SYLLABUS DEL CORSO

Metodi per le Indagini Campionarie M

2122-1-F8204B007

Learning objectives

This course aims at providing the students with all the necessary notions to face statistical inference for finite populatins. In the first part of the course, the teacher will introduce the most important sampling schemes for finite populations and he will define different kinds of estimators for means/totals. In the last part of the course some important applications will be discussed, with a special focus on the most recent techniques to privatize a dataset.

Contents

The first part of the course aims at providing the students with a solid theoretical background to face statistical inference for finite populations. More precisely the most important sampling schemes will be analyzed: simple random sampling, stratified, systematic, cluster sampling, multi-stage design, etc. Besides different kinds of estimators for means, totals and proportions will be defined and studied, among them we consider the ratio and regression estimators. In the second part of the course, some applications will be analyzed, among which the randomized response techinque, disclosure risk assessment and differential privacy. ⎯⎯⎯⎯⎯⎯

Detailed program

1. INTRODUCTION AND BASIC NOTIONS
Historical background, the difference between a sample survey and a census. The notion of population, sample, variable. The underlying probability space in a sample survey and the notion of sampling design. Non-probabilistic sampling schemes.

2. SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT
The Horvitz-Thompson estimator for the total for simple random sampling without replacement. Evaluation of the variance and the unbiased estimator. Hàjek theorem (without proof) and the construction of asymptotic confidence

intervals. Estimators for means and proportions. The problem of sample size in simple random sampling.

3. SIMPLE RANDOM SAMPLING WITH REPLACEMENT
The Hansen-Hurwitz estimator: derivation in the general case. Analysis of the estimator for simple random sampling with replacement. The notion of design effect.

4. VARYING PROBABILITY SAMPLING

The variance of the Hansen-Hurwitz and the Horvitz-Thompson estimator. The notion of auxiliary variable. Different sampling schemes with varying probabilities.

5. STRATIFIED SAMPLING

The definition of the sampling scheme. Estimators for mean, total in the stritified sampling scheme.
Optimum and proportional allocation. Poststratification.

6. RATIO ESTIMATOR
The use of auxiliary variables for defining different and more efficient estimators. The ratio estimator: definition, the method of linearization to estimate the variance, comparision with simple random sampling.
Ratio estimator for stratified random sampling.
Regression estimator: definition, variance.

8. CLUSTER AND MULTISTAGE SAMPLING
Cluster sampling: basic properties. The unbiased estimator and the ratio estimator in cluster sampling. The analysis of variance: variability within and across clusters, total variability of the population, the index of homogeneity within clusters.
Multistage design: definition, estimator of the total and evaluation of the variance.

9. APPLICATIONS
Disclosure risk assessment. Randomized response techniques. Differential Privacy.

10. NON-SAMPLING ERRORS
The notion of non-sampling errors.
The randomized response technique: Warner's and Simmons's methods.

Prerequisites

It is recommended the knowledge of the arguments of Mathematical Analysis and Statistics, taught during the bachelor degree.

Teaching methods

Traditional lessons and class exercises.
During the emergency Covid-19 the lessons will be live via Webex.

Assessment methods

The exam is written, the oral test is not mandatory. In the written test, _____

_____

Textbooks and Reading Materials

As for the first part, the following books are recommended:

1) G. Cicchitelli, A. Herzel, G.E. Montanari. Il campionamento statistico. Il Mulino, 1997.

2) P.L. Conti, D. Marella. Campionamento da popolazioni finite. Springer-Verlag Mailand, 2012.

3) S. Thompson. Sampling. Wiley, 2012.

As for the second part (disclosure risk assessment and differential privacy):

1) Dwork, C., Roth A. The Algorithmic foundations of Differential Privacy. 2014.

2) Papers suggested during the course.

Semester

Spring semester.

Teaching language

The lessons will be held in Italian.