

REGOLAMENTO DIDATTICO E ORGANIZZATIVO DEL CORSO *MINOR* IN

Analisi dei dati

PARTE I – INFORMAZIONI GENERALI

Proposta di attivazione

Prima attivazione

Anno accademico

2023-24

Dipartimento di riferimento

Economia

Corso interdipartimentale

Scienze Politiche; Scienze della Formazione

Organo di gestione

Consiglio di Dipartimento di Economia

Collaborazione con ente esterno

nessuno

PARTE II – ORGANIZZAZIONE DIDATTICA E AMMINISTRATIVA

Il corso in breve

Il minor si propone di arricchire gli iscritti di solide competenze statistiche, in particolare nell'ambito dell'analisi dei dati, finalizzate all'interpretazione e al supporto dei processi decisionali riguardanti fenomeni di natura socio-politica ed economica, e attori pubblici o privati.

Si propongono due percorsi distinti, uno più orientato alle metodologie di analisi legate all'econometria e alla finanza, uno più legato all'analisi dei dati in ambito manageriale, ambientale e in particolare alle tecniche di elaborazione di *big data*.

In entrambi i percorsi verrà data enfasi alle applicazioni pratiche e ai casi di studio con l'ausilio di software.

Il progetto formativo è costruito in modo da includere almeno un insegnamento di Laurea Magistrale di ciascuno dei dipartimenti proponenti, in maniera tale che gli studenti iscritti a corsi di studio dei tre dipartimenti che si iscrivano al minor possano vedere riconosciuti i crediti relativi ad insegnamenti già superati.

I due percorsi hanno un insegnamento comune e i restanti 18 crediti differenziati, oltre che per tematiche, per crediti degli insegnamenti previsti. Ciò è reso necessario dalla diversa "metrica" degli insegnamenti dei tre dipartimenti.

Percorso A:	
9 cfu	Analisi statistiche per le scienze sociali/Statistica per il management/Statistical methods in economics
9 cfu	Metodi statistici per l'economia e la finanza
9 cfu	Advanced topics in statistical learning

Percorso B:	
9 cfu	Analisi statistiche per le scienze sociali/Statistica per il management/ Statistical methods in economics
6 cfu	Statistical learning/Statistica ambientale
6 cfu	Big data and machine learning/Statistica per la valutazione
6 cfu	Metodi statistici di analisi dei dati

Lista delle attività didattico-formative che compongono il corso

Codice AD	attività didattico- formativa e relativo SSD	Sem.	docente e relativo SSD	ore di didattica assistita	CFU
21210418	Advanced topics in statistical learning SECS-S/01	1	Maria Maddalena Barbieri SECS-S/01	60	9
21210001	Analisi statistiche per le scienze sociali SECS-S/01	1	Silvia Terzi Francesca Fortuna SECS-S/01	60	9
21810762	Big data and machine learning SECS-S/01	1	Domenico Cucina SECS-S/01	36	6
22902266	Metodi statistici di analisi dei dati SECS-S/01	2	Giuseppe Bove SECS-S/01	36	6
21210457	Metodi statistici per l'economia e finanza SECS-S/01	2	Alessia Naccarato SECS-S/01	60	9
21810418	Statistica ambientale SECS-S/05	2	Francesco Lagona SECS-S/05	36	6
21201420	Statistica per il management SECS-S/01	2	Paola Vicard SECS-S/01	60	9
22910114	Statistica per la valutazione SECS-S/01	1	Rosa Capobianco SECS-S/01	36	6
21210229	Statistical learning SECS-S/01	2	Francesca Fortuna SECS-S/01	40	6
21210065	Statistical methods in economics SECS-S/01	1	Caterina Conigliani SECS-S/01	60	9

Ulteriori informazioni sulle attività didattico-formative

1. Analisi statistiche per le scienze sociali – Silvia Terzi, Francesca Fortuna

(9 cfu - Primo semestre)

Obiettivi formativi:

Il corso si propone di fornire allo studente un'adeguata preparazione sui concetti e sui metodi statistici utilizzati nella rilevazione, elaborazione e sintesi dei dati relativi a fenomeni economici e sociali del Paese; privilegia gli aspetti sostantivi e interpretativi dei modelli statistici e si propone come manuale d'istruzioni utile all'impiego corretto delle tecniche multivariate. Per ciascun argomento si dedicherà particolare attenzione alle condizioni specifiche di applicabilità delle varie tecniche, chiarendone le potenzialità analitiche e dedicando ampio spazio agli esempi concreti trattati. Il corso ha un taglio applicativo.

Prerequisiti:

Tra i prerequisiti del corso figurano esclusivamente i concetti-base della statistica descrittiva univariata; ulteriori strumenti matematici o statistici verranno forniti o ripresi durante il corso.

Programma:

Argomenti trattati nel primo modulo (prof.ssa S.Terzi):

Le prime lezioni verranno dedicate al riepilogo di concetti base della statistica (quali varianza, covarianza, correlazione, combinazioni lineari) e all'algebra delle matrici.

Si passerà poi allo studio di possibili sintesi della matrice dei dati: analisi delle componenti principali; cluster analysis.

Infine dopo un riepilogo sulla regressione semplice (cioè la retta) si passerà allo studio della regressione multipla.

Si porrà molta enfasi sui casi di studio.

Argomenti trattati nel secondo modulo (prof.ssa F. Fortuna):

Introduzione ai comandi base di R e dell'interfaccia R-studio;

analisi delle componenti principali, cluster analysis e regressione semplice e multipla in R. Discussione di casi di studio.

Svolgimento:

Il corso prevede lezioni frontali in aula e in laboratorio nonché discussione di casi di studio.

Valutazione

La valutazione avviene a seguito di una prova pratica (elaborazione di un dat set) e di un colloquio orale.

Di norma la prova pratica si svolge in laboratorio.

Per gli studenti frequentanti e' prevista la possibilità di sostenere l'esame mediante due prove di esonero da superare entro la fine del corso consistenti nell'elaborazione di una tesina e in un colloquio orale.

Testi:

Tecniche e modelli di analisi multivariata, G. Di Franco, Franco Angeli 2017

<https://cran.r-project.org/doc/contrib/DellOmodarme-esercitazioni-R.pdf>

verranno inoltre fornite dispense e slides a cura dei docenti

2. Statistica per il management – Paola Vicard

(9 cfu - Secondo semestre)

Obiettivi formativi

Il corso intende fornire gli strumenti necessari per la progettazione, l'elaborazione e l'analisi dei dati in campo manageriale. Il corso introduce lo studente alle metodologie statistiche per le ricerche di mercato, per l'analisi della soddisfazione e per le decisioni aziendali. La parte teorica sarà affiancata da una parte applicata su dati reali e case studies provenienti dal marketing e dal management mediante l'uso di opportuni software statistici (quali Minitab, SPSS).

Agli studenti verrà insegnato non solo l'aspetto teorico dei modelli ma anche i principali contesti applicativi e il loro utilizzo mediante opportuni software statistici.

Il corso di Statistica per il management, insegna a gestire un'indagine statistica dalla sua pianificazione all'analisi e commento dei dati.

Prerequisiti:

Gli studenti devono possedere le conoscenze di statistica di base (esame del triennio), ossia statistica descrittiva, elementi di calcolo delle probabilità e elementi di inferenza statistica.

3. Statistical methods in economics – Caterina Conigliani

(9 cfu - Primo semestre – lingua inglese)

Course learning objectives

The main objective of the course is to provide the fundamental tools for the application of statistical methods to the analysis of economic data. The theoretical part will be supported by an applied part devoted to the analysis of real data sets by means of the software R. A student that has completed the course should be practiced in the application of advanced statistical methods, should be able to interpret the results of a statistical analysis, and should be aware of limitations and possible sources of errors in the analysis.

Detailed course schedule

Part 1: Introduction to data analysis and exploratory techniques

- Cluster analysis
- Principal component analysis

Part 2: Normal linear regression and its generalizations

- Polynomial regression
- Multiple regression
- Logistic regression
- Beta regression
- Poisson and negative binomial regression
- Spatial regression models

Part 3: Panel data analysis

- Balanced and unbalanced panel, micro and macro panel
- Modeling the level of a dependent variable
- Modeling the change of a dependent variable
- Fixed effects and random effects models for categorical variables and continuous variables
- Spatial regression models for panel data

Text books

- Chatterjee, S. and Hadi, A.S. (2012), Regression Analysis by Example, 5th Edition, Wiley. Chapters: 1, 2, 3 (excluding 3.9), 4 (excluding 4.3, 4.9.2, 4.9.3, 4.10, 4.12, 4.13, 4.14), 5 (excluding 5.6 and 5.7), 6 (excluding 6.6 and 6.7), 9, 11, 12 (excluding 12.8.3 and 12.8.4), 13 (excluding 13.5, 13.6, 13.7).
- Fox, J. and Weisberg, S. (2010), An R companion to applied regression, 2nd Edition, SAGE publications Inc.
- Andreb, H-J, Golsch, K., Schmidt, A.W. (2013), Applied panel data analysis for economic and social surveys, Springer. Chapters: 1, 2, 3, 4.

Final assessment

Attending students will develop and discuss a short dissertation (January/February 2023 only). For non-attending students the course assessment will be based on a written exam held in the computer lab, that will involve the analysis of different data sets using the methods and models studied during the course.

4. Metodi statistici per l'econometria e la finanza - Alessia Naccarato

(9 cfu - Secondo semestre)

Obiettivi formativi:

L'obiettivo di introdurre alle principali tecniche dell'econometria, il cui uso è ormai divenuto pratica corrente nel lavoro empirico in molti ambiti dell'analisi economica, finanziaria e aziendale. l'attenzione è focalizzata sull'intuizione che sta alla base dei diversi approcci e sulla loro rilevanza pratica. nel corso vengono introdotti e discussi esempi e applicazioni empiriche tratte da ambiti di analisi come l'economia del lavoro, la finanza, l'economia internazionale, l'economia ambientale, la macroeconomia, il management. l'utilizzo delle diverse procedure viene illustrato attraverso esempi di natura pratica basati sull'uso di dati tratti da casi reali, con l'impiego di un idoneo software (e-views, r).

Programma

Richiami di inferenza e algebra lineare. Modello di regressione lineare classico: ipotesi di base del modello, stima dei minimi quadrati, stima di massima verosimiglianza, test sui parametri del modello, linearità, eteroschedasticità, autocorrelazione, multicollinearità, regressori endogeni e stimatori alle variabili strumentali, previsione lineare, errori di specificazione, stabilità della funzione di regressione.

Modelli per dati panel ad effetti fissi e casuali. Analisi delle serie storiche: aspetti descrittivi, modelli AR, MA, ARMA, modelli a ritardi distribuiti.

Testi Adottati

Introduzione all'econometria

James H. Stock - Mark W. Watso

Ed. Pearson

Econometria Marno Verbeek

Ed. Zanichelli

Modalità Erogazione

Lezioni frontali in aula

Modalità Valutazione

Colloquio orale sugli argomenti del corso

5. Advanced topics in Statistical learning – Marilena Barbieri

(9 cfu - primo semestre – lingua inglese)

Obiettivi formativi

The aim of the course is to provide students with a coverage of a set of methods used in the analysis of economic data to answer a variety of specific, quantitative questions and with the computational tools to be used in the empirical applications. The program includes further topics on the classical regression model, time series analysis and some recent proposal to deal with applications having many observations and/or many predictors relative to the number of observations.

The course applies the widely used freeware programming environment for statistical analysis, known as R (through the RStudio interface).

Programma

A refresher on multiple linear regression. Further topics in regression analysis: Instrumental variables, Differences-in-Differences, Regression Discontinuity estimators Heckman's sample selection model, Mediation analysis, Quantile regression. Introduction to time series and forecasting. Models for univariate time series: ARMA, ARIMA, ARCH, GARC. Time series regression with additional predictors. The autoregressive distributed lag model. Estimation of dynamic causal effects. The VAR model. Breaks. Cointegration and error correction. Prediction with many regressors and big data: Ridge regression, the Lasso, regression trees and random forests.

Lingua di svolgimento

inglese

Testi adottati

Stock J.H. and Watson M.W., Introduction to Econometrics, Global edition, 4th Edition, 2020, Pearson.

Angrist, J.D. ; Pischke, J.-S., Mostly harmless econometrics : an empiricist's companion, 2009, Princeton University Press.

Modalità di erogazione e di valutazione

Whole class teaching.

Attendance in classes is recommended, but not compulsory.

The final exam is composed of a written two hours closed-book and closed-note test, takes place in the Lab and consists in both analytical and computer-exercise questions and an oral exam. The oral exam is not compulsory and consists mainly in a discussion on the written test.

Each candidate may prepare a short essay on the analysis of a set of data, using the tools discussed during the classes, whose content will be argue during the oral exam.

6. Statistical learning – Francesca Fortuna – 6 cfu 40 ore

(6 cfu – Secondo semestre)

Obiettivi formativi:

Il corso si propone di introdurre metodi e modelli per estrarre informazioni rilevanti da grandi moli di dati, con particolare attenzione all'apprendimento statistico (statistical learning) in contesti di classificazione supervisionata (supervised classification). Al fine di fornire le competenze per l'analisi e la modellazione di dati reali, saranno trattati problemi applicativi attraverso l'uso di specifiche librerie del software statistico R.

Programma

Il corso affronterà le seguenti tematiche:

- Introduzione ai principali modelli di statistical learning (apprendimento statistico);
- problemi di previsione e classificazione: richiami su regressione lineare e sui principali metodi di classificazione non supervisionata;
- Supervised classification: K-Nearest-Neighbours;
- Errore di misclassificazione;
- Metodi di ricampionamento: cross validation e bootstrap;
- Metodi basati su alberi decisionali: regression trees, classification trees, bagging, random forests, boosting.
- Introduzione ai metodi di classificazione Semi-supervisionata;
- Uso dell'ambiente statistico R

L'insegnamento è strutturato in 40 ore di didattica frontale, con lezioni teoriche e pratiche (con utilizzo del software RStudio).

7. Statistica ambientale – Francesco Lagona

(6 cfu – Secondo semestre)

Obiettivi formativi

Essere in grado di scegliere il più appropriato modello statistico per l'analisi di fenomeni ambientali.

Familiarizzare con l'ambiente statistico R per la stima del modello e la valutazione della sua bontà di adattamento.

Essere in grado di comunicare in modo efficace i risultati ottenuti.

Programma

R-studio ed R: installazione e principali caratteristiche. La creazione di un project. Primi passi con R: operazioni elementari, statistiche descrittive, grafici. Richiami di inferenza statistica. Distribuzioni campionarie, intervalli di confidenza e test. Applicazioni ed esempi in R. Stima di massima verosimiglianza. Modelli lineari: stima dei parametri, selezione del modello e valutazione della bontà di adattamento. Modelli lineari generalizzati: stima dei

parametri, selezione del modello e valutazione della bontà di adattamento. Modello di regressione logistica. Modello di regressione di Poisson.

Modalità di valutazione

Elaborazione di una tesina e discussione.

Testo

Carsten Dormann (2020) Environmental Data Analysis: An introduction with Examples in R, Springer.

8. Big data and machine learning – Domenico Cucina

(6 cfu – Secondo semestre)

Obiettivi Formativi

Il corso intende fornire allo studente le conoscenze metodologiche e applicative necessarie per risolvere problemi relativi all'apprendimento automatico e analizzare insiemi di dati potenzialmente grandi.

Lo studente acquisisce capacità teoriche e pratiche che gli consentono di utilizzare e sviluppare strumenti di machine-learning per analizzare dati di grandi dimensioni.

Programma

Caratteristiche dei big data - Modelli di programmazione per i big data: Hadoop MapReduce e Apache Spark - Algoritmi di machine learning – Apache Spark con R: sparklyr, dplyr, ggplot2.

Modalità di valutazione

Elaborazione di una tesina e discussione.

Testo

Jared Dean. Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, 2014, Wiley.

9. Metodi statistici di analisi dei dati - Giuseppe Bove

(6 cfu – Secondo semestre)

Prerequisiti

Conoscenze statistiche di base

Programma

Regressione lineare semplice. Intensità dell'effetto e potere predittivo della retta di regressione. Casi influenti. Variabili indipendenti qualitative. Limiti dell'analisi bivariata e potenzialità dell'analisi multivariata. Cenni alla regressione multipla. Campionamento casuale. Errore di campionamento e intervallo di confidenza. Dimensione campionaria. Campionamento probabilistico e non probabilistico.

Modalità di svolgimento

Corso in presenza

Modalità di valutazione

E' prevista una unica prova finale scritta di un'ora, con 20 quesiti di tipo misto (domande aperte con brevi risposte; domande chiuse), riguardanti il complesso del programma.

Testi adottati e bibliografia di riferimento

CORBETTA P., GASPERONI G., PISATI M., STATISTICA PER LA RICERCA SOCIALE, IL MULINO, BOLOGNA, 2001.

Capitoli e paragrafi da studiare sul testo di Corbetta, Gasperoni e Pisati

Capitolo 6: tutto (escluso regressione lineare piece-wise a pag. 171) capitolo 7: tutto (escluso definizione e utilizzo dei coefficienti dxy ed escluso paragrafo 5.) Capitolo 8: tutto (escluso paragrafo 2.3) capitolo 10 tutto (nei paragrafi 3.1 e 3.2 escluse le formule che non riguardano la stima della media).

10. Statistica per la valutazione – Rosa Capobianco

(6 cfu – Primo semestre)

Prerequisiti

Nessun prerequisito richiesto

Obiettivi Formativi

Il corso presenta allo studente alcuni dei concetti di base della statistica descrittiva e inferenziale. Particolare attenzione è rivolta ai test statistici e all'analisi della varianza. La finalità del corso è dare agli studenti gli strumenti per effettuare l'analisi dei dati.

Programma

- Richiami di statistica descrittiva.
- Campionamento e misurazione: definizione di popolazione e campione. Campionamenti probabilistici: campionamento casuale semplice con e senza reinserimento. Campionamento stratificato. Campionamento a grappoli. Campionamento non probabilistico.
- Cenni alla teoria della probabilità. Variabili casuali discrete e continue. Distribuzione bernoulliana e distribuzione normale.
- Inferenza statistica: la stima. Stima puntuale e stima intervallare. Intervallo di confidenza per una proporzione. Intervallo di confidenza per la media. Scelta della dimensione campionaria.
- Inferenza statistica: test di significatività. Definizione del sistema d'ipotesi e del test di significatività. Gli errori di prima e di seconda specie. Test di significatività per la media. Test di significatività per la proporzione. Calcolo del p-value.
- Confronto tra due gruppi: Confronto tra due proporzioni. Confronto tra due medie. Confronto per dati appaiati.
- Analisi della varianza
- Tabelle di contingenza e test chi-quadrato di indipendenza.

Modalità di valutazione

E' prevista una unica prova finale scritta di due ore, con 10 quesiti a risposta chiusa e 4 esercizi, riguardanti il complesso del programma.

Testi adottati e bibliografia di riferimento

Agresti, A., Finlay B. (2012). Metodi Statistici di Base e Avanzati per le Scienze Sociali, Pearson, Milano. (Capp. 1-9)

Numero minimo e massimo di iscritti ammissibili

numero minimo: 5

numero massimo: 50

Requisiti di ammissione

Aver acquisito nella propria carriera universitaria almeno 9 crediti in Statistica (Secs-S/01).

Criteri di selezione dei partecipanti

Laddove sia superato il numero massimo di domande, la selezione avverrà sulla base del numero di esami sostenuti e della relativa votazione (per quanto riguarda la selezione di studenti iscritti a corsi di studio dell'Ateneo).

Ai non studenti potrà essere riservata una quota del 20% del numero di immatricolati. Saranno selezionati in base al voto di laurea o laurea magistrale, e a parità di voto si darà la precedenza al candidato più giovane.

Contributi di iscrizione

Gli studenti regolarmente iscritti a un corso di laurea o di laurea magistrale dell'Ateneo, anche in qualità di studenti in mobilità internazionale in ingresso, possono iscriversi gratuitamente al corso minor per il medesimo anno accademico, fatto salvo il pagamento dell'imposta di bollo.

Coloro che non siano contemporaneamente iscritti a un corso di laurea o di laurea magistrale dell'Ateneo nel medesimo anno accademico sono tenuti al pagamento del contributo di iscrizione dell'importo di euro 480,00, oltre imposta di bollo.

Coloro i quali si trovino in condizioni di disabilità, con riconoscimento di handicap ai sensi dell'articolo 3, commi 1 e 3, della legge 5 febbraio 1992, n. 104, o con un'invalidità pari o superiore al 66%, sono esonerati dal pagamento dei contributi di iscrizione al corso e versano esclusivamente l'imposta di bollo.

Trattandosi di corso interdipartimentale, i proventi derivanti dai contributi di iscrizione saranno ripartiti nella misura del 40% al Dipartimento di Economia (dipartimento di riferimento per questo corso), del 30% al Dipartimento di Scienze della Formazione e del 30% al Dipartimento di Scienze Politiche.

Eventuali agevolazioni economiche

Nessuna

Prova finale

Non è prevista una prova finale