UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Linear Models for Categorical Data

2223-1-F8203B010-F8203B011M

Obiettivi formativi

Il corso ha quale obiettivo la trattazione dei modelli lineari per dati categoriali secondo due diverse impostazioni. La prima riguarda il modello lineare generale (GLM), in particolare i modelli ANOVA ad una o più vie e il modello ANCOVA. La seconda impostazione riguarda i modelli lineari generalizzati, in particolare il modello log-lineare di Poisson per dati di conteggio e il modello logistico binomiale, in un'ottica di GLM. L'analisi dei casi empirici è svolta con il software SAS.

Conoscenza e comprensione. Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- Principali modellistiche a natura lineare in presenza di variabili categoriali, siano esse sul versante della variabile dipendente o sul versante delle variabili indipendenti o di entrambi i versanti
- Metodi per l'approfondimento delle analisi mediante opportuna definizione di funzioni dei parametri del modello, finalizzate in particolare al confronto fra specifici gruppi o categorie di unità statistiche
- Principali procedure implementate nel software SAS per la costruzione di modelli lineari per dati categoriali e il successivo approfondimento delle analisi con le relative rappresentazioni grafiche
- Lettura e interpretazione degli output delle analisi prodotte con SAS.

Capacità di applicare conoscenza e comprensione. Alla fine dell'insegnamento gli studenti saranno in grado di:

- Scegliere il modello lineare più adeguato in base alla natura della variabile dipendente
- Stabilire l'opportunità di applicazione di uno specifico modello lineare per dati categoriali in base a scopi formulati a priori anche in relazione al tipo di studio e alla natura dei dati a disposizione
- Interpretare i parametri di interazione inseriti nel modello relativamente a due variabili categoriali considerate congiuntamente e approfondire i risultati delle analisi in base a strategie che tengano conto della significatività o meno di tali interazioni
- Utilizzare le principali procedure di SAS dedicate alla modellistica per dati categoriali superando le impostazioni di default e usando gli statement più avanzati per la personalizzazione e l'approfondimento delle analisi.

L'insegnamento consente allo studente di acquisire le principali basi teoriche e applicative relativamente alla specificazione e alla costruzione dei modelli lineari per dati categoriali necessarie in qualsiasi contesto lavorativo in cui si utilizzino file di dati e che rappresentano una base imprescindibile per il proseguimento del percorso universitario.

Contenuti sintetici

Modello lineare generale (GLM), modello ANOVA a una e più vie e modello ANCOVA. Modelli lineari generalizzati (GzLM), modello logistico binomiale e modello log-lineare di Poisson. Applicazioni a dati reali e sperimentali con il software SAS.

Programma esteso

- La teoria del modello lineare generale (GLM): specificazione del modello, inversa generalizzata, funzioni stimabili, ipotesi testabili. Relazione con il metodo di stima dei minimi quadrati vincolati: approcci *sum-to-zero* e *set-to-zero linear constraints*. Parametrizzazione degli effetti e della categoria di riferimento. Contrasti
- Casi particolari di GLM: modelli ANOVA ad effetti fissi a una e a più vie, modello ANCOVA. PROC GLM di SAS
- Modelli lineari generalizzati (GzLM): legge di distribuzione della variabile risposta, funzione link, specificazione del modello, metodo di stima di massima verosimiglianza, proprietà degli stimatori, bontà di adattamento del modello, intervalli di confidenza e verifica di ipotesi
- Casi particolari di GzLM: modello log-lineare di Poisson per dati di conteggio e modello logistico binomiale, in un’ottica di GLM. PROC GENMOD di SAS

Prerequisiti

Per questa attività formativa è consigliata la conoscenza degli argomenti trattati nei corsi di base di Inferenza Statistica e di Modelli Statistici.

Metodi didattici

Lezioni teoriche in aula ed esercitazioni pratiche in laboratorio statistico-informatico con il software SAS.

Modalità di verifica dell'apprendimento

L'esame consiste nella preparazione di un'analisi statistica di dati con il software SAS (secondo le modalità specificate sulla piattaforma e-learning del corso), che dovrà essere poi discussa in sede d'esame, e in una prova scritta (durata: 2 ore) che ha ad oggetto argomenti sia a natura teorica sia a natura pratica.

Le domande a natura teorica sono di carattere generale e consentono di verificare le conoscenze teoriche acquisite in merito alla logica e agli aspetti più avanzati della costruzione della modellistica in presenza di dati categoriali (siano essi sul versante della variabile dipendente e/o delle variabili indipendenti), della distinzione fra le varie forme di parametrizzazione del modello, dei problemi legati alle nozioni di stimabilità e di testabilità di funzioni dei parametri, e della conduzione dell'inferenza statistica nell'ambito di tale modellistica. Permettono inoltre di verificare la capacità di utilizzare in autonomia il linguaggio simbolico-formale statistico e di fornire in modo appropriato le definizioni. Le parti con natura più metodologica sono oggetto di una domanda facoltativa che permette di verificare la capacità di dimostrare analiticamente i principali risultati teorici più avanzati.

Le domande pratiche riguardano sia l'individuazione, la costruzione e l'utilizzo della modellistica più opportuna con riferimento a problemi tratti da situazioni e da dati reali, sia la definizione del disegno dell'analisi più adeguato a soddisfare obiettivi di studio definiti a priori. L'analisi statistica di dati da preparare prima e presentare poi in sede d'esame costituisce la parte dell'esame in cui questi aspetti hanno maggior enfasi, poiché richiede allo studente di lavorare in modo critico e in piena autonomia, soprattutto nella definizione e nella realizzazione degli obiettivi dello studio. Le domande pratiche consentono in definitiva di verificare la capacità di comprensione delle problematiche sottoposte e di proporne soluzioni in termini di analisi, la competenza nel leggere e interpretare i risultati delle analisi, e l'abilità nel realizzare le analisi richieste mediante le procedure di SAS.

Per quanto riguarda più nel dettaglio l'analisi statistica di dati da svolgere con SAS, la metodologia da utilizzare viene assegnata nominalmente e in modo casuale (mediante l'ausilio di un generatore di numeri casuali) a ciascuno studente iscritto alla piattaforma e-learning del corso. L'analisi statistica deve essere preparata prima della prova d'esame seguendo una traccia specifica relativa alla metodologia assegnata e pubblicata alla fine del corso sulla piattaforma e-learning. In sede d'esame si dovrà poi presentare la stampa dell'output secondo le modalità specificate sulla piattaforma e-learning del corso.

Considerata l'abbondanza di materiale didattico messo a disposizione dalla docente sulla piattaforma e-learning del corso, non si prevede alcuna distinzione fra esami per studenti frequentanti ed esami per studenti non frequentanti. Infine non si prevedono prove in itinere.

Testi di riferimento

- Materiale didattico della docente pubblicato sul sito e-learning del corso (ad accesso riservato)
- Agresti, A. (2002), Categorical Data Analysis, Second Edition, New York: John Wiley & Sons
- Dobson, A., and Barnett, A. (2018), An Introduction to Generalized Linear Models, Boca Raton, FL: Chapman Hall/CRC, Fourth edition
- Littell, R. C., Freund, R. J., and Spector, P. C. (2002), SAS for Linear Models, 4th Edition, Cary, NC: SAS Institute Inc.
- Searle, S. R., and Gruber, M.H.J. (2017), Linear Models, 2nd Edition, John Wiley & Sons, Hoboken, New Jersey

Periodo di erogazione dell'insegnamento

I Semestre, II periodo

Lingua di insegnamento

Italiano

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ