



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Metodi per le Indagini Campionarie M

2526-1-F8206B005

Obiettivi formativi

Questo corso si propone di introdurre gli strumenti fondamentali della teoria dei campioni necessari per l'inferenza da popolazioni finite. Nella prima parte del corso verranno analizzati i più importanti piani di campionamento probabilistici e verranno definite diverse tipologie di stimatori per i caratteri di interesse della popolazione. Nella seconda parte del corso, verranno discusse alcune applicazioni, in particolare verranno affrontate le tecniche più moderne per la privatizzazione dei dati. Gli obiettivi formativi del corso sono fondamentali per l'area Statistica. Il corso contribuisce a consolidare le capacità di apprendimento e di aggiornamento nell'ambito dei metodi avanzati di inferenza statistica, in coerenza con l'area di apprendimento "Statistica" del corso di laurea magistrale in Scienze Statistiche ed Economiche.

Contenuti sintetici

La prima parte del corso intende fornire agli studenti una solida base teorica dei metodi di campionamento da popolazioni finite. In particolare verranno presentati diversi piani di campionamento: casuale semplice, stratificato, sistematico, a grappoli, multistadiale, piani di campionamento non probabilistici. Contestualmente saranno introdotte diverse tipologie di stimatori per totali, medie e proporzioni, tra cui gli stimatori quoziente e per regressione. Nella seconda parte del corso verrà analizzato il metodo delle risposte casualizzate, il problema della privatizzazione dei dati (in particolare il concetto di differential privacy). Verrà introdotto il concetto di indice di disclosure (divulgazione) per quantificare la rischiosità di ledere la riservatezza dei dati forniti dal rispondente quando essi vengono pubblicati da un ufficio statistico. Infine verrà affrontato il problema degli errori non campionari, tra cui quello delle mancate risposte, ed il metodo della ponderazione dei dati. Il corso prevede lezioni teoriche ed esercizi.

Programma esteso

1. INTRODUZIONE AL CORSO E NOZIONI DI BASE

La differenza tra indagini campionarie e censuarie. Cenni storici sulle indagini campionarie. Il concetto di indagine statistica, popolazione, campione, caratteri. Lo spazio campionario e la nozione di piano di campionamento. I campionamenti non probabilistici.

2. CAMPIONAMENTO CASUALE SEMPLICE SENZA RIPETIZIONE

Lo stimatore di Horvitz-Thompson del totale e della media nel campionamento casuale semplice senza ripetizione. Calcolo della varianza dello stimatore e stima corretta della varianza. Cenni al Teorema di Hájek (senza dimostrazione) e alla costruzione di intervalli di confidenza asintotici per medie e totali. Gli stimatori per le proporzioni. Il problema della stima della dimensione campionaria nel campionamento casuale semplice.

3. CAMPIONAMENTO CASUALE SEMPLICE CON RIPETIZIONE

Lo stimatore di Hansen-Hurwitz del totale e della media: derivazione generale dello stimatore nel caso di probabilità di estrazione variabili. Analisi degli stimatori per il campionamento casuale semplice con ripetizione: varianza dello stimatore e stima corretta della varianza. Il concetto di design effect.

4. CAMPIONAMENTI A PROBABILITÀ VARIABILI

Calcolo della varianza per lo stimatore di Hansen-Hurwitz. Calcolo della varianza per lo stimatore di Horvitz-Thompson.

Il concetto di misure d'ampiezza. Diversi metodi di campionamento a probabilità variabili.

5. CAMPIONAMENTO STRATIFICATO

Il concetto di stratificazione. Stimatori della media e del totale nel campionamento stratificato. Stratificazione con allocazione proporzionale e allocazione ottima delle unità. La poststratificazione.

6. STIMATORE RAPPORTO

L'uso delle variabili ausiliarie per definire stimatori più efficienti. Lo stimatore rapporto: definizione, approssimazione della varianza mediante linearizzazione, confronto con il campionamento casuale semplice. Lo stimatore rapporto nel campionamento stratificato: stimatore quoziente combinato e separato. Lo stimatore per regressione: definizione, analisi della varianza.

7. CAMPIONAMENTO A GRAPPOLI E MULTISTADIALE

Il piano di campionamento a grappoli: generalità. Lo stimatore corretto del totale e lo stimatore quoziente nel campionamento a grappoli. Analisi della varianza dello stimatore: variabilità nei grappoli, variabilità tra i grappoli, variabilità complessiva della popolazione, indice di omogeneità nei grappoli. Efficienza dello stimatore in funzione dell'indice di omogeneità nei grappoli. Il campionamento sistematico come caso particolare del campionamento a grappoli.

Il campionamento multistadiale: definizione dello stimatore del totale e calcolo della varianza approssimata.

8. APPLICAZIONI

La valutazione del rischio associato alla divulgazione dei dati a fini statistici, alcuni metodi di privatizzazione dei dati (differential privacy e tecnica delle risposte randomizzate). Diversi indici per misurare il rischio nella divulgazione dei dati, quando essi sono nella forma di tabelle di frequenze.

9. I DOMINI DI STUDIO

Il concetto di dominio di studio, classificazione dei domini in base alla dimensione. Stima dei parametrine dei domini di studio maggiori e minori. Cenni alle problematiche dei mini-domini e domini rari.

10. GLI ERRORI NON CAMPIONARI

Diverse tipologie di errori non campionari: errori di copertura, errori da mancate risposte ed errori di misurazione. Metodi per la riduzione degli errori non campionari.

Metodo delle risposte casualizzate: metodo di Warner e metodo di Simmons.

11. I METODI DI CALIBRAZIONE

Generalità sulle strategie di calibrazione, stimatori calibrati, calibrazione lineare, metodo del raking ratio.

Prerequisiti

Per seguire in modo proficuo il corso di Metodi per le Indagini Campionarie si consiglia la conoscenza degli argomenti trattati nei corsi di Analisi Matematica e Statistica a livello di una laurea triennale in Statistica.

Metodi didattici

Sono previste circa 40 ore di lezioni frontali, in cui verranno affrontati numerosi esempi ed esercizi. Le lezioni saranno svolte in modalità erogativa in presenza.

Modalità di verifica dell'apprendimento

L'esame è costituito da una prova scritta, l'orale è facoltativo. La prova scritta è costituita da esercizi e da alcune domande di teoria. Gli esercizi mirano ad accertare la comprensione degli argomenti trattati e la capacità dello studente di sfruttare gli strumenti di teoria dei campioni per risolvere problemi concreti. Le domande di teoria servono a verificare la conoscenza dei concetti di base di teoria dei campioni. Una ed una sola delle domande di teoria concerne una dimostrazione vista durante il corso.

L'orale è facoltativo e può essere chiesto sia dallo studente che dal docente. L'esame orale verte su tutto il programma del corso e deve essere svolto pochi giorni dopo lo scritto, in base alle disponibilità del docente. In tal caso il voto finale è una media della prova scritta e della prova orale.

Durante lo scritto è consentito l'uso della calcolatrice scientifica, ma non è ammesso l'uso di appunti, libri e strumenti tecnologici.

Testi di riferimento

Per la prima parte del corso consigliano i seguenti testi:

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.

Per la seconda parte del corso (divulgazione dei dati a fini statistici e valutazione del rischio)

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

2. Articoli indicati dal docente durante il corso

Periodo di erogazione dell'insegnamento

Il corso verrà erogato nel primo semestre, secondo ciclo.

Lingua di insegnamento

L'insegnamento è erogato in lingua italiana.

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
