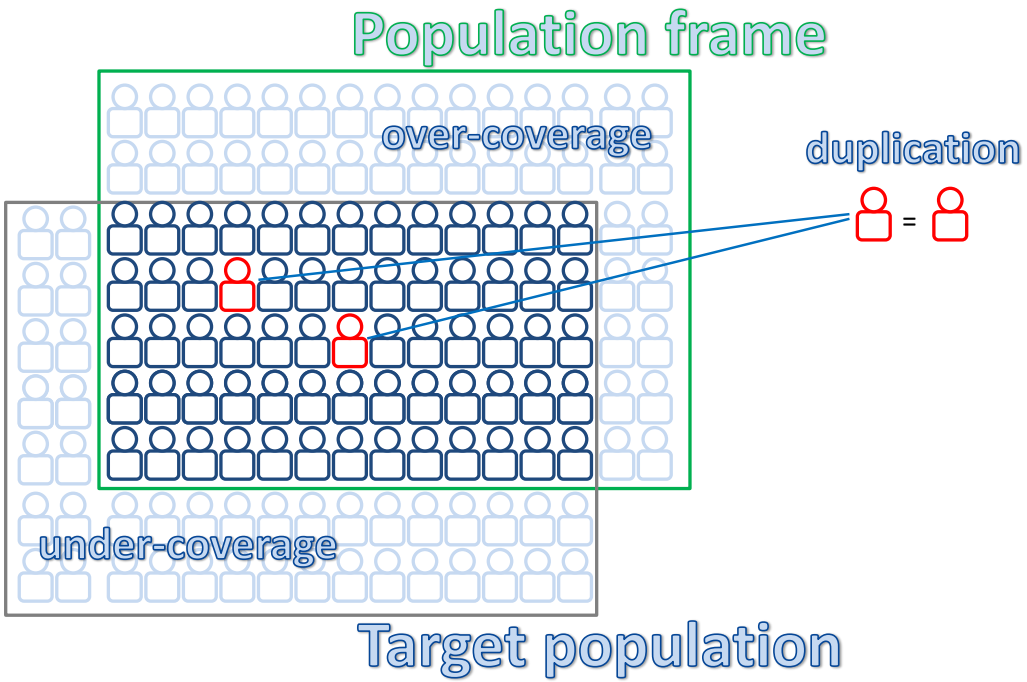
Sampling and non-sampling errors

Coverage Error: the survey frame errors
(Lessler & Kalsbeek, 1992, cap. 2; Nicolini *et al.*, 2014, cap. 4)



Sampling and non-sampling errors

Coverage Error: the survey frame errors
(Lessler & Kalsbeek, 1992, cap. 2; Nicolini *et al.*, 2014, cap. 4)



Sampling and non-sampling errors

Correcting the under-coverage error

- Improve frame
- Multiple frame sampling
- Estimation methods usign auxiliary information from target population

Non-response

Non-response is the failure to obtain complete measurements on the eligible survey sample

- **Unit Non-response**
  A sample unit does not provide any of the data required by the survey.
  Reasons:
  - Failure to locate or identify the sample unit
  - Failure to make contact the s.u.
  - Refusal of the sample unit to participate
  - Inability of the sample unit to participate
  - Inability to communicate
  - Accidental loss of the data

Non-response

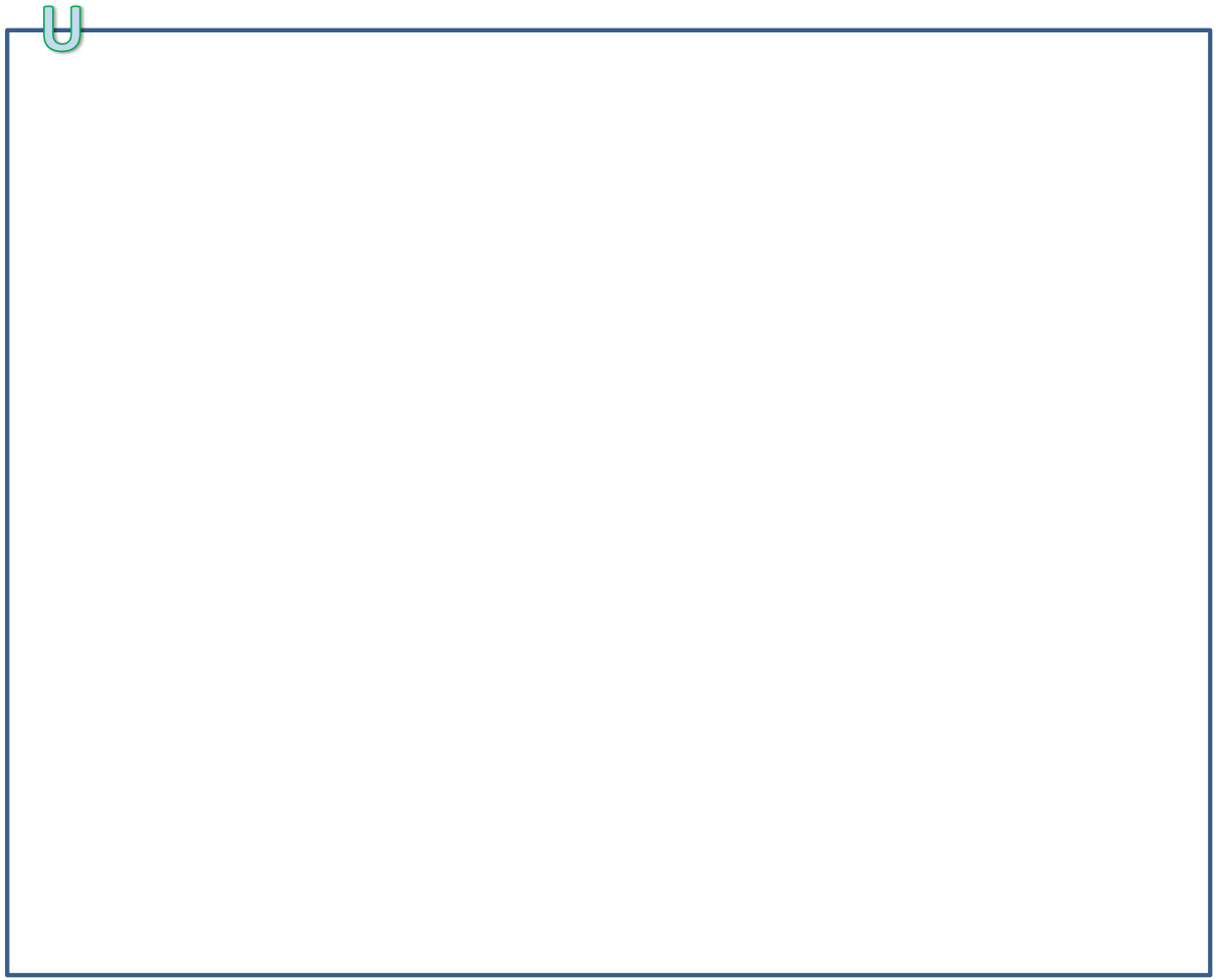
Non-response is the failure to obtain complete measurements on the eligible survey sample

- **Item Non-response**
  A sample unit participate in the survey but data for some survey items are not available for the analysis.

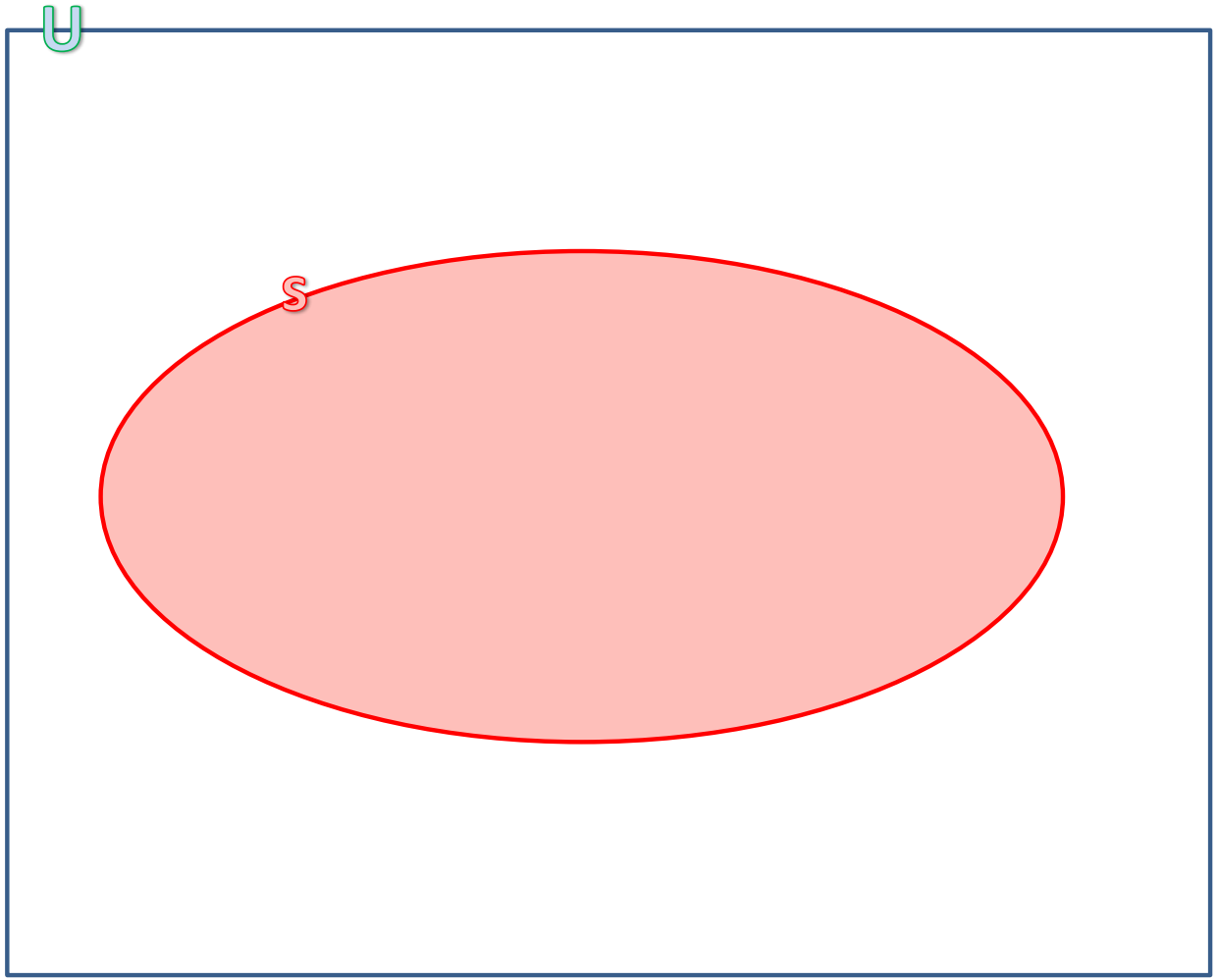
  Reasons:

  - Refusal to provide an answer
  - Inability to provide an answer
  - Other failure to provide an answer
  - Inadequate quality of provided answer (incomplete, implausible, etc.)

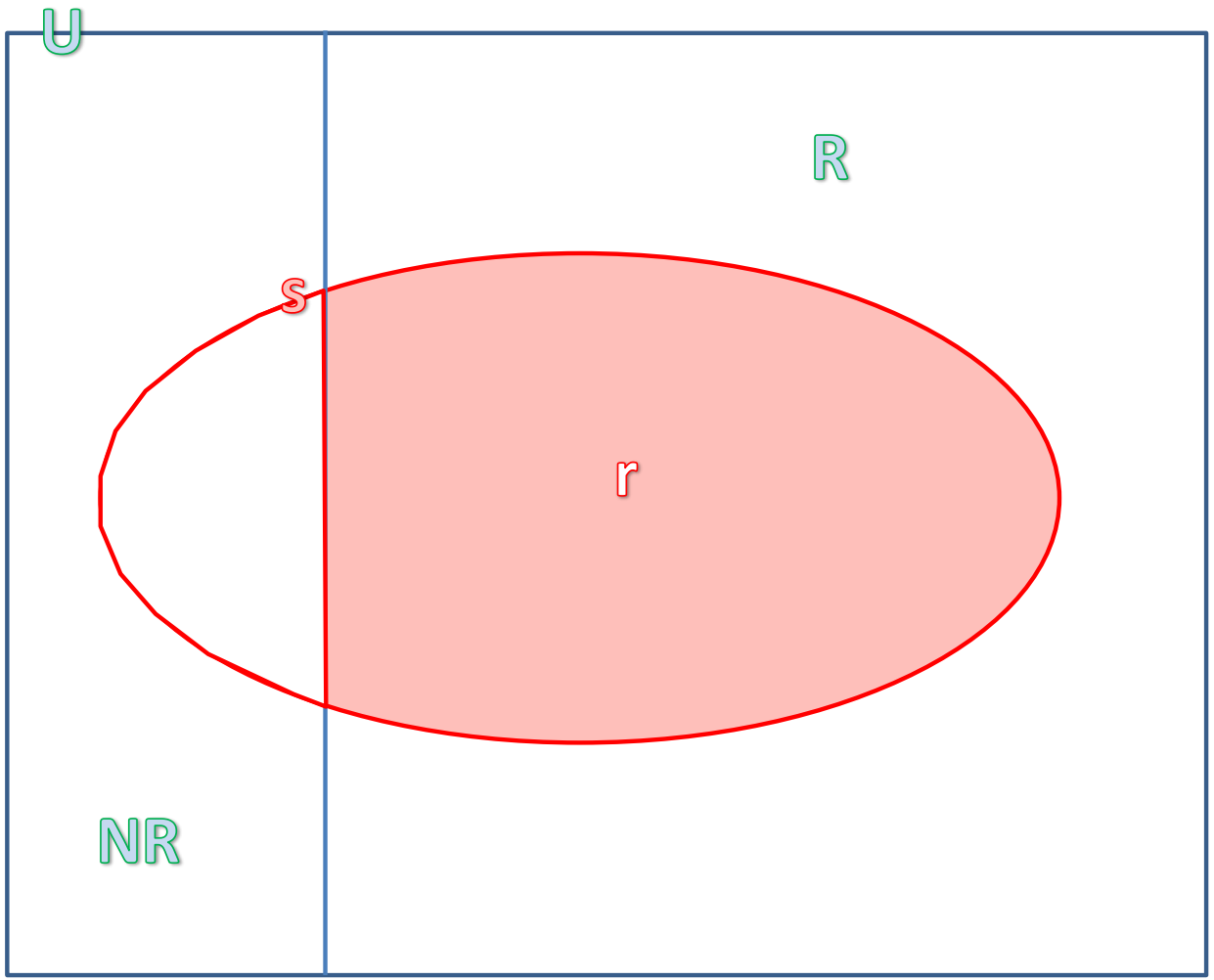
Non-response



Non-response



Non-response



Non-response

Non-response bias

$$\hat{t}_Y = N \frac{\sum_{k \in r} d_k \; y_k}{\sum_{k \in r} d_k} = N \frac{\hat{t}_{Y_R}}{N_R}$$

$$\mathbb{E}\left[\hat{t}_Y\right] = N \; \bar{y}_R$$

$$\begin{aligned} B\left(\hat{t}_Y\right) &= \mathbb{E}\left[\hat{t}_Y\right] - t_y \\ &\cong N \; \bar{y}_R - t_y \\ &= N \; \bar{y}_R - N_R \; \bar{y}_R - N_{NR} \; \bar{y}_{NR} \\ &= N_{NR} \left(\bar{y}_R - \bar{y}_{NR}\right) \end{aligned}$$

Non-response

- is a normal, but undesirable feature of essentially all sample surveys today
- those who happen to respond are often not "representative" for the population for which we wish to make inferences
- causes bias in the estimates
- bias is never completely eliminated, but we strive to reduce it as far as possible
- small variance no consolation, because (bias)2 can be the dominating part of MSE
- increases survey cost; follow-up is expensive
- will increase the variance, because fewer than desired will respond. But this can be compensated by anticipating the NR rate and allowing "extra sample size"

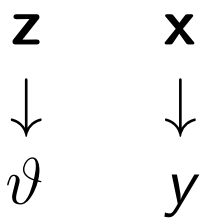
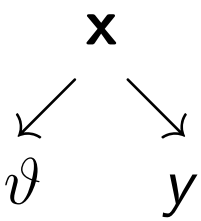
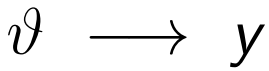
Non-response

Preventive actions

- extra sample size
- follow-up
- substitution
- sub-sampling

Non-response

Groves (2006)

separate causes model	common cause model	survey variable cause model
		
MCRA	MAR	MNAR

Little & Rubin (2002)

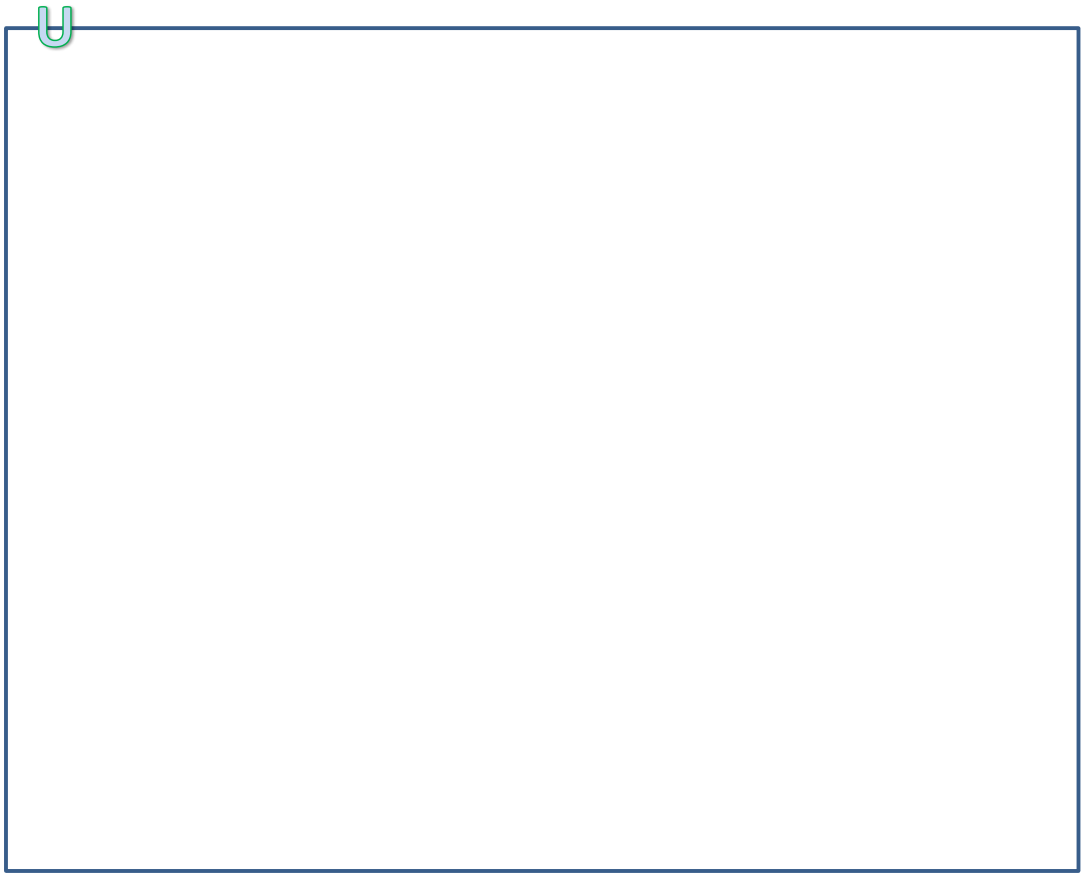
Non-response mechanism

Little & Rubin (2002)

- **MCAR (Missing Completly at Random)**
  Non-response is independent of all survey variables
  $\Longrightarrow$ no bias
- **MAR (Missing at Random)**
  Non-response depends on auxiliary variables only
  $\Longrightarrow$ correctable bias
- **NMAR (Not Missing at Random)**
  Non-response depends directly on the target variable of the survey
  $\Longrightarrow$ bias

Non-response

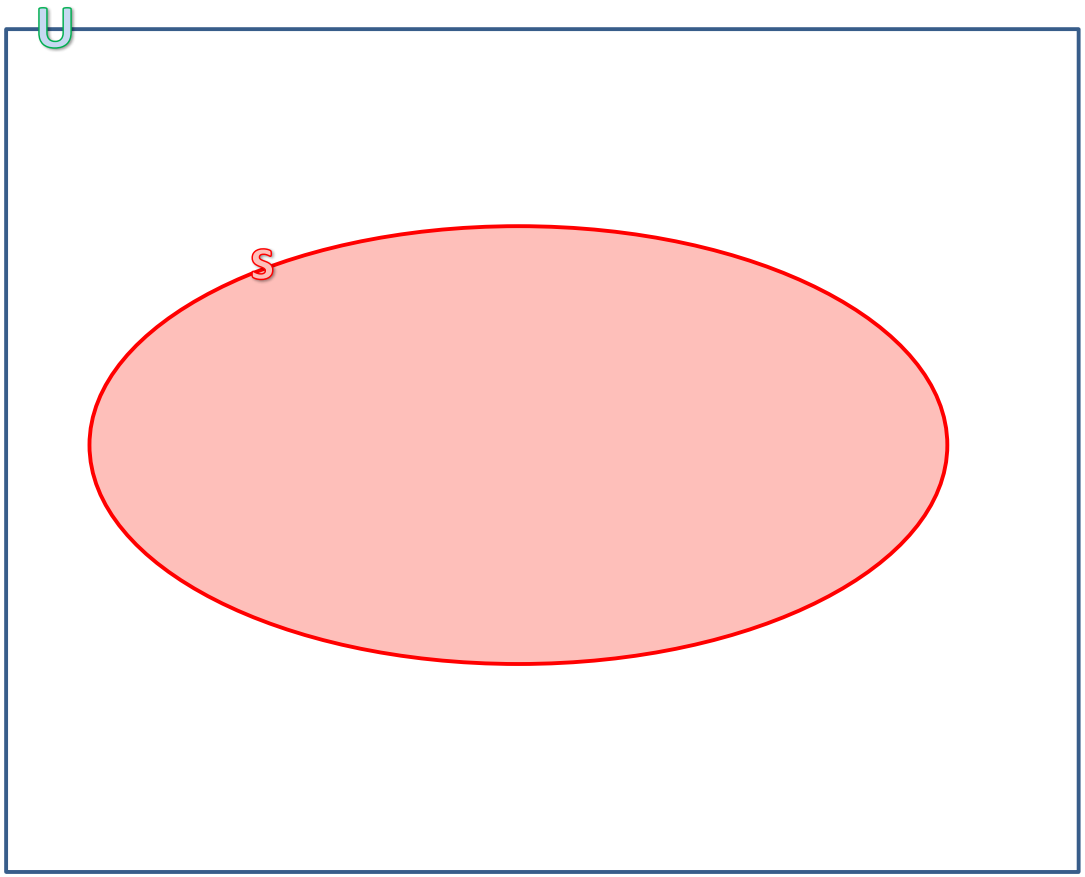
Two phases of selection approach



risposta1.pdf

Non-response

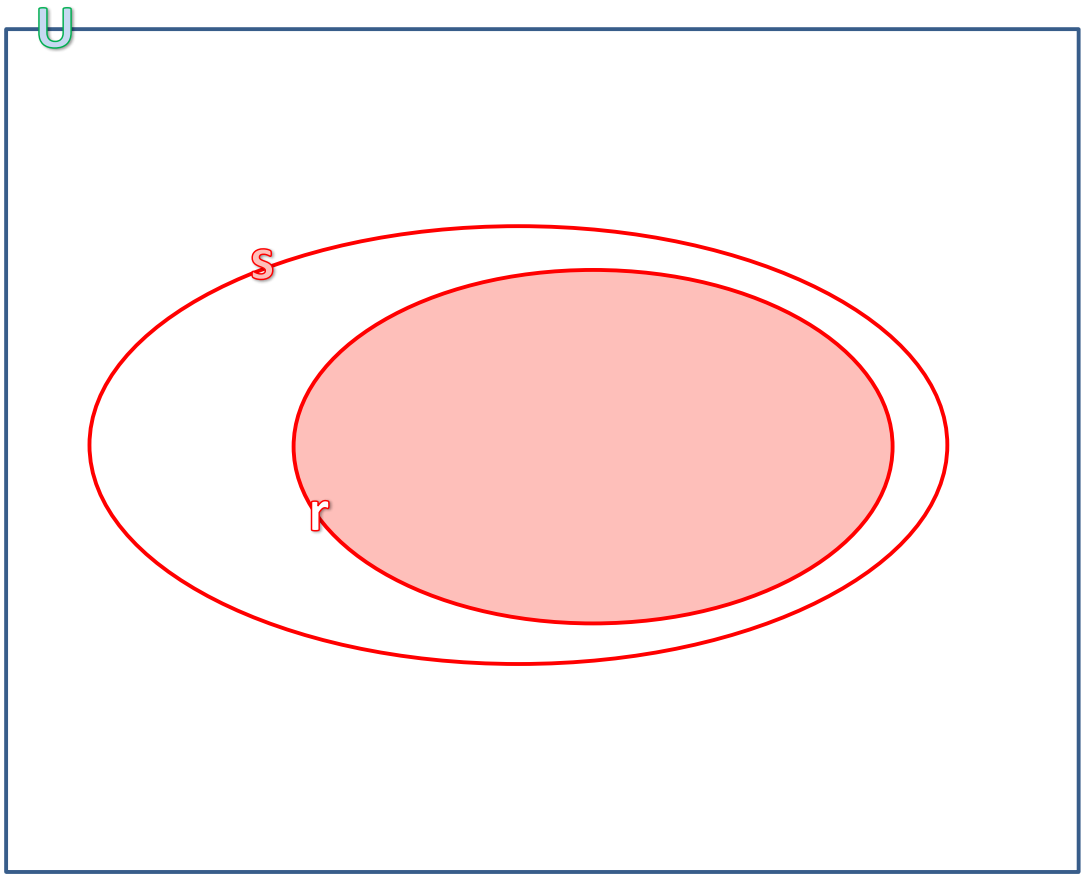
Two phases of selection approach



risposta22.pdf

Non-response

Two phases of selection approach



risposta23.pdf

Non-response

Two phases of selection approach

Inclusion probability

$$Pr\{k \in s\} = \pi_k$$

Response probability

$$Pr\{k \in r \mid k \in s\} = \vartheta_k$$

$$\Downarrow$$

$$\begin{aligned} Pr\{k \in r\} &= Pr\{k \in s\}\ Pr\{k \in r \mid k \in s\} \\ &= \pi_k\ \vartheta_k \end{aligned}$$

$$\hat{t}_Y = \sum_{k \in s} d_k^*\ y_k = \sum_{k \in r} \frac{y_k}{\pi_k\ \vartheta_k} = \sum_{k \in r} d_k\ \vartheta_k^{-1}\ y_k$$

$$\hat{t}_Y - t_Y = \left(\sum_{k \in s} \frac{y_k}{\pi_k} - \sum_{i=1}^{N} y_i\right) + \left(\sum_{k \in r} \frac{y_k}{\pi_k\ \vartheta_k} - \sum_{k\ \in\ r} \frac{y_k}{\pi_k}\right)$$

$$\begin{aligned} Var\left(\hat{t}_Y\right) &= Var\left(\mathbb{E}\left[\hat{t}_Y \mid s\right]\right) + \mathbb{E}\left[Var\left(\hat{t}_Y \mid s\right)\right] \\ &= Var\left(\hat{t}_Y\right) + \sum_{k \in r} (1 - \vartheta_k)\ \frac{y_k^2}{\pi_k^2} \end{aligned}$$

Method 1 - Constant probability

$$\vartheta = \vartheta_k = \frac{N_R}{N}$$

Method 2 - Logistic model

$$\vartheta_k = h\left(\mathbf{x}_k; \beta\right)$$

Method 3 - Regression estimate

$$\mathbf{x}_k \; \zeta = \vartheta_k^{-1} \qquad k = 1, \ldots, N$$

Method 4 - Calibrated estimator

$$d_k^* = 1 + \left(t_X - \hat{t}_{X_{HT_R}}\right) \left(\sum_{k \in r} d_k \; \mathbf{x}_k{}^t \; \mathbf{x}_k\right)^{-1} d_k \; \mathbf{x}_k{}^t$$

Non-response

Response homogeneity groups (RHG)

The elements in the sample (and those in the response set) are divided into groups about which it is assumed that all elements in the same group respond with the same probability

l	groups, $l = 1, \ldots, L$
n_l	l^{th} group size
r_l	respondent in the l^{th} group
$\bar{y}_l$	mean in the l^{th} group

Non-response

Response homogeneity groups (RHG)

Information at the population level

$\hat{t}_y = \sum_{l=1}^{L} N_l \, \bar{y}_l$

$B\left(\hat{t}_Y\right) = \sum_{l=1}^{L} N_{NR_l} \left(\bar{y}_{R_l} - \bar{y}_{NR_l}\right)$

Information at the sample level

$\hat{t}_Y = \sum_{l=1}^{L} \hat{N}_l \, \bar{y}_l$

Choice of weighting classes

- Choice of suitable auxiliary information
- Mean between respondent and non-respondent in the groups
- Sample sizes in classes not "too small"

Estimation and sampling error

Estimation

Estimation is the procedure of linking the information gathered from the sample back to the overall population, to determine a likely value for a parameter in the survey population

Sampling error

For every estimate the sampling error has to be evaluated by calculating the sampling variance: the average squared deviation about the estimator's average value, across al possible sample

Estimation and sampling error

HT **estimator**

Horvitz-Thompson Estimator (1952)

$$\hat{t}_{Y_{HT}} = \sum_{k \in s} \frac{y_k}{\pi_k} = \sum_{k \in s} y_k \; d_k$$

$$var\left(\hat{t}_{Y_{HT}}\right) = \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}$$

where $\boldsymbol{\Delta}_s = \left\{\left(\pi_{kl} - \pi_k \; \pi_l / \pi_{kl} \; \pi_k \; \pi_l\right)\right\}_{k,l \in s}$

Estimation and sampling error

HT **estimator**

- homogeneous linear estimator
- design unbiased;
- has minimum variance among the unbiased homogeneous linear estimators;
- its variance is null when the values of y in the population are proportional to the inclusion probabilities.

Estimation and sampling error

Generalized Regression (*GREG*) estimator

> *" [...] If the auxiliary variables are correlated with the target variables, an estimator that is more precise the the Horvitz-Thompson estimator can be constructed."*
>
> *Bethlehem & Keller (1987), p. 143*

It is implicitly assumed a linear regression superpopulation model

$$\xi : \mathbf{y} = \mathbf{XB} + \mathbf{e}$$

where,

- $\mathbf{e} = (e_1 \cdots e_k \cdots e_N)^t$
- $\mathbb{E}_\xi[e_k] = 0 \qquad k = (1, \cdots, N)$
- $Var_\xi[e_k] = \sigma_k^2 = q_k^{-1} \qquad k = (1, \cdots, N)$
- $Cov_\xi[e_k, e_l] = 0 \qquad k = (1, \cdots, N)$ e $l \neq k$

Estimation and sampling error

Generalized Regression (*GREG*) estimator

cfr. Cassel *et al.* (1979), Fuller (2002)

$$\hat{t}_{Y_{GREG}} = \hat{t}_{Y_{HT}} + \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_{HT}}\right)^t \hat{\mathbf{B}}$$

$$var\left(\hat{t}_{Y_{GREG}}\right) = \sum_{k \in s}\sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{\hat{e}_k}{\pi_k} \frac{\hat{e}_l}{\pi_l}$$

with

$$\hat{\mathbf{B}} = \left(\sum_{k \in s} \frac{q_k \mathbf{x}_k \mathbf{x}_k^t}{\pi_k}\right)^{-1} \sum_{k \in s} \frac{q_k \mathbf{x}_k y_k}{\pi_k}$$

$$\hat{\mathbf{e}}_s = \left(e_k = y_k - \mathbf{x}_k^t \hat{\mathbf{B}}\right)_{k \in s}$$

Properties of *GREG* estimator

- Model unbiased
  $\mathbb{E}_\xi [\mathbf{y}] = \mathbf{XB}$
- Asymptotically design unbiased (ADU)
  $\mathbb{E}_p \left[\hat{t}_{Y_{GREG}}\right] \asymp t_Y$
- Asymptotically design consistent
  $MSE_p \left[\hat{t}_{Y_{GREG}}\right] \asymp 0$
- Consistent with auxiliary totals
  $\hat{\mathbf{t}}_{\mathbf{X}_{GREG}} = \hat{\mathbf{t}}_{\mathbf{X}_{HT}} + \left(\mathbf{t}_\mathbf{X} - \hat{\mathbf{t}}_{\mathbf{X}_{HT}}\right)^t \hat{\mathbf{B}} = \mathbf{t}_\mathbf{X} \qquad \text{con } \hat{\mathbf{B}} = \mathbf{I}$

Estimation and sampling error

Calibration estimator *CAL*

Deville & Särndal (1992), Särndal (2007)

$$\hat{t}_{Y_{CAL}} = \sum_{k \in s} y_k \; w_k = \sum_{k \in s} y_k \; d_k \; \gamma_k$$

" [...] a strong correlation between the auxiliary variables and the study variable means that the weights that perform well for the auxiliary variables also should perform well for the study variables."

Deville & Särndal (1992) p. 376.

Estimation and sampling error

Calibration estimator *CAL*

Final weights

$$w_k = d_k + \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_{HT}}\right) \left(\sum_{k \in s} \frac{q_k \ \mathbf{x}_k \mathbf{x}_k^t}{\pi_k}\right)^{-1} \frac{q_k \ \mathbf{x}_k^t}{d_k}$$

The variance estimator

$$var\left(\hat{t}_{Y_{CAL}}\right) = \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \left(w_k \ \hat{e}_k\right) \left(w_l \ \hat{e}_l\right)$$

Estimation and sampling error

Calibration estimator *CAL*

$$\begin{cases} \min_{w_k} \left\{ \sum_{k \in s} G_k \left(w_k - d_k \right) / q_k \right\} \\ \\ \sum_{k \in s} \mathbf{x}_k \; w_k = \mathbf{t_x} \end{cases}$$

$G_k \left(w_k - d_k \right)$ is a pseudo-distance function

- $w_k \mapsto G_k \left(w_k - d_k \right)$ is non negative
- strictly convex
- continuously differentiable with respect to w_k
- $G_k \left(d_k - d_k \right) = 0$

Calibration estimator *CAL*

Le funzioni di distanza
Deville & Särndal (1992), Singh & Mohl (1996)

	Funzione	$G_k\,(w_k - d_k)$	γ_{ks}	Range w_k
a.	Lineare	$\frac{(w_k-d_k)^2}{2d_k}$	$1+\mathbf{x}_k^t\,\boldsymbol{\lambda}$	$-\infty \leq w_k \leq +\infty$
b.	Logaritmica	$w_k\ ln\left(\frac{w_k}{d_k}\right) - w_k + d_k$	$\exp(\mathbf{x}_k^t\,\boldsymbol{\lambda})$	$0 \leq w_k \leq +\infty$
c.	Chi-quadrato	$\frac{1}{2\ w_k}\left(\frac{w_k}{d_k}-1\right)^2$	$(1-2\ \mathbf{x}_k^t\,\boldsymbol{\lambda})^{-1/2}$	$0 \leq w_k \leq +\infty$
d.	Minima entropia	$-d_k\ ln\left(\frac{w_k}{d_k}\right) + w_k - d_k$	$(1-\mathbf{x}_k^t\,\boldsymbol{\lambda})^{-1}$	$0 \leq w_k \leq +\infty$
e.	Hellinger	$2\ d_k\left(\sqrt{\frac{w_k}{d_k}}-1\right)^2$	$(1-\frac{1}{2}\ \mathbf{x}_k^t\,\boldsymbol{\lambda})^{-2}$	$0 \leq w_k \leq +\infty$
f.	Logaritmica Troncata	$\left(\frac{w_k}{d_k}-L\right) ln\left(\frac{\frac{w_k}{d_k}-L}{1-L}\right) + \left(U-\frac{w_k}{d_k}\right) ln\left(\frac{U-\frac{w_k}{d_k}}{U-1}\right)$	$\frac{L(U-1)+U(1-L)\ \exp\left(\frac{U-L}{(U-1)(1-L)}\ \mathbf{x}_k^t\,\boldsymbol{\lambda}\right)}{(U-1)+(1-L)\ \exp\left(\frac{U-L}{(U-1)(1-L)}\ \mathbf{x}_k^t\,\boldsymbol{\lambda}\right)}$	$L \leq w_k \leq U$

$$w_k = d_k\,(1 + a\ \mathbf{x}_k^t \boldsymbol{\lambda})^{\frac{1}{a}} \qquad k = (1, \cdots, n)$$

When considering the Chi-squared pseudo-distance

$$\boldsymbol{\lambda} = \left(\mathbf{X}^t \boldsymbol{\Pi}^{-1} \mathbf{X}\right)^{-1} \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_{HT}}\right)$$

$$\mathrm{a} = 1$$

we have

$$\mathbf{w}_s = \mathbf{d}_s + \boldsymbol{\Pi}^{-1} \mathbf{X}_s \left(\mathbf{X}_s^t \boldsymbol{\Pi}^{-1} \mathbf{X}_s\right)^{-1} \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_{HT}}\right)$$

$$\begin{aligned} \hat{t}_{Y_{CAL}} &= \mathbf{y}_s^t \, \mathbf{d}_s + \mathbf{y}_s^t \boldsymbol{\Pi}^{-1} \mathbf{X}_s \left(\mathbf{X}_s^t \boldsymbol{\Pi}^{-1} \mathbf{X}_s\right)^{-1} \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_{HT}}\right) \\ &= \hat{t}_{Y_{HT}} + \hat{\mathbf{B}}^t \left(\mathbf{t_x} - \hat{\mathbf{t}}_{\mathbf{x}_{HT}}\right) \\ &= \hat{t}_{Y_{GREG}} \end{aligned}$$

Evaluation of sampling error

In the dissemination phase it is not possible to publish each estimate with its sampling error.

To provide the user with a tool to evaluate the precision of the estimate we usually make use of the regression model method.
A mathematical function (Generalized Variance Function, GVF) that relates each estimate with its sampling error is determined.

Evaluation of sampling error

Model used for categorical variables

$$log\left(\hat{\epsilon}^2\left(\hat{t}_{Y_d}\right)\right) = a + b\ log\left(\hat{t}_{Y_d}\right)$$
$$\hat{\epsilon}^2\left(\hat{t}_{Y_d}\right) = \sqrt{exp\left(a + b\ log\left(\hat{t}_{Y_d}\right)\right)}$$

- **Theoretical basis**: the coefficients of variation (cv) of absolute frequency estimates are decreasing function of the values of the estimates themselves.
- **Basic assumption**: estimates in the same domain share the same design effect
- Parameter a and b are estimated by the minimum least square method
- the model is fitted to the set of pairs $\left(\hat{t}_{Y_d}, \hat{\epsilon}^2\left(\hat{t}_{Y_d}\right)\right)$

Evaluation of sampling error

Model used for continuous variables

For estimates of totals of continuous variables a theoretical basis does not exist, so it is only possible to use empirical models, choosing the best fitting one. A model that usually fits well is

$$\hat{\sigma}\left(\hat{t}_{Y_d}\right) = a + b\ \hat{t}_{Y_d} + c\ \hat{t}_{Y_d}$$

Evaluation of sampling error

Example of table of parameter for synthetic presentation of sampling errors

Geographical Area	a	b	$r^2(\%)$
ITALY	8.484000	-1.096278	96.2
North-West	8.717029	-1.112776	95.0
North-Est	8.505412	-1.122544	96.1
Center	8.400121	-1.110399	96.1
South	7.502174	-1.036004	93.9
Islands	7.755317	-1.055478	92.6

Evaluation of sampling error

Example of table of interpolated values of sampling errors

Estimate	Italy	North-West	North-Est	Center	South	Islands
20,000	30.5	31.6	27.1	27.3	25.2	26.0
30,000	24.4	25.2	21.6	21.8	20.4	21.0
40,000	20.9	21.5	18.4	18.6	17.6	18.0
50,000	18.5	19.0	16.2	16.4	15.7	16.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
400,000	5.9	6.0	5.0	5.2	5.3	5.3
500,000	5.2	5.3	4.4	4.6	4.8	4.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2,500,000	2.4	2.4	2.0	2.1	2.3	2.3
5,000,000	1.5	1.5	-	-	-	-

Bibliography

Sampling Theory

Ardilly A., Tillé Y. (2005). *Sampling methods: exercises and solutions*. Springer, New York.

Bethel J.W. (1989). Sample allocation in multivariate surveys. *Survey Methodology*, 15(1):47-57.

Cicchitelli G., Herzel A., Montanari G.E. (1992). *Il campionamento statistico*. Il Mulino, Bologna.

Cochran W. (1977). *Sampling Teechniques*. Wiley & Sons Ltd., New York, 3rd edition.

Conti P., Marella D. (2011). *Campionamento da popolazioni finite: teoria e tecnica*. Springer, Milan.

Falorsi P.D., Ballim M., De Vitiis C., Sceoi G. (1998). Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte in Istat. *Statistica Applicata*, 10(2):235-257.

Lavalée P.(2007). *Indirect Sampling*. Springer, New York.

Sampling errors

Wolter K.M. (2006). *Introduction to variance estimation*. Springer, New York, 2nd edition.

Non sampling errors

Lessler J.T., Kalsbeek W.D. (1992). *Non sampling error in surveys*. Wiley & Sons Ltd., New York.

Nicolini G., Marasini D., Montanari G., Pratesi M., Ranalli M., Rocco E. (2013). *Metodi di stima in presenza di errori non campionari*. Springer-Verlag, Milano.

Nonresponse

Groves R.M. (1976). Nonresponse rates and nonresponse bias in household surveys. *Public opinion quarterly*, 70(5):646-675.

Rubin D.B. (1976). Inference and missing data. *Biometrika*, 63:581-592.

Rubin D.B. (1992). *Multiple imputation for non response in survey*. Wiley & Sons Ltd., New

Särndal C.-E., Lundström S. (1992). *Estimation in survey with nonresponse*. Wiley & Sons Ltd., New York. York.

Bibliography

Estimators

HT estimator

Horvitz D., Thompson D. (1952). A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663-685.

GREG estimator

Bethlehem J. G., Keller J. W. (1987). Linear weighting of sample survey data. *Journal of Official Statistics*, 3:141–153.

Cassel C. M., Särndal C.-E., Wretman J. H. (1979). Prediction theory for finite populations when model-based and design-based principles are combined. *Scandinavian Journal of Statistics*, 6:97–106.

Fuller W. A. (2002). Regression estimation for survey samples. *Survey Methodology*, 28.

CAL estimator

Deville J.-C., Särndal C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.

Särndal C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33: 99–119.

Singh A.C., Mohl C.A. (1996). Understanding calibration estimator in survey sampling. *Survey Methodology*, 22: 107–115.

Surveys

Istat (2006). Il sistema di indagini sociali multiscopo. *Metodi e Norme*, 31.

Istat (2006). La rilevazione sulle Forze di Lavoro: contenuti, metodologie, organizzazione. *Metodi e Norme*, 32.

Istat (2008). L'indagine europea sui redditi e le condizioni di vita delle famiglie (It-Silc). *Metodi e Norme*, 37.