

Capitolo 1

Introduzione

Presentazione del corso

Il corso di laurea triennale in Statistica e Gestione delle Informazioni (nel seguito SGI) prepara all'uso di metodi statistici fortemente integrati con gli strumenti informatici; insegna a gestire e trattare grandi quantità di dati per interpretare fenomeni demografico-sociali, biostatistici o relativi agli aspetti gestionali di diverse realtà.

Profili professionali e sbocchi occupazionali

Il laureato in Statistica e Gestione delle Informazioni è in grado di esercitare funzioni ed attività coerentemente con gli obiettivi formativi ed i risultati di apprendimento attesi corrispondenti ai seguenti sbocchi occupazionali e professionali:

- pianificazione e gestione del territorio e delle dinamiche della popolazione
- ricerca sperimentale ed osservazionale
- proiezioni elettorali e sondaggi d'opinione
- gestione dei sistemi informativi e delle basi di dati
- consulenza statistica
- data mining

La laurea in Statistica e Gestione delle Informazioni consente l'inserimento nei seguenti settori economici: Terziario avanzato, Istituti di ricerca (biologici, ambientali, clinici, epidemiologici, demo-sociali, etc.), Aziende farmaceutiche, ospedaliere e sanitarie locali, Industria, Pubblica Amministrazione.

Qualche dato sul corso di laurea

Qualche numero relativo al Corso di Laurea in SGI. È nato nel 2005 ed ha avuto un numero di immatricolazioni pressoché costante negli ultimi anni a partire dal 2008 come riportato in tabella 1.

a.a.	2010	2011	2012	2013	2014	2015	2016	2017
immatricolati	34	40	60	44	50	87	134	123

Tabella 1.1: Immatricolazioni per anno accademico

Fonte: *Cruscotto della didattica dell'Università degli Studi di Milano-Bicocca Estrazione del 17 settembre 2018*

Negli ultimi anni, gli studenti intervistati in merito alla soddisfazione per lo svolgimento dei corsi di SGI hanno evidenziato un buon livello di soddisfazione. In tabella 1.2 sono riportate il valo-

re medio dei giudizi riportati per la domanda: "Sono complessivamente soddisfatto di come è stato svolto questo insegnamento?"

Anno accademico	2013	2014	2015	2016	2017
Giudizio medio*	2.10	2.07	2.11	2.36	2.07

Tabella 1.2: Giudizio medio sulla soddisfazione per anno accademico.

*Si ricorda che la scala della risposta è 0 - 3 dove valori elevati indicano soddisfazione

Fonte: *Cruscotto della didattica dell'Università degli Studi di Milano-Bicocca*

Estrazione del 17 settembre 2018

Gli abbandoni al primo anno sono riportati nella tabella 1.3. Il valore è variato tra il 16% del 2013 e il 29.4% del 2010. Il valore medio di tutte le triennali dell'Università di Milano-Bicocca è pari a 19.9%.

La percentuale di laureati in corso si trova nella tabella 1.4. In ogni anno accademico la percentuale di laureati in corso è superiore alla percentuale media di laureati di tutte le triennali dell'Ateneo (27.8%)

Statistiche e Lavoro

Come si inseriscono professionalmente gli Statistici dopo aver conseguito la laurea? Quale tipo di lavoro svolgono? Quanto tempo trascorre dal conseguimento del titolo al primo lavoro?

Anno accademico	2010	2011	2012	2013	2014	2015	2016
% di abbandoni	29.4	22.5	25.0	15.9	18.0	28.7	38.3

Tabella 1.3: Percentuali di abbandoni al primo anno per anno accademico

Fonte: *Cruscotto della didattica dell'Università degli Studi di Milano-Bicocca*

Estrazione del 17 settembre 2018

n.d.=non disponibile

Anno accademico	2010	2011	2012	2013	2014	2015	2016
% di laureati in corso	41.2	47.5	33.3	45.5	42.0.	n.d.	n.d.

Tabella 1.4: Percentuali di laureati in corso per anno accademico

Fonte: *Cruscotto della didattica dell'Università degli Studi di Milano-Bicocca*

Estrazione del 17 settembre 2018

n.d.=non ancora disponibile

Per rispondere a queste ed altre domande, si interrogano periodicamente i laureati con l'indagine "Statistiche e lavoro" che ha lo scopo di contestualizzare al meglio la figura dello Statistico nel mondo del lavoro.

La situazione dei laureati in SGI ad un anno dal conseguimento del titolo è decisamente buona come riportato in tabella 1.5: circa un terzo degli laureati in SGI, a un anno dalla laurea, è assorbito

velocemente dal mondo del lavoro e nessuno è in cerca di lavoro.

Percentuale degli intervistati	
Lavora e non è iscritto alla magistrale	12.5
Lavora ed è iscritto alla magistrale	25.0
Non lavora ed è iscritto alla magistrale	62.5
Non lavora, non è iscritto alla magistrale e non cerca	-
Non lavora, non è iscritto alla magistrale ma cerca	-

Tabella 1.5: Condizione lavorativa degli studenti che si sono laureati nel 2016

Fonte: <https://www2.almalaurea.it/>. Indagine “Condizione occupazionale dei laureati”. Anno di indagine: 2017

Il ruolo dello statistico medico

”Cos’è la statistica medica? Si tratta di individuare i metodi statistici più adeguati in un contesto particolare come quello della medicina.

Il clinico svolge fondamentalmente due attività: curare i propri pazienti con gli interventi terapeutici farmacologici più adeguati per la malattia e per il paziente e tentare di prevenire l’insorgenza della malattia, rimuovendo nei propri assistiti tutti i fattori di rischio.

Un medico, per poter curare il proprio paziente, dovrebbe riconoscere la malattia e saper individuare, tra tutte le tecniche disponibili, quella più adeguata. Per poter fare questo, qualcuno de-

ve aver studiato quali sono gli interventi terapeutici farmacologici più adeguati. Dietro ogni farmaco, che vedete o consumate, ci sono spesso decine di anni di sperimentazioni cliniche che, in modo rigoroso e più oggettivo possibile, studiano, indagano e forniscono prove sull'efficacia e sull'adeguatezza dei farmaci per la cura di una determinata malattia. Ecco è in questo contesto che interviene lo statistico che, insieme ad altri professionisti, pianifica lo studio.

Spostiamoci adesso all'altra attività del medico quella della prevenzione. Per poter rimuovere i fattori di rischio bisogna essere in grado di riconoscerli. Anche in questo caso lo statistico interviene, insieme ad altri professionisti, nel pianificare lo studio sui fattori di rischio ed analizzare i dati.”¹

Il ruolo del demografo

Cos'è la demografia? In che cosa consiste? La demografia fornisce gli strumenti attraverso i quali rispondere a domande fondamentali relative alla popolazione. Quali ad esempio: quanti sono gli abitanti? Quale è la densità degli abitanti sul territorio? Come mai una popolazione si comporta in un certo modo rispetto alla sua dimensione quantitativa. La demografia si occupa della misurazione di questi fenomeni, non solo contando i nati e morti, ma cercando anche di capire cosa c'è dietro. Come mai le nascite, in un paese come l'Italia, sono passate da un milione di 30 anni fa a cinquecentomila oggi? Cosa guida il comportamento italiano in questa direzione? Come mai cent'anni fa un neonato aveva davanti a sé una

¹Tratto dal video di presentazione del Prof. Corrao. Il video è consultabile al seguente <http://www.youtube.com/watch?v=AMOVa8Wka78>

speranza di vita di 30 anni, mentre oggi la speranza di vita è più di 80 anni? Come è successo? Cosa ha determinato questa evoluzione? Queste sono le domande a cui il demografo ha la presunzione di rispondere. Il massimo dell'aspirazione demografica è immaginare il futuro. Non solo contare, spiegare gli andamenti del passato e del presente, ma servirsi di questi elementi per anticipare il futuro. Quindi raccontare il passato, interpretare il passato, esporre il presente anche alla luce del passato, ma, soprattutto, anticipare il futuro sono le tre dimensioni temporali che caratterizzano il lavoro del demografo. Non dimentichiamo che, come diceva un grande demografo francese, ciascuna popolazione ha iscritto il suo futuro dentro di sé, è importante quindi saperlo leggere, interpretare e magari anche raccontare a chi in prospettiva dovrà prendere decisioni in termini di programmazione di interventi e, più in generale, di politica.²

Il ruolo dello statistico

Spesso si considera la Statistica come una materia esclusivamente tecnica e poco creativa. Ma è davvero così? In realtà rappresenta una sfida affascinante finalizzata a conoscere meglio alcuni aspetti della realtà che ci circonda, altrimenti difficilmente approcciabili, come i fenomeni di massa, i mercati, le azioni politiche.

L'indagine richiede cultura generale e grande immaginazione per analizzare i fenomeni tenendo conto di tutti i fattori in gioco e

²Tratto dal video di presentazione del Prof. Blangiardo. Il video è consultabile al seguente <http://www.youtube.com/watch?v=EN8K9fA8Dkc>

dandogli il giusto peso. "La Statistica è una scienza relativamente giovane che in tre secoli ha avuto un enorme sviluppo. Il suo nome deriva da "Stato": statistici erano coloro che discutevano delle cose relative al governo dello Stato sulla base di "numeri, pesi e misure ...". Compito degli statistici del XVII e del XVIII secolo era dunque quello di raccogliere metodicamente i dati relativi a popolazioni numerose, alle loro ricchezze, ai loro comportamenti, per sintetizzarli in informazioni utili al fine di conoscere e, quindi, di assumere decisioni razionali Gli statistici, partendo da queste tecniche antichissime e avvalendosi dei progressi della matematica affinarono via via le procedure volte a migliorare le informazioni numeriche sulle popolazioni e sui diversi fenomeni che le riguardavano, in vista delle decisioni, del controllo o del confronto con altri Stati. Le sintesi da essi operate, anche se comportavano sempre una perdita di informazioni rispetto alla ricchezza dei dati individuali, si dimostrarono strumenti potenti per la conoscenza e il governo delle popolazioni. Per questo motivo nei secoli XIX e XX l'impiego delle procedure statistiche si estese progressivamente allo studio di altri insiemi numerosi di elementi empirici dei quali poteva essere utile sintetizzare caratteristiche e comportamenti ai quali si dette il nome di popolazioni o di collettivi: da quello delle molecole che compongono un gas a quello delle particelle elementari della materia; da quello delle misure ripetute di grandezze fisiche o chimiche a quello dei risultati di numerosi esperimenti in campo medico o farmaceutico; da quello dei prodotti realizzati in serie a quello dei potenziali clienti delle imprese o elettori delle democrazie; da quello dei prezzi dei beni di consumo a quello dei lavoratori

occupati e disoccupati e così via”.³

Prima di iscriversi

Prepararsi agli studi universitari significa valutare le proprie capacità e mettersi nelle condizioni per ottenere i migliori risultati, senza dimenticare che si studia bene solo ciò di cui si capisce l'utilità e la bellezza.

Un corso scientifico come SGI dà ottime possibilità di lavoro, ma richiede impegno e capacità. La maggior parte degli studenti che iniziano i corsi universitari scientifici non incontra gravi ostacoli negli esami di matematica. Tuttavia per altri la matematica è fonte di difficoltà e fatica a nostro avviso riconducibili a due carenze, che è utile distinguere.

Contenuti: Negli insegnamenti universitari non si assumono conoscenze di Analisi Matematica, mentre si considerano acquisite l'Algebra elementare (disequazioni di II grado e irrazionali, esponenziali e logaritmi), la Geometria Euclidea elementare, la Geometria Analitica elementare, la Trigonometria (funzioni goniometriche, equazioni e disequazioni) e le proprietà elementari dei numeri interi e dei numeri razionali.

Metodo: La fatica incontrata da molti studenti nello studio della matematica è però dovuta anche al particolare metodo di studio che la matematica richiede: bisogna comprendere la necessità di un linguaggio preciso e saperlo utilizzare quan-

³Tratto da "La Statistica" del Prof. Marco Martini per l'articolo completo vedi: <http://sgi.dismeq.unimib.it/> sezione Orientamento

do necessario, familiarizzarsi con gli strumenti logici ed il simbolismo matematico, comprendere le giustificazioni dei risultati (esercizi, teoremi, tecniche di calcolo) verificandone i passaggi, esemplificandole e riutilizzandole in situazioni analoghe.

Il Corso di Laurea in SGI propone, agli studenti interessati, di lavorare per tempo sulle difficoltà descritte sopra attraverso il progetto disponibile in forma di e-learning alla pagina

<https://open.elearning.unimib.it/> o sui seguenti testi:

- M. Bramanti, Precalculus, Progetto Leonardo, Società Editrice Esculapio (1999),
- M. Bramanti e G. Travaglini, Matematica. Questione di metodo, Zanichelli (2009),

o seguendo i consigli dei propri docenti di Matematica della Scuola Superiore.

Modalità di ammissione

A partire dall'anno accademico 2018-2019 per potersi immatricolare al corso di laurea è necessario sostenere il TOLC-E del CISIA (Consorzio Interuniversitario Sistemi Integrati per l'Accesso). Il test è diviso in sezioni: linguaggio matematico di base, logica, comprensione del testo, inglese ed è erogato in modalità informatica come prova unica a livello nazionale e può essere svolto presso qualunque sede universitaria nazionale accreditata da CISIA in più date durante l'anno. Per le modalità di iscrizione, le date definitive dei test e le modalità di assegnazione del punteggio ad ogni domanda si rimanda al sito del corso di laurea o su

<http://www.cisiaonline.it/area-tematica-tolceconomia/la-prova-line-2>

ATTENZIONE: Lo studente può immatricolarsi al corso di Laurea di SGI con qualsiasi punteggio del test. Tuttavia gli studenti che hanno ottenuto un punteggio complessivo inferiore a 15 (somma dei punteggi delle diverse sezioni tranne quella di inglese) non potranno sostenere altri esami prima di aver superato l'esame di Statistica I e Analisi matematica I del primo anno di corso.

1.0.1 Informazioni generali

Come raggiungere la sede del corso

La sede del corso è situata nell'edificio U7, in Via Bicocca degli Arcimboldi 8 a Milano. Gli spazi dell'edificio U7 sono aperti dal lunedì al venerdì, dalle ore 8.00 alle ore 20.00.

Il complesso universitario Bicocca, situato nella zona Nord di Milano, è facilmente raggiungibile:

<http://www.unimib.it/go/8210263877271840055/Home/Italiano/Menu/Dove-siamo/Come-raggiungerci>

I Rappresentanti degli studenti

Oltre ai docenti anche alcuni studenti iscritti ai corsi di laurea in statistica, eletti periodicamente, partecipano alle sedute del Consiglio di Coordinamento Didattico.

Docenti di riferimento per attività di orientamento

Dott.ssa Stefania Rimoldi

tel. +39 02.6448.5844

mail: stefania.rimoldi@unimib.it

Studiare all'estero

L'Unione Europea ha istituito diversi programmi riguardanti l'internazionalizzazione degli studi universitari.

1. Il programma Erasmus+ ai fini di studio permette di trascorrere un periodo di studio (fino a dodici mesi) presso un'Università straniera con un contributo da parte della Comunità Europea. Per l'anno accademico 2018-2019, il Corso di Laurea prevede lo scambio di studenti con le Università di seguito indicate:
 - **Spagna:**
 - Universitat de Barcelona (Corso di Laurea a ciclo unico in Statistica)
 - Universidad Miguel Hernandez de Elche (Corso di Laurea in Statistica per l'impresa)
 - Universidad Carlos III de Madrid (Corso di Laurea in Statistica per l'impresa)
 - **Norvegia:**
 - Universitetet I Oslo (Facoltà di Matematica e Scienze Naturali, Facoltà di Medicina)
 - **Svezia:**

- Stockholms Universitet (Dipartimento di Matematica, Divisione di Statistica Matematica, Master in Statistica Matematica)

Nel bando saranno indicati i requisiti richiesti e tutte le informazioni per presentare la candidatura.

Non è possibile sostenere all'estero i seguenti esami: Analisi Matematica I, Analisi Matematica II, Algebra Lineare, Statistica I, Calcolo delle Probabilità, Statistica II, Analisi Statistica Multivariata.

2. Il programma Erasmus+ ai fini di Traineeship permette agli studenti di svolgere stage e/o tesi all'estero in qualsiasi sede consorziata in base ad accordi/contatti personali di docenti del dipartimento con colleghi/università/centri di ricerca/aziende all'estero. Lo studente riceve un contributo comunitario per un periodo di tirocinio che va da un minimo di 3 a un massimo di 6 mesi.

Docenti di riferimento per Erasmus e internazionalizzazione:

Prof. Rino Bellocco

tel. +39 02.6448.5831

mail: rino.bellocco@unimib.it

Dott.ssa Simona C. Minotti

tel. +39 02.6448.5836

mail: simona.minotti@unimib.it

Sito del corso di laurea

Tutte le informazioni contenute in questa Guida, i Regolamenti didattici dei corsi, i servizi, le pagine personali dei docenti, sono disponibili nel sito web del corso di laurea, accessibile dall'indirizzo <http://sgi.dismeq.unimib.it>.

Pagina Facebook:

<https://www.facebook.com/SGI.UniversitaBicocca>

Iscrizioni

Le modalità e i termini di scadenza delle immatricolazioni sono indicati nel sito di Ateneo www.unimib.it, sezione Segreteria Studenti.

Trasferimenti, seconde lauree, riconoscimento dei crediti

Ai fini del trasferimento da un altro Corso di Laurea o dell'iscrizione a seconda laurea, è possibile richiedere al Comitato di Coordinamento del Corso di Laurea la ricostruzione della carriera pregressa. Il Comitato provvederà a calcolare il valore in CFU delle attività formative da riconoscere e le relative tipologie, determinando l'anno di corso al quale gli studenti saranno iscritti. Più precisamente possono essere iscritti al secondo anno solo coloro ai quali siano state riconosciute attività formative per almeno 30 CFU, al terzo anno coloro ai quali siano state riconosciute attività formative per almeno 60 CFU. Il Comitato di Coordinamento didattico concorderà con gli studenti un piano di studi che specifichi tutte le attività formative residue necessarie per il conseguimento

mento della laurea. I termini e le modalità di presentazione delle domande sono indicati nel sito ufficiale dell'Ateneo.

Docente di riferimento per ricostruzioni carriere:

Prof. Piergiorgio Lovaglio

tel. +39 02 6448.3217

mail: piergiorgio.lovaglio@unimib.it

Per altre informazioni: Sportello telefonico - Servizio orientamento di Ateneo

Sportello Telefonico

tel. 02.6448.6448

mail:

dal lunedì al venerdì dalle 10.00 alle 12.00 e dalle 14.00 alle 16.00.

Richieste via email: sono disponibili indirizzi email per diverse tematiche

- informazioni sulle immatricolazioni e iscrizioni, procedure e scadenze, servizi e le opportunità: mail: orientamento@unimib.it
- informazioni sul tirocinio formativo attivo: mail: tfa@unimib.it
- informazioni sugli stage: mail: stage@unimib.it
- informazioni sul job placement: mail: vulcano@unimib.it
- informazioni su esami e prove di accertamento di lingua straniera: mail: segreteria.lingue@unimib.it
- informazioni su esami e prove di accertamento di informatica: mail: informatica.ateneo@unimib.it

Front office: Edificio U17, Piazzetta Difesa per le Donne (adiacente a via Padre Beccaro).

Organizzazione del corso di laurea

Elenco delle attività secondo il regolamento del Corso di Laurea in Statistica e Gestione delle Informazioni anno 2018-2019 sono elencate di seguito

Insegnamento Moduli	CFU	SSD
PRIMO ANNO		
Algebra lineare	6	MAT/02
Analisi Matematica I	9	MAT/05
Calcolo delle probabilità	9	SECS-S/01
Demografia	9	SECS-S/04
Informatica	9	ING-INF/05
Statistica I	6	SECS-S/01
Statistica Medica	6	MED/01
Laboratorio di informatica	3	
Lingua straniera	3	
TOTALE PRIMO ANNO	60	
SECONDO ANNO		
Analisi dei dati	6	SECS-S/01
Analisi matematica II	6	MAT/05
Analisi statistica multivariata		
Analisi esplorativa	7	SECS-S/01
Modelli statistici	8	SECS-S/01
Basi di dati	6	INF/01
Statistica I - Complementi	6	SECS-S/01
Statistica II	12	SECS-S/01
Statistica Sociale	9	SECS-S/04
TOTALE SECONDO ANNO	60	
TERZO ANNO		
Data mining e Statistica computazionale		
Data mining	9	SECS-S/01
Statistica computazionale	6	SECS-S/01
Sistemi informativi	9	ING-INF/05
Attività formative affini o integrative	18	
Insegnamenti a scelta	12	
TOTALE TERZO ANNO	60	

Il credito formativo universitario (CFU) è una modalità utilizzata nelle università italiane per misurare il carico di lavoro richiesto allo studente. E' stato calcolato che uno studente può dedicare ogni anno 1500 ore del proprio tempo allo studio (studio individuale, lezioni, laboratori, stage). Convenzionalmente 1 CFU rappresenta 25 ore di lavoro (come studio personale o come frequenza a laboratori o lezioni). Per conseguire la laurea occorrono 180 CFU; per la laurea magistrale 120.

I settori scientifico-disciplinari sono raggruppamenti di discipline, stabiliti dal Ministero dell'Istruzione, dell'Università e della Ricerca con il DM 4 ottobre 2000 e poi rivisitati successivamente. Ad esempio MAT/05 indica Analisi Matematica, e SECS-S/01 indica Statistica. L'obiettivo dei settori è garantire i contenuti degli insegnamenti e la competenza specifica dei docenti.

Attività affini o integrative

Gli studenti possono acquisire tali crediti scegliendo gli insegnamenti tra quelli proposti. Tali insegnamenti sono riconducibili a tre aree tematiche. Almeno due insegnamenti devono appartenere alla stessa area.

Area Statistica. Settore SECS-S/01:

- Piano degli esperimenti (6 CFU)
- Data science e Modelli statistici per il trattamento dei dati non strutturati (6 CFU)
- Statistica spaziale ed ambientali (6 CFU)
- Complex Data Analysis (6 CFU)

Area Biostatistica. Settore MED/01:

- Elementi di biostatistica (6 CFU)
- Epidemiologia (6 CFU)
- Modelli lineari generalizzati in epidemiologia e medicina (6 CFU)

Area Demografica. Settore SECS-S/04:

- Demografia sociale (Paesi in via di sviluppo) (6 CFU)
- Demografia sociale (Mobilità e migrazioni) (6 CFU)
- Popolazione, territorio e società I (6 CFU)

L'area statistica fornisce una solida preparazione relativa ai metodi e ai modelli statistici che, unitamente a conoscenze sulle più moderne tecniche computazionali, viene impiegata per trattare e interpretare dati provenienti da svariati ambiti con particolare attenzione a quello ambientale.

L'area biostatistica si focalizza sulla pianificazione, gestione, analisi e interpretazione statistica di studi osservazionali e sperimentali nel contesto della ricerca biomedica. Fornisce approfondimenti sulle tecniche statistiche e computazionali che trovano prevalente applicazione nello studio dei fattori che condizionano la salute dell'uomo (ambiente fisico e sociale, patrimonio genetico, trattamenti e interventi medici, etc.).

L'area demografica si propone di descrivere ed interpretare gli eventi ed i processi propri della realtà demografica e sociale, ovvero di pianificare e gestire le dinamiche della popolazione e di condurre ricerche interdisciplinari sulla popolazione e sulle tematiche di contesto in ambito sociale.

Attività formative a scelta

Gli studenti possono acquisire i 12 CFU previsti per tali attività con una delle seguenti modalità:

- con il superamento della verifica di profitto relativa ad insegnamenti scelti fra quelli impartiti presso il Corso di laurea in SGI, in altre strutture dell'Università degli Studi di Milano-Bicocca o in altre Università convenzionate con l'Università degli Studi di Milano-Bicocca;
- con lo svolgimento dello stage/project work

Tra le attività formative, gli studenti potranno scegliere di frequentare una “Summer School” in matematica, statistica o informatica previa valutazione del programma annuale da parte del Coordinatore del proprio corso di laurea.

Stage/tirocini formativi

Tra le attività formative a scelta, lo studente può decidere di svolgere uno stage/project work, presso un'Azienda o un Ente convenzionato (stage esterno) o presso il Dipartimento di Statistica e Metodi Quantitativi (stage interno). Possono accedere allo stage/project work gli studenti iscritti al secondo o al terzo anno del corso di laurea che abbiano superato la metà degli esami previsti nel piano di studi. Il project work è limitato agli studenti lavoratori.

Per dare luogo all'attribuzione dei CFU previsti per tale attività, gli stage devono avere la durata minima di tre mesi. L'attribuzione dei CFU è subordinata ad un colloquio finale con il proprio tutor accademico e all'acquisizione da parte dell'Ufficio Stage del Corso di laurea in Statistica e Gestione delle Informazioni, del questiona-

rio di valutazione del tutor aziendale che sarà sottoposto anche al tutor accademico.

Per gli aspetti organizzativi si consiglia di rivolgersi all'Ufficio stage di Ateneo.

Di seguito i nomi degli enti, aziende o società coinvolti negli stage durante gli ultimi anni.

Elenco:

Ist. di ricerche farmacologiche Mario Negri, Menarini ricerche spa, Ist. clinici di perfezionamento, A.O. Niguarda Ca' Granda, Ist. Ortopedico Galeazzi, A.O. Osp. Riuniti di Bergamo, A.O. Osp. di Lecco, A.O. di Desio e Vimercate, Osp. pediatrico Bambin Gesù (Roma), Ist. nazionale per lo studio e la cura dei tumori, Ist. Neurologico Besta, Osp. San Raffaele, Asl Como, Asl Alessandria, Visup srl, AB research, Medtronic italia spa, Citroen italia spa, Consultraining di Bresso, Etro spa, Synergia, Ipsos srl, Patheon italia spa, Accenture spa, Camera di Commercio di Sondrio, TNS italia srl, Siae microelettronica spa, Anbsc agenzia nazionale dei beni sequestrati alla mafia, Nielsen srl, Simbologica srl, Micro semiconductor italia srl, Celio italia spa, Comune di Milano, Comune di Brescia, Comune di Piacenza, Comune di Legnano, Comune di Bergamo, Comune di Como, Comune di Vimercate, Comune di Arona, Comune di Bormio, Comune di Treviglio, Provincia di Bergamo, Dip. di Scienze Chirurgiche (unimib), Dip. di Sociologia (unimib), Opis srl, Ist. per le tecnologie della costruzione del CNR, Fondazione Banco Alimentare onlus, Parco regionale Spina Verde, Consorzio per le valutazioni biologiche e farmacologiche, Sesric (Ankara), Sas Institute, Dip. di Scienze Umane (unimib), CRISP (unimib).

Idoneità lingua straniera

La verifica della conoscenza della lingua straniera (3 CFU) è effettuata con le modalità previste dalla Commissione linguistica di Ateneo. Per tali modalità si rimanda al sito web di Ateneo, <https://www.unimib.it/didattica/lingue-unimib/idoneit%C3%A0-ateneo-e-accertamento-linguistico>. In conformità con la delibera del Senato Accademico del 3 luglio 2006, i CFU previsti per la lingua straniera devono essere acquisiti prima di sostenere gli esami del secondo e del terzo anno.

Commissione Lingue del Corso di Laurea in SGI:

Prof. Maria Elena Regonesi
tel. +39 02.6448.3437
mail: mariaelena.regonesi@unimib.it

Prof. Stefanie Karin Vogler
tel. +39 02.6448.3196
mail: stefanie.vogler@unimib.it

Abilità informatiche

Tali CFU possono essere acquisiti attraverso il superamento della prova di idoneità relativa all'insegnamento "Laboratorio di Informatica".

Prova finale

Per la prova finale sono possibili due alternative, ciascuna delle quali comporta l'acquisizione di 6 CFU. La scelta tra le due dipende dalla presenza o meno dello stage nel piano di studio.

È prevista la discussione di un elaborato, realizzato sotto la guida di un docente del Corso di laurea in Statistica e Gestione delle Informazioni, o di un altro docente di materie statistiche dell'ateneo.

Per gli studenti che abbiano effettuato uno stage è possibile, in alternativa, la discussione di una relazione scritta concernente l'esperienza di stage, predisposta con l'assistenza di un docente del Corso di laurea in Statistica e Gestione delle Informazioni, o di un altro docente di materie statistiche dell'ateneo.

In ogni caso la discussione ha luogo in seduta pubblica, di fronte ad una Commissione composta da professori e ricercatori del Corso di laurea in Statistica e Gestione delle Informazioni, o da altri docenti di materie statistiche dell'ateneo, che esprime la valutazione finale in centodecimi, con eventuale lode, tenendo conto sia dello svolgimento della prova finale sia dell'intera carriera universitaria dello studente. L'eventuale attribuzione della lode, in aggiunta al punteggio massimo di 110 punti, è subordinata alla valutazione unanime della Commissione.

Il calendario delle prove finali è disponibile sul sito del Corso di Laurea alla sezione Didattica, Esami e Lauree.

Elenco dei titoli relativo alle recenti prove finali:

- Analisi sui social media di compagnie aeree low cost
- Acquisizioni di cittadinanza e intenzioni migratorie. Analisi del caso lombardo(2013)
- Analisi del ruolo dell'angiogenina nella SLA
- Analisi della letteratura scientifica contenuta nell'archivio PubMed per mezzo di Graph Databases
- Analisi di dati biologici volta alla validazione di un farmaco bioequivalente
- Analisi e messa in qualità di un dataset usando algoritmi basati su word embeddings
- Bioinformatica e statistica per il trattamento di dati NGS in condizioni di campionamento limitato.
- Carte di controllo data-mining driven: teoria e applicazione.
- Confronto tra metodologie di selezione delle variabili in una coorte di pazienti obesi
- Creazione di dataset di training per la classificazione di offerte di lavoro testuali pubblicate sul web con tecniche di Machine Learning
- Cutting-Edge Deep Learning Neural Networks for advanced image classification: A comparative analysis between the Residual and the Capsules approaches
- Dinamiche demografiche e previsioni comunali: il caso di Boltiere (BG) dal 1982 al 2015
- I social media nelle strategie di marketing: SMA nel settore Coatings
- I movimenti tra luogo di residenza e struttura di ricovero dei cittadini lombardi: un'analisi basata sulla distanza
- Il p-value: da gold standard a "fuorilegge" della statistica

- Il fenomeno migratorio in Italia
- Integrazione delle carte di controllo tramite p-value
- L'approccio di sequence mining: un'applicazione in farmacoepidemiologia
- La competenza in materia di AIDS e la discriminazione nei confronti dei soggetti affetti dal virus.
- La relazione tra patient satisfaction e outcome negli ospedali lombardi
- Le cifre dei rifugiati e richiedenti asilo in Italia
- Le domande controfattuali, un caso di studio: cambiamenti nelle intenzioni di mobilità tra gli stranieri in Italia.
- Le intenzioni di mobilità degli immigrati residenti in Italia
- Liste civiche 2011-2016: Nuove evidenze. Analisi delle corrispondenze multiple dei risultati elettorali amministrativi
- Metodi di campionamento per il bilanciamento delle classi
- Modelli di regressione per l'analisi di risposte ordinali longitudinali. Un'applicazione ai dati di uno studio clinico in ambito oncologico.
- Rappresentazioni grafiche multidimensionali: il caso del Delirium Day
- Stimatori Bayesiani Empirici per il liscio di mappe
- Tecniche di classificazione dei tweets: confronto di algoritmi di clustering e sentiment analysis
- Tecniche per la correzione del selection bias: il caso della residenzialità alcolica

Le opinioni degli ex studenti

Alice Corbella

Sono Alice e ho iniziato nel 2009 il mio percorso universitario e statistico iscrivendomi al corso di laurea triennale in Statistica e Gestione delle Informazioni. Fin dai primi mesi di studio delle basi teoriche e pratiche della statistica ho apprezzato delle prime (la matematica, il calcolo delle probabilità, ...) la formalità che richiedevano: questo rendeva tutto molto più chiaro e ordinato; invece affrontando l'informatica e i primi accenni di trattamento dei dati ho iniziato ad accorgermi della potenza della statistica: la quantificazione di un fenomeno permette di leggere la realtà e trarne indicazioni utili per supportare decisione. Non nego che questo inizio è stato anche molto difficile perché ha coinciso con l'inizio dello studio universitario, tutto diverso da quello liceale, che richiedeva (fortunatamente) continuo confronto con compagni di corso e professori per poter imparare veramente e di più.

Proseguendo (specialmente durante il terzo anno) è stato per me interessante scoprire quanti metodi diversi esistano per approssimare lo stesso problema, e scoprire che serve il mio personale occhio critico per decidere quale tecnica usare. Un'altra parte che mi ha appassionato dell'applicazione dei modelli è quella che riguarda l'interpretazione dei risultati perché è richiesta sia l'interazione con chi conosce meglio di me l'ambito in cui vengono applicate le tecniche (siano medici o direttori d'azienda o ...) che la mia capacità di leggere i numeri. In questi aspetti, alla fine del

percorso, mi sono accorta di aver raggiunto una certa autonomia; questa è stata messa alla prova durante lo stage e il lavoro di tesi che ne è uscito: muoversi, col supporto del relatore, in una analisi tutta mia è stato un bel punto di verifica per tutto quello che avevo imparato fino a quel momento.

Durante tutto questo percorso le cose che ho gustato di più sono state il veder emergere la mia capacità di mettermi in discussione (non solo nel confronto con docenti e compagni ma soprattutto rispetto ai dati con cui avevo a che fare) e l'opportunità di incontrare professori da cui imparare questo mestiere.

Andrea Riganti

Ciao mi chiamo Andrea e ho frequentato il corso di Statistica e Gestione delle Informazioni qualche anno fa, ora sono a metà di un percorso di dottorato.

La cosa che ho apprezzato di più del mio studio della statistica, e che vedo ancora essere determinante per il mio lavoro di ricerca, è stato l'imparare a ragionare su un problema reale, non semplicemente prendendo i dati e buttandoli dentro ad un programma che facesse i conti al posto mio, ma provando a pensare criticamente come avrei potuto comportarmi e che relazioni avrei potuto indagare. Infatti penso che la caratteristica del corso di laurea che più mi ha affascinato è stato l'insegnamento da parte di molti docenti di un modo per affrontare le domande di ricerca in maniera, appunto, critica. La grande possibilità che ho avuto è stata quella di trovare molti insegnanti che non si accontentavano di darmi le formule e insegnarmi come applicarle, ma che volevano far emergere

la creatività di ciascuno nell'impostare un certo problema, valorizzando le idee e il modo che ciascuno per tentativi voleva provare a indagare una certa relazione.

Sicuramente un grande aiuto è stata la possibilità, visto il grande numero di docenti se paragonato all'esiguo numero di studenti, di poter approfondire e consolidare il rapporto tra docente e studenti, facendo sì che le domande mie e dei miei compagni potessero essere prese in considerazione a lezione, si potesse in alcuni casi approfondire un determinato argomento e si potesse essere effettivamente seguiti.

Luca Cilumbriello

Mi è sempre piaciuto “dare i numeri”. Per questo motivo mi sono iscritto nel 2009 presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Milano-Bicocca. Una scelta che a distanza di anni mi convince sempre di più.

Nel 2012 ho conseguito la Laurea Triennale in Statistica e Gestione delle Informazioni al termine di un triennio durante il quale ho avuto l'opportunità di conseguire la certificazione in SAS Base Programming in collaborazione con SAS Institute di Milano e di svolgere uno stage semestrale presso il Dipartimento di Epidemiologia dell'Istituto di Ricerche Farmacologiche “Mario Negri” di Milano. Qui ho redatto la mia prima tesi, investigando i fattori di rischio per lo sviluppo della Dermatite Atopica nei neonati e approfondendo la veridicità della cosiddetta Hygiene Hypothesis.

Nel 2014 ho conseguito la Laurea Magistrale in Biostatistica e Statistica Sperimentale al termine di uno stage annuale presso il

Settore Monitoraggi Ambientali dell’Agenzia Regionale Protezione Ambiente (ARPA) della Lombardia. Insieme al team dell’Unità Operativa per le Acque abbiamo sviluppato e pubblicato la nostra “Analisi geo-spaziale sulla concentrazione dei metalli pesanti nelle acque sotterranee degli acquiferi della Regione Lombardia e Proposta di definizione di Valore di Fondo Naturale basato sul Kriging”.

Il mondo della ricerca scientifica mi ha dato tanto e cerco sempre di restare aggiornato sui temi che ho avuto modo di approfondire durante i miei studi. Ho sempre i miei appunti a portata di mano. Il mondo del lavoro mi ha chiamato a poche settimane dal termine degli studi. Le proposte ricevute sono state variegiate (“voi statistici potete andare dappertutto” mi sono sentito dire più volte). La mia scelta si è concentrata su un’assunzione a tempo indeterminato. In Italia, a Milano, la mia città. Ora sono inquadrato come Statistical Consultant e porto avanti differenti progetti in SAS e in R in collaborazione con i team di Advanced Analytics and Data Science principalmente in ambito Healthcare a servizio delle più importanti strutture sanitarie pubbliche e private del Nord Italia.

Elenco degli insegnamenti e dei programmi

Indicazioni generali

Dall’anno accademico 2008-2009 è entrata in vigore la Riforma universitaria, la cosiddetta 270⁴.

⁴Per maggiori informazioni consultare il sito: www.miur.it

Con la nuova Riforma, alcuni insegnamenti sono articolati in moduli. Per superare l'esame previsto per ciascuno insegnamento organizzato in moduli, lo studente dovrà sostenere una o più prove, secondo le modalità che verranno precisate, all'inizio delle lezioni, dai docenti interessati.

I programmi illustrati nelle pagine successive sono gli insegnamenti:

- del primo anno di SGI, come da Regolamenti 2018 - 2019
- del secondo anno di SGI, come da Regolamenti 2017 - 2018
- del terzo anno di SGI, come da Regolamenti 2016 - 2017.

Si consiglia di visitare le pagine personali dei docenti per il materiale didattico dei corsi nonché per altre informazioni riguardanti la didattica.

Si accede alle pagine personali degli insegnamenti attraverso la piattaforma e-learning:

<http://elearning.unimib.it/> seguendo il percorso Economia e statistica, Corso di Laurea triennale, Statistica e Gestione delle Informazioni.

In caso di difformità riscontrate tra la Guida e i Regolamenti dei Corsi di Laurea, fare riferimento ai Regolamenti.

Capitolo 2

Corsi

2.1 Algebra Lineare

Docente: Bandini Elena

Mail: elena.bandini@unimib.it

Ore: 42

Tipo di attività: Obbligatorio

Settore disciplinare: MAT/02

Obiettivi formativi

L'obiettivo è fornire le conoscenze propedeutiche di algebra lineare ai corsi Calcolo delle Probabilità e Analisi statistica multivariata.

Contenuti sintetici

Spazi vettoriali e applicazioni lineari; matrici e diagonalizzazione; determinanti; similitudine di matrici.

Programma esteso

\mathbb{R}^n e i suoi sottospazi vettoriali. Spazi vettoriali. Sistemi di generatori e basi. Dimensione. Prodotto scalare standard e disuguaglianza di Cauchy-Schwartz. Norma e disuguaglianza triangolare. Angolo tra vettori. Elementi di geometria analitica. Applicazioni lineari. Teorema nullità più rango. Studio di applicazioni lineari mediante matrici e sistemi lineari. Basi ortonormali. Proiezioni ortogonali. Determinante e suo calcolo mediante operazioni per righe. Invertibilità. Autovalori e autovettori. Diagonalizzabilità. Matrici ortogonali e simmetriche. Teorema spettrale.

Prerequisiti

Nessun prerequisito formale richiesto.

Metodi didattici

Lezioni.

Modalità di verifica dell'apprendimento

Prova scritta a base di esercizi e domande teoriche per verificare la capacità di risoluzione di semplici problemi mediante l'applicazione della teoria acquisita.

Orale facoltativo (su richiesta del docente o dello studente).

Testi di riferimento

T.M. Apostol, Calcolo, volume secondo (Geometria), Bollati Boringhieri, 2003.

Periodo di erogazione dell'insegnamento

II semestre, III e IV ciclo (periodo approssimativo da marzo a giugno).

Lingua di insegnamento

Italiano.

2.2 Analisi Matematica I

Docente: Travaglini Giancarlo

Mail: giancarlo.travaglini@unimib.it

Ore: 63

Tipo di attività: Obbligatorio

Settore disciplinare: MAT/05

Obiettivi formativi

Il principale obiettivo di questo insegnamento è fornire una preparazione rigorosa sul calcolo differenziale ed integrale in una variabile.

Contenuti sintetici

Linguaggio comune e linguaggio matematico.

Numeri reali.

Successioni e serie.

Calcolo differenziale in una variabile.

Sviluppi e serie di Taylor.

Integrale di Riemann in una variabile.

Funzioni integrali e funzioni di ripartizione.

Programma esteso

Linguaggio comune e linguaggio matematico. Proposizioni e proprietà, variabili logiche. Il linguaggio degli insiemi. Implicazioni,

dimostrazioni e contresempi. Negazioni e dimostrazioni indirette. Sostituzione di una variabile in una formula. Uso degli indici: sommatorie. Lo studio di un libro di Matematica. Definizioni astratte ed esempi. Studio di una dimostrazione: verifica dei passaggi, considerazione di opportuni esempi, applicazione a situazioni analoghe.

Numeri reali. Proprietà metriche ed aritmetiche. Potenze con esponente reale.

Equazioni e disequazioni. Estremo superiore. Limiti di successioni. Successioni monotone. Forme di indecisione. Il numero e . Serie numeriche. La serie geometrica.

Limiti di funzioni e proprietà delle funzioni continue. Funzioni composte e loro limiti. Derivate. Studio del comportamento locale e globale di una funzione. Il teorema del valor medio. Derivate successive. Convessità. Sviluppi e serie di Taylor. La serie esponenziale.

Integrale di Riemann. Teorema fondamentale del Calcolo Integrale. Tecniche di integrazione.

Integrale di Riemann generalizzato: criteri di convergenza. Serie numeriche e integrali generalizzati. La funzione Gamma. Funzioni integrali e loro grafici. Funzioni di ripartizione.

Prerequisiti

Nessun prerequisito formale richiesto. È però necessario avere familiarità con i seguenti argomenti.

- + Algebra elementare (disequazioni di II grado e irrazionali, esponenziali e logaritmi),
- + Geometria Euclidea elementare,

- + Geometria Analitica elementare,
- + Trigonometria (funzioni goniometriche, equazioni e disequazioni),
- + Proprietà elementari dei numeri interi e dei numeri razionali.

Metodi didattici

Lezioni in aula. L'insegnamento è affiancato da attività di tutoraggio in cui si svolgono esercitazioni pratiche. Sono periodicamente assegnati numerosi esercizi da svolgere a casa, che vengono corretti individualmente.

Modalità di verifica dell'apprendimento

Esame scritto ed esame orale.
Non ci sono prove intermedie.

Testi di riferimento

M. Bramanti, C. Pagani, S. Salsa, *Analisi Matematica I*, Zanichelli.

M. Bramanti, G. Travaglini, *Matematica. Questione di Metodo*, Zanichelli.

M. Bramanti, *Precalculus*, Progetto Leonardo, Esculapio.

M. Bramanti, *Esercizi di Calcolo Infinitesimale e Algebra Lineare*, Seconda Edizione, Progetto Leonardo, Esculapio.

M. Boella, *Analisi matematica e algebra lineare*, vol.1, Pearson.
Appunti, video e centinaia di esercizi svolti a

<http://elearning.unimib.it/enrol/index.php?id=19508> <http://elearning.unimib.it/enrol/index.php?id=19508>

Periodo di erogazione dell'insegnamento

Primo semestre (da settembre a gennaio)

Lingua di insegnamento

Italiano

2.3 Analisi Matematica II

Docente: Travaglini Giancarlo

Mail: giancarlo.travaglini@unimib.it

Ore: 42

Tipo di attività: Obbligatorio

Settore disciplinare: MAT/05

Obiettivi formativi

Lo scopo di questo insegnamento è fornire una preparazione rigorosa sulle serie di Fourier e sul calcolo differenziale ed integrale in d variabili.

Contenuti sintetici

Serie di Fourier.

Calcolo differenziale in R^d .

Integrazione in R^d .

Programma esteso

Serie di Fourier. Legge di Benford.

Calcolo differenziale in più variabili. Derivate parziali, differenziabilità, gradiente e piano tangente.

Massimi e minimi liberi. Derivate successive, polinomi di Taylor, matrice Hessiana. Retta di regressione.

Massimi e minimi vincolati. Funzioni definite implicitamente.
Metodo dei moltiplicatori di Lagrange.

Funzioni convesse.

Integrazione in R^d . Cambi di variabili.

Metodo Monte Carlo.

Integrali generalizzati. Integrazione di funzioni radiali su R^d .

Prerequisiti

Avere superato gli esami di Analisi Matematica I e Algebra Lineare.

Metodi didattici

Lezioni in aula. L'insegnamento è affiancato da attività di tutoraggio in cui si svolgono esercitazioni pratiche. Sono periodicamente assegnati numerosi esercizi da svolgere a casa, che vengono corretti individualmente.

Modalità di verifica dell'apprendimento

Esame scritto ed esame orale.

Non ci sono prove intermedie.

Testi di riferimento

M. Bramanti, C. Pagani, S. Salsa, Analisi Matematica 2, Zanichelli.

M. Bramanti, Esercizi di Calcolo Infinitesimale e Algebra Lineare, Seconda Edizione, Progetto Leonardo, Esculapio.

M. Boella, Analisi Matematica 2, Pearson.

Appunti e centinaia di esercizi svolti a <http://elearning.unimib.it/course/view.php?id=19499>

Periodo di erogazione dell'insegnamento

Primo semestre. Primo ciclo (da settembre a novembre).

Lingua di insegnamento

Italiano

2.4 Analisi Statistica Multivariata

2.4.1 Modulo: Analisi Esplorativa

Docente: Solaro Nadia

Mail: solaro.nadia@unimib.it

Ore: 49

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il modulo di Analisi Esplorativa introduce i principali metodi statistici descrittivi per lo studio di due o più fenomeni osservabili congiuntamente su un insieme di unità statistiche. Si tratta di metodi finalizzati all'esplorazione dei dati multivariati per individuare la struttura soggiacente e ridurre la dimensionalità in modo da preservare le caratteristiche principali osservate. Dal punto di vista applicativo l'analisi dei dati viene affrontata con il ricorso al software R in ambiente RStudio.

Contenuti sintetici

Introduzione all'analisi statistica multivariata, matrici di dati quantitativi, qualitativi e misti, rappresentazioni grafiche per dati multidimensionali. Cluster Analysis: metodi di raggruppamento gerarchici e non gerarchici. Analisi delle componenti principali. Analisi fattoriale esplorativa. Uso integrato dei metodi esplorativi

di analisi multivariata. Applicazioni a dati reali con il software R in ambiente RStudio.

Programma esteso

- Introduzione all'analisi statistica multivariata: scuola francese e scuola anglosassone, classificazione delle metodologie di analisi multivariata
- Matrici di dati quantitativi, qualitativi e misti. Principali sintesi e trasformazioni. Rappresentazione dei dati, spazio degli individui e spazio delle variabili. Dissimilarità e distanze fra unità, distanze fra variabili
- Cluster Analysis: metodi di raggruppamento gerarchici e non gerarchici, bontà della classificazione, applicazione a variabili quantitative e qualitative
- Analisi delle componenti principali: estrazione delle componenti principali, criteri di arresto, valutazione della variabilità riprodotta, interpretazione delle componenti principali, applicazioni
- Analisi fattoriale esplorativa: modello fattoriale, studio delle correlazioni, metodi di estrazione dei fattori, rotazione dei fattori, punteggi fattoriali, applicazioni
- Uso integrato delle tecniche esplorative di analisi multivariata
- Analisi di casi empirici con RStudio

Prerequisiti

Superamento degli esami degli insegnamenti propedeutici di I anno: Statistica I, Analisi Matematica I, Algebra Lineare, Calcolo delle Probabilità

Metodi didattici

Lezioni teoriche in aula ed esercitazioni pratiche in laboratorio statistico-informatico con il software R in ambiente RStudio

Modalità di verifica dell'apprendimento

L'esame consiste in una prova scritta (durata complessiva: 2 ore) con tre quesiti (articolati in più punti) che riguardano gli aspetti sia teorici sia applicativi delle metodologie trattate nel corso. Un quarto quesito è facoltativo e riguarda la programmazione con il software R. Le domande a natura teorica riguardano gli aspetti metodologici degli argomenti trattati al corso. Le domande a natura applicativa riguardano sia esercizi numerici, sia la lettura e il commento di parti di output di R.

La prova orale è facoltativa (su richiesta del docente o dello studente) e riguarda argomenti sia teorici sia pratici. L'accesso alla prova orale è subordinato al superamento della prova scritta con un esito di almeno 18/30. Si fa presente che la prova orale può comportare sia l'aumento, sia il mantenimento, che la diminuzione della valutazione conseguita alla prova scritta.

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.

Testi di riferimento

- Materiale didattico della docente pubblicato sul sito e-learning del corso
- Frosini, B.V. (2014). Complementi di analisi statistica multivariata, EDUCatt, Milano
- Zani, S., Cerioli, A. (2007). Analisi dei dati e data mining per le decisioni aziendali, Giuffrè Editore, Milano
- Bolasco, S. (1999). Analisi multidimensionale dei dati: strategie e criteri di interpretazione, Carocci, Roma
- Dillon, W.R., Goldstein, M. (1984). Multivariate Analysis, J. Wiley, New York
- Everitt, B.S., Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R, Springer, Berlin

Periodo di erogazione dell'insegnamento

I Semestre, II ciclo

Lingua di insegnamento

Italiano

2.4.2 Modulo: Modelli Statistici

Docente: Pennoni Fulvia

Mail: fulvia.pennoni@unimib.it

Ore: 57

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il corso intende sviluppare le conoscenze teoriche e applicative circa il modello di regressione lineare multipla.

Vengono trattati aspetti di analisi grafica, e analisi computazionale utilizzando la notazione matriciale. Il corso presenta un taglio sia teorico che applicativo basato sulla discussione dei risultati delle tecniche esposte per l'analisi dei dati aventi natura multivariata e provenienti da varie fonti informative.

Contenuti sintetici

Durante il corso vengono trattati gli argomenti secondo la seguente sintesi.

Lo studente viene introdotto alle analisi grafiche dei dati e all'utilizzo dei coefficienti di correlazione totali e parziali.

Si considerano le assunzioni del modello di regressione lineare multipla ed il metodo dei minimi quadrati per la stima dei parametri. Si espongono i concetti per l'inferenza statistica sui parametri del modello.

Viene introdotta la distribuzione di Gauss bivariata e multivariata e le relative proprietà vengono illustrate anche con esempi applicativi su dati reali e simulati.

Si considera il modello a fini presivisi e si impara a valutare il modello utilizzando sia gli indici di adattamento per la scelta del numero di variabili esplicative, sia le analisi grafiche dei residui, sia i criteri d'informazione.

Programma esteso

Nel presente corso vengono trattate delle tecniche avanzate di inferenza statistica utilizzando le distribuzioni multivariate tra cui la distribuzione di Gauss bivariata e multivariata. In particolare si illustrano le matrici di varianza e covarianza, di correlazione e delle correlazioni parziali. Viene anche introdotto l'ellissoide di concentrazione. Le realizzazioni simulate da queste distribuzioni sono illustrate con i grafici a dispersione sia in due sia in tre dimensioni.

Il modello di regressione lineare multipla è introdotto anche nella notazione matriciale. Si illustra la scomposizione della devianza totale. Nel caso del modello con due variabili esplicative si illustra il metodo del determinante per ottenere le stime dei parametri ed il calcolo dell'indice di determinazione lineare. Viene considerato il miglioramento in varianza residua passando dalla retta di regressione lineare al piano di regressione. Viene illustrato l'indice di inflazione delle varianze.

Sono fornite le proprietà degli stimatori in base alle assunzioni del modello l'inferenza sui parametri viene presentata sia a livello

del singolo parametro che per coppie di parametri.

Si illustrano varie analisi diagnostiche riguardanti i residui per la valutazione del modello ed alcuni criteri statistici tra cui il criterio di informazione Bayesiano per la selezione del modello. Viene illustrato il metodo di previsione per unità esterne al campione.

Nel corso si accennano anche i seguenti aspetti : i) metodo di stima della massima verosimiglianza; ii) trasformazione delle variabili; iii) variabili esplicative categoriali; iv) modelli con ordini di interazione tra variabili esplicative; v) variabile risposta categoriale con particolare riferimento a di variabili di conteggio.

Gli argomenti trattati a livello teorico sono affiancati dall'illustrazione di numerose applicazioni su dati reali e simulati sviluppate tramite l'ambiente statistico R, Rstudio, RMarkdown, ed il software SAS.

Prerequisiti

Si richiede di aver superato gli esami degli insegnamenti propedeutici come da regolamento didattico:

Statistica I, Analisi Matematica I, Algebra Lineare, Calcolo delle Probabilità

Metodi didattici

Sono previste lezioni frontali sulla parte di teoria. Le lezioni si svolgono in laboratorio informatico in modo che congiuntamente alla teoria vengono proposte delle esercitazioni con applicazioni degli aspetti teorici a problemi concreti relativi a dati multivaria-

ti reali e simulati riferiti a ambiti applicativi diversi (es. Biologia, Astronomia, Economia, etc...). Durante le esercitazioni con l'ausilio di R nell'ambiente RStudio e l'interfaccia di RMarkdown e SAS gli studenti sono incoraggiati ad affrontare il problema applicativo con lo scopo ulteriore di sviluppare l'apprendimento cooperativo. In questo modo tramite l'interazione sono incentivati a riconoscere la problematica dell'esercizio, individuare la metodologia adatta, applicare le analisi richieste e commentare i risultati delle analisi in modo critico.

Modalità di verifica dell'apprendimento

L'esame ha durata complessiva di un ora e trenta minuti. Si svolge presso il laboratorio informatico e consiste in una prima parte a libro chiuso riguardante la teoria con domande aperte da svolgere senza l'ausilio del calcolatore. La seconda parte si svolge al computer e lo studente deve dimostrare la sua abilità nell'applicare il modello di regressione lineare multipla rispetto a dati reali o simulati. Lo studente deve fornire dei commenti dettagliati al codice impiegato e ai risultati ottenuti con l'ausilio dell'ambiente R oppure con il software SAS. Lo studente è chiamato ad illustrare gli aspetti rilevanti dei risultati sia dal punto di vista statistico che di contenuto specifico inerente l'ambito applicativo. E' possibile utilizzare il codice di R e SAS fornito durante le lezioni.

La prima parte pesa 1/3 sul totale del punteggio.

La prova orale è facoltativa e possono richiederla solo coloro che hanno un esito di almeno 18/30 alla prova scritta. L'esito della prova scritta viene pubblicato sulla rispettiva pagina di elearning

e lo studente deve presentarsi nella data stabilita per la visione del compito.

Testi di riferimento

Il materiale didattico si compone di diapositive dispense e programmi di calcolo. Gli studenti sono invitati a studiare principalmente le dispense redatte dal docente caratterizzanti sia la parte di teoria che di applicazioni. Tutto il materiale didattico è reso disponibile da parte del docente presso la pagina della piattaforma e-learning dell'ateneo dedicata al corso.

I principali testi di riferimento oltre a quelli che costituiscono i riferimenti nelle dispense sono i seguenti:

Faraway, J. J. (2014). *Linear models in R*, Second Edition, Chapman & Hall, CRC Press.

Johnson, R. A., and Wichern, D. W. (2002). *Applied multivariate statistical analysis*, Pearson Education International, Prentice Hall. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*, New York, Springer.

Nolan, D., & Lang, D. T. (2015). *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*. Chapman & Hall, CRC Press.

SAS/STAT 12.1. *User's guide*, SAS Institute, 2012.

Zenga, M. (1995). *Metodi statistici per l'economia*, Giappichelli.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Periodo di erogazione dell'insegnamento

Il Semestre, III Ciclo: Febbraio - Aprile 2018, secondo un calendario di 4 giorni.

Lingua di insegnamento

Il corso viene erogato in lingua italiana.

Gli studenti Erasmus hanno in supporto il materiale didattico in Inglese. Questi hanno anche la possibilità di svolgere l'appello d'esame in lingua inglese.

2.5 Analisi dei Dati

Docente: Vittadini Giorgio
Mail: giorgio.vittadini@unimib.it
Ore: 47
Tipo di attività: Obbligatorio
Settore disciplinare: SECS-S/01

Obiettivi formativi

Il corso affronta il tema delle principali tecniche di analisi dei dati sia a partire da dati quantitativi che qualitativi. A lezioni frontali si alterneranno esercitazioni in laboratorio in modo da permettere l'apprendimento di pacchetti statistici e poter quindi applicare le tecniche apprese.

Contenuti sintetici

Analisi della correlazione canonica
Analisi discriminante
Analisi delle corrispondenze
Multidimensional scaling

Programma esteso

Introduzione analisi dei dati
A Analisi della correlazione canonica
1 Scopo

- 2 Modo per ricavare 1° variabile canonica e successive
- 3 Lettura risultati
- 4 Esempi
- 5 Esercitazioni con SAS
- B Analisi discriminante
- 1 Scopo
- 2 I dati
- 3 Modo per ricavare 1° variabili discriminante e successive
- 4 Lettura risultati
- 5 Modo alternativo per ricavare variabili discriminanti
- 6 Esempi
- 7 Esercitazioni con SAS
- C Analisi corrispondenze
- 1 Scopo
- 2 Matrici profili riga e colonna
- 3 Analisi corrispondenze in ottica correlazione canonica e componenti principali
- 4 Lettura risultati: scomposizione χ^2
- 5 Lettura altri risultati
- 6 Analisi delle corrispondenze multiple
- 7 Esempi
- 8 Esercitazioni con SAS
- D Multidimensional scaling
- 1 Multidimensional scaling metrico con distanze euclidee
- 2 Teorema di Torgerson
- 3 Lettura risultati
- 4 Esempi
- 5 Esercitazioni con SAS
- 6 Multidimensional scaling con dissimilarità metriche

7 Multidimensional scaling con dissimilarità ordinali

8 Esempi

9 Esercitazioni con SAS

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento dell'esame di Analisi Statistica Multivariata.

Metodi didattici

Il corso verrà erogato sia mediante lezioni frontali sia mediante esercitazioni pratiche in laboratorio informatico con il linguaggio SAS. Durante le lezioni verranno fornite le conoscenze teoriche dei metodi statistici oggetto del corso, mentre nelle esercitazioni si applicheranno queste tecniche ad alcuni casi reali.

Modalità di verifica dell'apprendimento

Esame scritto, prova pratica in laboratorio (da svolgersi durante l'esame scritto) e esame orale. L'esame scritto consiste in due domande teoriche sui metodi oggetto del corso, mentre la prova pratica consiste in un esercizio da svolgere con il software SAS, in analogia a quanto svolto durante le esercitazioni. L'esame orale consiste nella discussione dell'esame scritto.

Testi di riferimento

Dispense e lucidi del docente.

Vitali O., Statistica per le scienze applicate. Vol. 2, Cacucci Editore, 1993 (per correlazione canonica e analisi discriminante).

Zani S., Cerioli A. Analisi dei dati e data mining per le decisioni aziendali. Giuffrè, 2007 (per analisi delle corrispondenze e multidimensional scaling).

Periodo di erogazione dell'insegnamento

2° semestre 2° ciclo

Lingua di insegnamento

Italiano

2.6 Basi di Dati

Docente: Mezzanzanica Mario

Mail: mario.mezzanzanica@unimib.it

CFU: 6

Ore: 47

Tipo di attività: Obbligatorio

Settore disciplinare: ING-INF/05

Obiettivi formativi

Il Corso intende creare le necessarie conoscenze, sotto il profilo tecnico e metodologico, che consentano la comprensione del modello di una base di dati relazionale per il supporto decisionale (Database e Data warehouse) quale risorsa strategica essenziale al raggiungimento degli obiettivi di un'organizzazione aziendale. Il corso inoltre intende fornire le competenze tecniche per l'interrogazione mediante il linguaggio SQL.

Contenuti sintetici

Progettazione di basi di dati

La progettazione concettuale e logica

La normalizzazione

Basi di dati per il supporto alle decisioni

Eventuali variazioni saranno comunicate dal docente in aula

Programma esteso

- Progettazione di basi di dati
 - o Metodologie e modelli per il progetto
 - o Introduzione alla progettazione: il ciclo di vita dei sistemi informativi, metodologie di progettazione e basi di dati.
 - o Il modello Entità - Relazione: i costrutti principali del modello, panoramica finale del modello E-R
 - o Documentazione di schemi E-R, regole aziendali, tecniche di documentazione.
 - La progettazione concettuale e logica
 - o L'analisi e la specifica dei requisiti
 - o Strategie di progetto: strategia top-down, strategia bottom-up, strategia inside-out, strategia mista
 - o Qualità di uno schema concettuale
 - o Strumenti CASE per la progettazione di basi di dati
 - La normalizzazione
 - o Ridondanze e anomalie
 - o Dipendenze funzionali
 - o Forma normale di Boyce e Codd: definizione e decomposizione in forma normale di Boyce e Codd
 - o Progettazione di basi di dati e normalizzazione: verifiche di normalizzazione su entità e su associazioni, ulteriori decomposizioni di associazioni, ulteriori decomposizioni di schemi concettuali
- Basi di dati per il supporto alle decisioni
 - o Architetture e paradigmi per l'analisi dei dati
 - o Architettura dei data warehouse

- o Schemi dei data warehouse : schema a stella, schema a fiocco di neve
 - o Operazioni per l'analisi dei dati, interfacce per la formulazione di query, drill down e roll up, data cube.
 - o Data mining: il processo di data mining, problemi di data mining, prospettive del data mining
- Eventuali variazioni saranno comunicate dal docente in aula.

Prerequisiti

Nessuno

Metodi didattici

Lezioni frontali, seminari monotematici, esercitazioni, assegnamenti da svolgere a casa.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su una prova scritta e prova orale obbligatorie.

La prova scritta si svolge al computer ed è composta da domande aperte e chiuse e risposta multipla su tutti gli argomenti del corso.

In sede di valutazione viene considerata la capacità dello studente di rispondere a quesiti specifici facendo riferimento agli aspetti teorici e pratici (mediante esempi) connessi all'argomento richiesto.

La prova scritta è comune sia per gli studenti frequentanti sia per i non frequentanti.

La prova orale è mirata ad accertare la conoscenza teorica dello studente sugli argomenti del corso. Saranno quindi valutate le capacità di ragionare e approfondire le tematiche proposte in sede di esame e il rigore metodologico del loro sviluppo.

Testi di riferimento

Chianese, Moscato, Picariello e Sansone: Sistemi di basi di dati e applicazioni. Maggioli Editore, 2015.

Periodo di erogazione dell'insegnamento

I semestre, II ciclo

Lingua di insegnamento

Italiano

2.7 Calcolo delle Probabilità

Docente: Masiero Federica

Mail: federica.masiero@unimib.it

CFU: 6

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MAT/06

Obiettivi formativi

Presentare rigorosamente i concetti fondamentali della teoria della probabilità. Presentare in maniera sistematica le variabili aleatorie multidimensionali e la convergenza di variabili aleatorie, nozioni fondamentali utilizzate nei corsi di Statistica.

Contenuti sintetici

- Richiami sui concetti base della probabilità.
- Variabili aleatorie e variabili aleatorie multidimensionali.
- Funzioni caratteristiche.
- Indipendenza di variabili aleatorie
- Convergenza di variabili aleatorie.
- Leggi dei Grandi Numeri e Teorema del Limite Centrale.

Programma esteso

- Richiami sui concetti base della probabilità, misure di probabilità e loro proprietà.
- Definizione di variabile aleatoria, sua media e momenti pesanti; legge di una variabile aleatoria..
- Variabili aleatorie multidimensionali, densità congiunta, vettore delle medie e matrice di covarianza
- Funzioni caratteristiche: definizione, proprietà principali, corrispondenza tra distribuzioni e funzioni caratteristiche.
- Indipendenza di variabili aleatorie
- Convergenza di variabili aleatorie: quasi certa, in media pesanti, in probabilità e in legge
- Leggi dei Grandi Numeri e Teorema del Limite Centrale.
- Cenni ai processi stocastici.

Prerequisiti

Nessuna propedeuticità formale. Tuttavia, questa attività formativa richiede le conoscenze di un corso base di probabilità, dei corsi di analisi matematica e algebra lineare quali quelli impartiti in un CdL in Statistica.

Metodi didattici

Lezioni frontali, con l'obiettivo di dare i fondamenti teorici del Calcolo delle Probabilità; verranno inoltre svolti esercizi sugli argomenti del corso.

Modalità di verifica dell'apprendimento

Esame scritto, con esercizi e una domanda di teoria, e orale, con domande teoriche; l'orale si svolge nello stesso giorno dello scritto.

Testi di riferimento

- P. Baldi, Calcolo delle Probabilità
- F. Caravenna, P. Dai Pra, Probabilità, Un'introduzione attraverso modelli e applicazioni
- A. F. Karr, Probability

Periodo di erogazione dell'insegnamento

Primo semestre, primo ciclo.

Lingua di insegnamento

Italiano

2.8 Complex Data Analysis

Docente: Fattore Marco

Mail: marco.fattore@unimib.it

Ore: 45

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Obiettivi formativi

Rendere gli studenti in grado di affrontare l'impostazione, la conduzione e la gestione di progetti di data science ed estrazione di informazioni da sistemi complessi di dati, integrando le proprie competenza tecniche con elementi di project management e di comunicazione.

Il corso si concentra in particolare sull'impostazione e conduzione di progetti di estrazione dell'informazione in ambiti "aperti", cioè in assenza di obiettivi analitici precisi, come nella pratica statistica "classica", ponendosi nelle condizioni tipiche dei contesti attuali (sia in ambito aziendale che istituzionale), dove i flussi dati costituiscono la base per l'ideazione, il design e l'implementazione di nuovi servizi, richiedendo al "data scientist" capacità creative, abilità organizzativa e comunicativa e solidità metodologica.

Contenuti sintetici

Il corso illustra la tipologia di attività che il data scientist è chiamato a svolgere nel contesto attuale, affronta la tematica della natura e della gestione di progetti di data science in ambito aziendale e istituzionale e si concreta nell'assegnazione di progetti di elaborazione dei dati (presi da ambiti reali) e generazione di contenuti informativi, che gli studenti, suddivisi in gruppi, devono condurre a termine nel periodo del corso, con la supervisione del docente.

Programma esteso

Il corso si suddivide in due parti logiche distinte. Nella prima, si effettueranno alcune attività "seminariali" di focalizzazione dei concetti base. Nella seconda, si assegneranno agli studenti i progetti che dovranno essere svolti nell'ambito del corso, con la supervisione del docente.

PARTE I

1. La data science: affinità e differenze con la statistica "classica".
2. Il contesto tecnologico ed economico attuale: complessità dei processi socio-economici e nuove esigenze conoscitive e di servizi a valore aggiunto.
3. Complessità dei dati e nuove fonti: web, e-commerce, Internet of Things, Smartphone...
4. Esempi di progetti di data science
5. Elementi di project management e caratteristiche/criticità dei progetti di data science: obiettivi, qualità dei dati, scelte tecnologiche.

PARTE II

Assegnazione dei progetti, impostazione delle modalità operative di conduzione delle attività e supervisione da parte del docente.

Prerequisiti

Non sono previsti prerequisiti formali, ma è necessaria una conoscenza delle tecniche di base della statistica inferenziale, dell'analisi multivariata e di data mining, nonché la conoscenza di base della programmazione R.

Metodi didattici

Lezioni frontali e supervisione periodica dei progetti di elaborazione dei dati.

Modalità di verifica dell'apprendimento

L'apprendimento verrà verificato attraverso:

1. Una valutazione progressiva del modo in cui gli studenti affronteranno le difficoltà nella conduzione dei progetti.
2. Una presentazione orale del lavoro ai docenti, con discussione critica

Questa modalità di verifica dell'apprendimento è motivata dall'obiettivo di mettere gli studenti nelle condizioni operative tipiche dell'attività lavorativa e di farne in particolare emergere le abilità soft (organizzative, comunicative, creative...).

Testi di riferimento

Computers Ltd. What they really can't do, Harel D. =xfordUniversity Press, 2000

Documenti e slide fornite dal docente.

Risorse web e TED Talks

Periodo di erogazione dell'insegnamento

Il semestre IV ciclo

Lingua di insegnamento

Italiano

2.9 Data Mining e Statistica Computazionale

Docente: Lovaglio Pietro Giorgio
Mail: piergiorgio.lovaglio@unimib.it
Ore: 105
Tipo di attività: Obbligatorio
Settore disciplinare: SECS-S/01

Obiettivi formativi

Statistica computazionale

L'obiettivo principale del corso è introdurre strumenti software avanzati e di alta complessità computazionale per disegnare ed eseguire analisi di dati e modellazione statistica complessa.

Data mining

Il corso intende fornire un'introduzione alle principali tecniche statistiche di Data Mining attraverso le più moderne tecniche e strategie per l'analisi di grandi moli di dati, illustrando le problematiche connesse.

Alla fine del corso lo studente ha la possibilità di proporre i principali algoritmi, discernendo pregi e difetti, essendo in grado di sperimentare ed applicare le conoscenze acquisite su dati reali.

Contenuti sintetici

Il corso affronta lo studio di tecniche modellistiche complesse e le principali problematiche e tecniche statistiche di Data Mining

Programma esteso

Statistica computazionale

- (1) SAS language
- (2) Interpretazione di Modelli lineari complessi (Anova, Ancova, GLM) con interazioni, trasformate,
- (3) Robust methods (Bootstrap, Jackknife, Robust Regression, IRLS, WLS, nonparametric regression, loess smoothing and splines)
- (4) Passi per costruzione di un modello Robusto
- (5) Influenza, diagnostiche, model selection,
- (6) Logistic Regression

Data mining

Il Data mining, robustezza, overfitting e problematiche di validazione dei risultati, Regole associative, Modelli statistici per la classificazione supervisionata (modello lineare, analisi discriminante parametrica, modello logistico polinomiale e ordinale), Algoritmi per la classificazione supervisionata (Naive Bayes, Nearest Neighbour, neural network, Alberi decisionali e Classificativi, PLS, Bagging, Boosting and Random forest)

Prerequisiti

Superamento esame di Analisi statistica Multivariata

Metodi didattici

Lezione frontale e sessioni di laboratorio

Modalità di verifica dell'apprendimento

PROVA SCRITTA

PROJECT WORK (Sviluppo di un progetto originale a partire da una semplice idea o dall'analisi di un caso esistente)

Lavoro applicativo da svolgere autonomamente o in gruppo di max 3 persone su dataset scelti dallo studente (R o SAS) su cui applicare i principali argomenti svolti a lezione .

I principali output del PROJECT WORK

(svolto nelle due/tre settimane precedenti la data dell'orale) vanno stampati e portati all'orale.

Di seguito le analisi da svolgere per i due moduli:

Di seguito le analisi da svolgere per le due analisi applicative (Sas base o R):

Statistica computazionale

1 PROJECT WORK completo con con target quantitativo

(analisi descrittive, modelli, trasformazioni, diagnostiche, modello robusto)

1 Analisi logistica con target binario (modello logistico con tutte le covariate e con model selection, controllare solo collinearity e separation)

Data mining (sas Enterprise Miner o R)

1 PROJECT WORK, analisi con con target binario

(ANALISI DA SVOLGERE: analisi descrittive, analisi classificazione, preprocessing, tuning modelli, confronto modelli, score di nuovi dati)

In totale tre project work su tre dataset differenti

Portali per la scelta dei dataset:

<https://archive.ics.uci.edu/ml/datasets>

www.kaggle.com

PROVE ORALI

COLLOQUIO DI DISCUSSIONE SULLO SCRITTO

L'esame orale, per ciascun modulo, consta di domande sulla TEORIA affrontata a lezione e sul commento degli output del lavoro applicativo per verificare la comprensione dei principali strumenti adottati e il conseguente "modus operandi" dell'analisi statistica svolta.

Non sono previste prove in itinere

Testi di riferimento

Statistica computazionale

Gareth, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R

Chapter 3 (no section 3.5), Chapter 4, 6,7

Lucidi sul moodle

Data mining

Gareth, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R

Chapter 2-3-4-5- 8

Lucidi sul moodle

Periodo di erogazione dell'insegnamento

I semestre, cicli I e II

Lingua di insegnamento

ITA

2.10 Data Science e Modelli Statistici per il Trattamento dei Dati Non Strutturati

Docente: Fattore Marco

Mail: marco.fattore@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Obiettivi formativi

Introdurre gli studenti alle moderne tecniche statistiche di classificazione non-supervisionata e riduzione della dimensionalità dei dati, con particolare enfasi su:

1. L'unità della struttura concettuale del problema e il legame tra questa e le tecniche statistiche.

2. La struttura logica e matematica sottostante gli algoritmi.

3. Le differenze e le diverse caratteristiche delle tecniche e degli algoritmi, con il relativo riflesso sui criteri di utilizzo nell'analisi dei dati.

4. I limiti intrinseci degli algoritmi statistici

In sintesi, l'obiettivo è fornire agli studenti una competenza allo stato dell'arte degli strumenti di classificazione non-supervisionata, insieme ad una comprensione profonda della struttura delle tecniche e ad una capacità critica per quanto riguarda il loro utilizzo.

Contenuti sintetici

Il corso introduce e motiva il problema della classificazione non-supervisionata e della riduzione della dimensionalità, mostrandone le sue applicazioni ad ambiti concreti, richiama gli strumenti matematici di base, principalmente algebrici, e illustra le principali tecniche e i principali algoritmi di classificazione/riduzione, sia di tipo lineare che non-lineare e sia per dati numerici che non-numerici e parzialmente ordinati. La spiegazione delle tecniche è supportata da numerosi esempi su dati reali o simulati e dall'illustrazione dei codici sw che ne permettono l'implementazione.

Programma esteso

1. Il problema della classificazione non-supervisionata e della riduzione della dimensionalità dei dati: esempi di analisi di dati sociali, di dati economici e di dati provenienti da discipline umanistiche.
2. Richiami di algebra lineare: spazi vettoriali, prodotti scalari, proiezioni ortogonali, norme matriciali.
3. Tecniche di analisi lineari:
 - Decomposizione a Valori Singolari (SVD) e suo legame con l'analisi delle Componenti Principali.
 - Non-negative Matrix Factorization e suo confronto con SVD.
 - Il lemma di Johnson-Lindestrauss e la riduzione della dimensionalità a distorsione limitata: Proiezioni Casuali
 - Multidimensional Scaling.
4. Tecniche non-lineari:

2.10. DATA SCIENCE E MODELLI STATISTICI PER IL TRATTAMENTO DEI DATI NON STRUTTURATI

- Dati su varietà differenziabili: Isomap
 - Self-organizing map (SOM)
 - Entropia, divergenza di Kullback-Liebler e riduzione della dimensionalità: SNE e t-SNE
5. Dati parzialmente ordinati:
- Estrazione di ranking da dati multidimensionali
6. I limiti degli algoritmi statistici: il teorema "no-free lunch".

Prerequisiti

Non sono formalmente previsti prerequisiti, ma è necessaria una competenza di base di algebra lineare, statistica descrittiva e analisi dei dati.

Metodi didattici

Lezioni frontali ed esercitazioni e simulazioni, mediante linguaggio R, condotte dal docente

Modalità di verifica dell'apprendimento

Esame orale.

La scelta di questa modalità di verifica è dettata da:

1. La tipologia di contenuti del corso, di natura metodologica.
2. L'importanza che gli studenti acquisiscano una capacità argomentativa e di organizzazione del pensiero e siano in grado di effettuare collegamenti analogici fra le parti del programma, sollecitati dalle domande del docente.

Testi di riferimento

Geometric Structure of High-Dimensional Data and Dimensionality Reduction, Wang J. - Springer 2012.

Methods of Multivariate Analysis, Rencher A. C., Wiley 2002

Introduction to Lattices and Order (second edition), Davey B.A., Priestley H. A., CUP 2002 (capitolo 1).

Dispense e articoli forniti dal docente

Periodo di erogazione dell'insegnamento

I semestre I ciclo

Lingua di insegnamento

Italiano

2.11 Demografia

Docente: Blangiardo Gian Carlo
Mail: giancarlo.blangiardo@unimib.it
Ore: 63
Tipo di attività: Obbligatorio
Settore disciplinare: SECS-S/04

Obiettivi formativi

Il corso si propone di fornire gli strumenti di base per studiare le trasformazioni quantitative e strutturali di una popolazione e approfondire i fattori che ne determinano l'evoluzione. Particolare attenzione viene dedicata alla presentazione dei metodi di analisi dei fenomeni demografici e alle tecniche di previsione, sia degli individui che delle famiglie. Ci si propone altresì di acquisire una professionalità nell'affrontare iniziative di documentazione, analisi e divulgazione dei temi riguardanti la popolazione e le sue dinamiche.

Contenuti sintetici

Dimensione e struttura di una popolazione.
Le fonti demografiche
Misure dell'incremento
L'analisi dei fenomeni demografici
Previsioni della popolazione

Programma esteso

Introduzione: che cosa è la demografia.

Le fonti demografiche, nazionali e internazionali

Dimensione e struttura di una popolazione.

Misure dell'incremento. Principali caratteristiche strutturali e fenomeni connessi: invecchiamento demografico, carico sociale, ecc.

Componenti che determinano l'evoluzione di una popolazione.

I fenomeni di movimento della popolazione: ruolo, importanza, problemi di misurazione e di confronto. Tassi generici e specifici. Confronto fra tassi.

L'analisi dei fenomeni demografici: strumenti e concetti di base, tassi e probabilità.

La mortalità (mortalità infantile e generale, tavole di mortalità), la nuzialità (formazione e dissoluzione familiare, tipologie di famiglia), la fecondità (intensità e caratteristiche, tendenze e confronti), migrazioni interne e internazionali (misura e caratteri delle migrazioni, la presenza straniera in Italia)

Previsioni della popolazione, metodo sintetico e analitico (sesso ed età). La scelta delle ipotesi di base

Tendenze demografiche in Italia e nel Mondo: valutazioni e aspetti problematici

Prerequisiti

Nessun prerequisito formale

Metodi didattici

Lezioni frontali teoriche con approfondimenti sugli scenari demografici attuali italiani ed esercitazioni finalizzate a mostrare le procedure di calcolo dei principali indicatori demografici e l'interpretazione dei risultati ottenuti.

Modalità di verifica dell'apprendimento

Esame scritto con domande teoriche ed esercizi pratici ed eventuale prova orale su aspetti teorici e sulle dinamiche demografiche italiane. La prova d'esame è mirata a verificare la capacità di trattamento di dati statistici, la costruzione ed interpretazione di indicatori demografici e la conoscenza dei principali metodi di analisi demografica.

Testi di riferimento

G. C. Blangiardo, Elementi di Demografia, Il Mulino, Bologna 1997 (ristampa 2006).

G.C. Blangiardo (a cura di) Sussidiarietà e crisi demografica. Rapporto sulla sussidiarietà 2016/2017. Fondazione Sussidiarietà. Milano 2017

Documentazione aggiuntiva verrà resa disponibile durante il corso sulla piattaforma e-learning

Periodo di erogazione dell'insegnamento

II semestre, III e IV ciclo (periodo approssimativo fine febbraio
- metà giugno)

Lingua di insegnamento

Italiano

2.12 Demografia Sociale (Mobilità e Migrazioni)

Docente: Terzera Laura

Mail: laura.terzera@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/04

Obiettivi formativi

Il corso si propone di descrivere il fenomeno della mobilità territoriale, le sue cause e caratteristiche. Verranno individuati gli strumenti adatti per valutare l'intensità dei flussi migratori ed introdotta la metodologia per quantificare la componente irregolare. Il tema verrà approfondito tramite lo studio del caso italiano

Contenuti sintetici

- Storia delle migrazioni
- Tipologie/Teorie
- Fonti ufficiali
- Tecniche di campionamento utilizzabili per la popolazione straniera
- Stima della numerosità
- Il caso italiano

Programma esteso

- Definizioni di mobilità e migrazione
- Migrazioni nelle società moderne
- Emigrazione internazionale e interna: il caso italiano
- Cicli, fasi e tipologie
- Teorie sulle migrazioni
- Fonti
- Campionamento cattura-ricattura
- Campionamento snowball- Risposte casualizzate
- Tecnica di campionamento delle unità abitative
- Tecnica di campionamento dei centri di aggregazione
- Metodi di stima dell'ammontare degli stranieri irregolari
- Le nuove frontiere: il caso africano

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento dell'esame di Demografia.

Metodi didattici

Lezioni frontali e laboratori (SPSS). Nelle prime lezioni il fenomeno delle migrazioni viene presentato da un punto di vista multidisciplinare, attraverso la presentazione teorica degli aspetti principali desumibili dall'approccio demografico, sociologico, economico ecc. Quindi ci si concentrerà sulla presentazione degli aspetti stati-

stici/demografici del fenomeno (fonti/tecniche di campionamento/metodi di analisi).

Nelle lezioni di laboratorio attraverso l'utilizzo di open data nazionali ed internazionali e di survey a livello nazionale o regionale ogni studente verrà guidato nella compilazione di un elaborato su un aspetto specifico del fenomeno (SPSS)

Modalità di verifica dell'apprendimento

Frequentanti: l'esame consiste in una prova orale di discussione di un elaborato (pratico) su un tema concordato con la docente e sugli argomenti trattati a lezione

Non frequentanti: l'esame consiste in una prova orale al fine di poter valutare la preparazione metodologica e la conoscenza del fenomeno in esame appresa dai materiali didattici indicati

Testi di riferimento

- C. Bonifazi, L'Italia delle migrazioni, Il Mulino, 2013
- M. Ambrosini, Sociologia delle migrazioni, Il Mulino, 2005
- International Migration Report 2017, UN, New York, 2018:

<http://www.un.org/en/development/desa/population/migration/publications/mi>

Materiale messo a disposizione durante le lezioni su piattaforma e-learning (slides, esempi, rapporti, ecc)

Periodo di erogazione dell'insegnamento

I semestre II ciclo (periodo approssimativo da fine novembre a metà gennaio)

Lingua di insegnamento

Italiano

2.13 Demografia Sociale (Paesi in Via di Sviluppo)

Docente: Farina Patrizia

Mail: patrizia.farina@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/04

Obiettivi formativi

L'obiettivo del corso consiste nell'approfondimento delle conoscenze delle dinamiche demografiche nei paesi in via di sviluppo. L'attività formativa prevede un costante richiamo agli aspetti applicativi dei temi trattati attraverso l'uso di dati e materiali disponibili a livello nazionale e internazionale.

Contenuti sintetici

- La transizione demografica
- La transizione epidemiologica e di fecondità
- La domanda e l'offerta di figli
- La regolazione della fecondità
- La contraccezione e il bisogno non soddisfatto di contraccezione
- Gli obiettivi sostenibili SDGs

Programma esteso

Il corso si propone di approfondire le dinamiche demografiche in corso e di fornire gli strumenti per valutare ed interpretare le manifestazioni e le cause delle differenziazioni territoriali dei fenomeni demografici. In particolare i temi demografici sono illustrati comparando le condizioni demografiche dei paesi a sviluppo economico avanzato e quelle dei paesi poveri. Tale chiave di lettura consente di mettere pienamente in luce le similitudini e le diversità nei processi demografici, storici e contemporanei che caratterizzano le aree del mondo. L'attività formativa prevede un costante richiamo agli aspetti applicativi dei temi trattati attraverso l'uso di dati e materiali disponibili a livello nazionale e internazionale. Tale obiettivo verrà perseguito anche con il ricorso ad esercitazioni di gruppo.

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento dell'esame di demografia

Metodi didattici

Lezioni frontali

Modalità di verifica dell'apprendimento

Per i frequentanti prova scritta sui contenuti teorici del corso e prova orale sul rapporto redatto su un progetto di ricerca concordato con la docente. Per i non frequentanti l'esame è orale sui testi indicati

Testi di riferimento

Per i frequentanti il materiale didattico è fornito dalla docente sulla piattaforma e-learning I non frequentanti devono preparare i seguenti testi:

M. Livi Bacci, *Storia minima della popolazione del mondo*, Il Mulino, Bologna ultima edizione

G. Micheli *Demografie*, Mc-Graw Hill, 2011, capitoli 1,4,5,6.

E uno fra:

Giovannini E. *L'utopia sostenibile* Laterza 2018

Angeli A Salvini S. *Popolazione mondiale e sviluppo sostenibile*, il Mulino 2018

Periodo di erogazione dell'insegnamento

Secondo semestre primo ciclo

Lingua di insegnamento

Italiano

2.14 Elementi di Biostatistica

Docente: Bagnardi Vincenzo

Mail: vincenzo.bagnardi@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MED/01

Obiettivi formativi

Obiettivo del corso è fornire allo studente le basi per pianificare accuratamente un esperimento o un'osservazione in ambito biomedico, per scegliere adeguatamente i metodi di raccolta dei dati e di analisi statistica, e per interpretare correttamente i risultati ottenuti.

Contenuti sintetici

- Introduzione al corso
- Analisi di studi con risposte continue
- Analisi di studi con risposte categoriche
- Analisi di studi con dati di sopravvivenza
- Approfondimenti sulla metodologia della ricerca clinica ed epidemiologica

Programma esteso

1. Introduzione

- 1.1 Introduzione al corso: i passi della ricerca in campo biologico e medico
- 2. Metodi per l'analisi di risposte continue
 - 2.1 Test t e analisi della Varianza (ANOVA)
 - 2.2 Verifica degli assunti per l'analisi parametrica e trasformazione di variabili
 - 2.3 Metodi non parametrici
 - 2.4 Regressione lineare semplice e multipla
- 3. Metodi per l'analisi di risposte binarie
 - 3.1 Analisi delle tabelle di contingenza
 - 3.2 Regressione logistica semplice e multipla
 - 3.3 Stima della relazione dose-risposta
- 4. Metodi per l'analisi dei tempi di sopravvivenza
 - 4.1 Caratteristiche dei dati di sopravvivenza
 - 4.2 Stima non parametrica della funzione di sopravvivenza (metodo di Kaplan-Meier)
 - 4.3 Modelli di regressione per dati di sopravvivenza: il modello per rischi proporzionali di Cox
- 5. Approfondimenti
 - 5.1 Campionamento casuale e assegnazione casuale, bias, confondimento, inferenza causale
 - 5.2 Calcolo della dimensione campionaria di uno studio clinico
 - 5.3 Gli studi esplorativi
 - 5.4 Il concetto di rischio. Rischio assoluto e rischio relativo

Prerequisiti

Nessuno

Metodi didattici

Lezioni

Esercitazioni al computer con applicazioni in SAS

Modalità di verifica dell'apprendimento

Prova scritta (solo per frequentanti)

Durante il corso saranno forniti dei dati raccolti in uno studio reale.

I dati dovranno essere analizzati utilizzando tutte le metodologie ritenute adatte al raggiungimento dell'obiettivo dello studio. I risultati dell'analisi dovranno essere presentati seguendo lo stile di un lavoro originale di ricerca.

Prova orale

Nella data dell'esame si svolgerà una valutazione orale sui contenuti dei capitoli indicati del libro di testo proposto.

Testi di riferimento

Martin Bland - Statistica Medica - Apogeo

Periodo di erogazione dell'insegnamento

Semestre I, Ciclo I

Lingua di insegnamento

Italiano

2.15 Epidemiologia

Docente: Zambon Antonella

Mail: antonella.zambon@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MED/01

Obiettivi formativi

Il corso si propone di fornire le basi concettuali e gli strumenti per la pianificazione e l'interpretazione statistica di uno studio epidemiologico. Alla fine del corso lo studente deve essere in grado di impostare correttamente il piano di uno studio epidemiologico orientandosi tra diversi disegni osservazionali e di fornire un contributo statistico alla stesura di un rapporto di ricerca. Per alcuni argomenti sono previste esercitazioni pratiche condotte in ambiente SAS.

Contenuti sintetici

Panoramica dei diversi disegni osservazionali

Principali indicatori epidemiologici: misure di frequenza, misure di associazione e misure di impatto potenziale

Introduzione alle distorsioni e al confondimento e metodi per controllarli

Programma esteso

Concetto di causa in epidemiologia

Metodi di osservazione in epidemiologia

Studi ecologici; l'errore ecologico; studi di correlazione geografica; studi di correlazione temporale

Studi analitici; studi di coorte (razionale; periodo di followup e tempi di induzione-latenza); studi caso-controllo (razionale; base dello studio; scelta dei casi e dei controlli); studi ambidirezionali (caso-coorte e caso-controllo innestati nella coorte)

Misure di frequenza di malattia. Misure di associazione. Misure di impatto potenziale

Validità e precisione delle stime: distorsione da selezione, da informazione (misclassificazione differenziale e non differenziale) e da confondimento

Metodi esatti e approssimati per la stima intervallare delle misure epidemiologiche

Introduzione alla meta-analisi

Introduzione alla farmacoepidemiologia.

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento degli esami di Statistica I e Statistica medica.

Metodi didattici

Lezioni frontali alternate a lezioni pratiche su dati reali erogate in laboratorio informatico

Modalità di verifica dell'apprendimento

Modalità prova finale per frequentanti

La prova finale sarà costituita da due momenti. Nel primo momento (tendenzialmente uguale per tutti) i gruppi di lavoro presenteranno un breve seminario dell'attività pratica svolta organizzato come un articolo scientifico: introduzione al problema, materiali e metodi usati, principali risultati, discussione e conclusioni. Ogni componente del gruppo avrà circa 5 minuti a disposizione per commentare le slide che ha prodotto. Il voto attribuito al seminario sintetizzerà le capacità organizzative e di comunicazione degli studenti oltre alla chiarezza espositiva e alla correttezza delle metodologie statistiche utilizzate e delle conclusioni tratte dalle analisi. In un secondo momento ciascuno studente sosterrà un esame orale in cui verrà verificata la conoscenza degli argomenti trattati nel corso (con minore enfasi per quelli già trattati nel seminario in cui si è cimentato lo studente), la capacità di comunicare con un linguaggio tecnico appropriato e la capacità di ragionamento in particolari scenari proposti dal docente. Anche l'orale produrrà un voto. Il voto finale attribuito allo studente sarà una media del voto del seminario e della prova orale purchè entrambe siano sufficienti.

Modalità prova finale per non frequentanti

La prova finale sarà identica a quella dei frequentanti per quanto riguarda la prova orale. Il seminario sarà sostituito da un lavoro

di approfondimento individuale (tesina) su un argomento a scelta dello studente tra quelli trattati nel corso. La tesina potrà avere un taglio teorico e/o pratico. La tesina dovrà essere inviata al docente contestualmente alla prova orale. Il voto finale attribuito allo studente sarà una media del voto della tesina e della prova orale purchè entrambe siano sufficienti.

Testi di riferimento

Materiale per i frequentanti.

-A. Zambon. Lucidi dell'anno accademico in corso. Materiale scaricabile dalla piattaforma didattica online. Per alcune parti del corso verrà fornito ulteriore materiale ad integrazione dei lucidi.

Materiale per i non frequentanti.

-Materiale disponibile per i frequentanti.
-Rothman KJ. 2004. Epidemiology: an introduction. New York: Oxford University Press

Periodo di erogazione dell'insegnamento

I semestre, II ciclo (periodo approssimativo da novembre a gennaio).

Lingua di insegnamento

L'insegnamento è completamente erogato in italiano. La maggior parte del materiale su cui gli studenti dovranno lavorare per la preparazione del seminario finale (ricerca bibliografica di articoli

scientifici, siti internazionali per la disponibilità di dati aggregati, etc) e le autovalutazioni online sono in inglese

2.16 Informatica

Docente: Pescini Dario

Mail: dario.pescini@unimib.it

Ore: 68

Tipo di attività: Obbligatorio

Settore disciplinare: ING-INF/05

Obiettivi formativi

Il corso introdurrà concetti di base riguardanti: l'architettura dei computer, l'hardware, il software, i sistemi operativi, il processo di sviluppo del software e le basi di dati relazionali.

Il corso si focalizzerà sulla costruzione di interrogazioni per l'accesso ai dati e di programmi per la manipolazione dell'informazione.

Al termine di questo corso, gli studenti saranno in grado di risolvere i problemi utilizzando un linguaggio di programmazione e gli strumenti di elaborazione automatica delle informazioni.

Contenuti sintetici

- Elaborazione dell'informazione
- Hardware/Software
- Sistemi operativi
- Linguaggi di programmazione
- Basi di dati e paradigma relazionale
- Linguaggio SQL:

- Estrazione di dati sparsi su più tabelle
- Operatori di aggregazione
- Operatori matematici
- Condizioni semplici e condizioni su valori aggregati
- Interrogazioni annidate
- Programmazione: linguaggio python:
 - Variabili, istruzioni condizionali, cicli
 - Strutture dati complesse
 - File
 - Procedure e funzioni

Programma esteso

- Elaborazione delle informazioni
- Cenni di Hardware/Software
- Sistemi Operativi
- Command Line Interface e Graphic User Interface
- Cenni di File System
- Differenze tra File System Unix e File System Windows
- Paradigmi di programmazione
- Basi di Dati:
 - Paradigma client/server
 - Gestione centralizzata dei dati
 - Il paradigma relazionale
 - Chiavi primarie ed esterne
 - Cenni di algebra booleana
- Linguaggio SQL (Data Definition and Data Manipulation Language):

- Anatomia di una query SQL
- Operatori di aggregazione
- Clausola Group By
- Query Annidate
- Introduzione agli algoritmi
- Dagli algoritmi alla programmazione
- Linguaggi di programmazione (Interprete e compilatore)
- Programmazione in Python
 - Ambienti di esecuzione e sviluppo per Python
 - Struttura di un file sorgente
 - Istruzioni
 - Variabili
 - Logica booleana
 - Istruzioni condizionali
 - Input Output
 - Funzioni
 - Parametri formali e parametri attuali
 - Moduli, Package, Package Manager
 - Strutture dati complesse (tuple, liste, dizionari)
 - Strutture dati complesse e iteratori
 - File
 - Trattamento dei file di testo

Prerequisiti

Nessun prerequisito formale richiesto.

Metodi didattici

Lezione frontale

Didattica in Blended E-Learning (video, quiz per fissare i contenuti, esercizi di autovalutazione)

Esercitazioni in laboratorio

Simulazioni d'esame (per permettere allo studente di prendere confidenza con la piattaforma d'esame e la propria preparazione)

Modalità di verifica dell'apprendimento

L'appello consiste in una prova scritta all'elaboratore.

Orale facoltativo (su richiesta del docente o dello studente). L'orale può comportare sia l'aumento che la diminuzione della valutazione della prova scritta.

Testi di riferimento

Le slide, gli esercizi ed in generale tutto il materiale presentato a lezione sarà pubblicato su questo sito web.

Libri

- Atzeni, P., Ceri, S., Paraboschi, S., & Torlone, R. Basi di dati. McGraw-Hill.

<https://www.mheducation.it/9788838694455-italy-basi-di-dati-5ed-con-connect>.

- T. Gaddis, Introduzione a Python. Pearson. Maggiori informazioni

http://www.pearson.it/opera/pearson/0-6270-introduzione_a_python.

- A. Downey, J. Elkner, C. Meyers. “Pensare da informatico, Imparare con Python”, traduzione italiana di “How to Think Like a Computer Scientist”, Green Tea Press, Wellesley, Massachusetts.
EBook disponibile in rete (basta effettuare una ricerca su google).
Alcuni link:
formato pdf http://www.python.it/doc/Howtothink/HowToThink_ITA.pdf,
formato html <http://www.python.it/doc/Howtothink/Howtothink-html-it/index.htm>.
- Le slide presentate a lezione saranno rese disponibili presso questo sito
- Durante il corso potrà essere indicato del materiale aggiuntivo

Periodo di erogazione dell’insegnamento

I semestre

Lingua di insegnamento

Italiano

2.17 Laboratorio di Informatica

Docente: Boselli Roberto
Mail: roberto.boselli@unimib.it
Ore: 24
Tipo di attività: Obbligatorio
Settore disciplinare: NN

Obiettivi formativi

Il corso fornisce le basi della programmazione software, al fine di far acquisire allo studente le conoscenze e le competenze necessarie per poter utilizzare strumenti di elaborazione automatica delle informazioni. Gli argomenti proposti saranno contestualizzati attraverso esempi espressi in un linguaggio di programmazione, cioè Python.

Contenuti sintetici

Il corso è basato sui seguenti argomenti: Strumenti per la programmazione Programmazione strutturata Programmazione ad oggetti Introduzione al Scientific Computing

Programma esteso

Il programma del corso è il seguente: Riepilogo dei fondamenti della programmazione strutturata con l'utilizzo del linguaggio di programmazione Python: variabili, istruzioni condizionali, cicli,

strutture dati complesse, file. Trattamento delle codifiche del testo in informatica e con Python. Programmazione ad oggetti: oggetti, classi e metodi in Python. Trattamento delle eccezioni in Python.

Prerequisiti

Nessuna propedeuticità formale. Si consiglia di aver superato l'esame di Informatica.

Metodi didattici

Il corso prevede lezioni frontali in aula ed esercitazioni in laboratorio. Le lezioni frontali sono dedicate all'approfondimento degli argomenti teorici inerenti il corso. Le esercitazioni in laboratorio informatico sono mirate alla risoluzione di problemi inerenti la programmazione. In particolare, gli studenti devono imparare a programmare in Python, identificando le soluzioni più idonee per lo svolgimento degli esercizi presentati dal docente. Le soluzioni sono poi commentate e discusse alla fine di ciascuna esercitazione. Sono infine previste della attività di tutoraggio volte a supportare lo studente che dovesse incontrare particolari difficoltà o colmare lacune su argomenti specifici tra quelli inerenti il corso.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su una prova scritta e una eventuale prova orale. La prova scritta si svolge al computer ed è mirata

ad accertare le capacità di programmare dello studente. Sarà quindi costituita da esercizi di tipo sia applicativo che teorico. In sede di valutazione viene considerata la capacità dello studente di: risolvere il problema proposto in termini di programmazione in Python, identificare le procedure idonee alla sua soluzione e discutere in modo critico le procedure utilizzate e i risultati conseguiti. Gli studenti che superano la prova scritta sono ammessi ad un eventuale prova orale. L'orale è facoltativo (su richiesta del docente o dello studente). Gli studenti apprendono l'esito della prova scritta tramite la pagina elearning del corso stesso. La prova orale dovrà necessariamente essere sostenuta nell'appello in cui si è tenuta la prova scritta. La prova orale è mirata ad accertare la conoscenza teorica dello studente sugli argomenti del corso. Saranno quindi valutate la capacità di formalizzare in termini di programmazione le tematiche proposte in sede di esame e il rigore metodologico del loro sviluppo. Il corso non prevede un voto in trentesimi ma un'approvazione.

Testi di riferimento

A. Downey, J. Elkner, C. Meyers, "Pensare da informatico, Imparare con Python", (traduzione italiana di "How to Think Like a Computer Scientist"), Green Tea Press, Wellesley, Massachusetts, 2002. T. Gaddis, *Introduzione a Python*, Pearson, 2016. Ulteriore materiale (esercizi e dispense su argomenti specifici) è messo a disposizione sulla pagina elearning del corso.

Periodo di erogazione dell'insegnamento

I ciclo, II semestre (periodo approssimativo da febbraio a aprile)

Lingua di insegnamento

Italiano

2.18 Modelli Lineari Generalizzati in Epidemiologia e Medicina

Docente: Bellocco Rino

Mail: rino.bellocco@unimib.it

Ore: 45

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MED/01

Obiettivi formativi

L'obiettivo del corso è quello di introdurre una ampia classe di modelli statistici, nota come Modelli Lineari Generalizzati, evidenziare le caratteristiche teoriche e mostrare l'applicazione in epidemiologia e medicina

Contenuti sintetici

- 1) Review dei principi di base di probabilità e inferenza
- 2) Il modello di regressione lineare
- 3) Il modello lineare generalizzato
- 4) Il modello di regressione logistica
- 5) Il modello di regressione logistica ordinale
- 6) Il modello di regressione di Poisson

Programma esteso

2.18. MODELLI LINEARI GENERALIZZATI IN EPIDEMIOLOGIA E MEDICINA 107

- Review di probabilità e inferenza statistica:
 - Variabili aleatorie
 - Distribuzioni
 - Metodi di stima
 - Intervalli di confidenza e test di ipotesi
- Introduzione a Stata
- Il modello di regressione lineare
 - Formulazione del modello
 - Analisi della varianza e della covarianza
 - Studio dell'adattamento del modello
 - Diagnostica
- Il Modello lineare generalizzato
 - Famiglia esponenziale
 - Stima di massima verosimiglianza
 - Definizione di devianza
 - Il test del rapporto della massima verosimiglianza
- Il modello di regressione logistica
 - Dati bernoulliani e binomiali
 - Formulazione del modello: funzione logit
 - Stima, intervallo di confidenza, test d'ipotesi del coefficiente di regressione
- Interpretazione dei parametri
- Godness of Fit
- Applicazione
- Il modello di regressione Logistica Ordinal
- Formulazione del modello delle odds proporzionali
 - Stima, intervallo di confidenza, test d'ipotesi del coefficiente di regressione
- Interpretazione dei parametri

- Studio dell'adattamento del modello
- Diagnostica
- Applicazione
 - Il modello di regressione di Poisson
- Definizione di tasso d'incidenza
- Anni persona
- Definizione della funzione di link logaritmica
- Stima, intervallo di confidenza, test d'ipotesi del coefficiente di regressione
- Interpretazione dei parametri
- Studio dell'adattamento del modello
- Diagnostica

Prerequisiti

Metodi didattici

Il materiale didattico verrà presentato in classe con lezioni frontali e applicazioni utilizzando dati reali provenienti da studi epidemiologici osservazionali, dove verranno implementati i modelli formulati a lezione con l'obiettivo duplice di mostrare l'utilizzo di uno dei software più comunemente usati in biostatistica e epidemiologia (Stata) e evidenziare come interpretare i modelli che vengono introdotti durante la lezione.

Modalità di verifica dell'apprendimento

2.18. MODELLI LINEARI GENERALIZZATI IN EPIDEMIOLOGIA E MEDICINA 109

L'esame scritto si basa su una applicazione guidata in Stata che permetta di valutare le conoscenze in termine di:

1) formulazione del modello statistico, 2) applicazione del modello) interpretazione dei parametri 4) inferenza statistica in termini di intervalli di confidenza e verifica delle ipotesi 5) diagnostica e bonta' di adattamento 6) capacita' di discriminazione tra modelli competitivi.

La discussione orale dell'esame serve per meglio discutere argomenti relativi alla prova scritta non totalmente chiari. L'orale può comportare sia l'aumento che la diminuzione della valutazione della prova scritta.

Testi di riferimento

Periodo di erogazione dell'insegnamento

II Semestre, II ciclo

Lingua di insegnamento

Italiano con materiale di insegnamento in Inglese

2.19 Piano degli Esperimenti

Docente: Chiodini Paola Maddalena

Mail: paola.chiodini@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il corso si propone di fornire le basi concettuali e gli strumenti sia per la costruzione di disegni campionari di base in ambito di popolazioni finite sia per la pianificazione, l'analisi e l'interpretazione statistica di un esperimento.

Alla fine del corso lo studente deve essere in grado di sapersi orientare basilamente nell'individuazione delle unità campionarie necessarie allo svolgimento di un disegno sperimentale.

Contenuti sintetici

Definizione di un piano campionario in presenza di popolazioni finite. Analisi di dati derivanti da una sperimentazione.

Programma esteso

- Campionamento da popolazioni finite
- Campione causale semplice
- Cenni alla stima della proporzione

- Campionamento stratificato
- Campionamento a grappoli
- Cenni al campionamento a panel
- Disegno completamente randomizzato (un solo fattore)
- ANOVA ad una via; ANOVA a due o più vie
- Disegni fattoriali 2^k
- Disegni a blocchi randomizzati.

Prerequisiti

Nessun prerequisito formale richiesto

Metodi didattici

Lezioni frontali di teoria ed esempi in aula.

Modalità di verifica dell'apprendimento

È prevista una prova scritta con domande di teoria ed esercizi ed una prova orale.

Testi di riferimento

Per la parte di campionamento:

Frosini B.V., Montinaro M., Nicolini G., Il campionamento da popolazioni finite, UTET, 1999 ; Cochran W.G., Sampling Techniques, J. Wiley, New York, 1977.

Per la parte di disegno degli esperimenti:

Cochran W.G., Cox M.G., *Experimental Designs*, II ed. Wiley, New York, 1992

Montgomery, D.C., *Progettazione e analisi degli esperimenti*, McGraw-Hill, Milano, 2005

Periodo di erogazione dell'insegnamento

II Semestre, III ciclo

Lingua di insegnamento

Italiano

2.20 Popolazione, Territorio e Società I

Docente: Rimoldi Stefania Maria Lorenza

Mail: stefania.rimoldi@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/04

Obiettivi formativi

L'obiettivo del corso è completare la formazione teorica e pratica sulle tecniche di analisi dei fenomeni demografici con particolare attenzione alle specificità e ai problemi introdotti dalle variabili spaziali

Contenuti sintetici

- Strumenti per l'analisi spaziale in demografia.
- Metodi per la classificazione del territorio.
- Introduzione all'autocorrelazione spaziale.
- Studi demografici territoriali in Italia.
- Introduzione all'uso dei GIS.

Programma esteso

Strumenti per l'analisi spaziale in demografia:

- distribuzione e concentrazione;
- accessibilità;

- composizione della popolazione;
- associazione geografica;
- migrazioni;
- diversità e segregazione.

Metodi per la classificazione del territorio:

- factorial ecology;
- cluster analysis per la definizione delle aree omogenee;
- ruolo della popolazione nella pianificazione delle infrastrutture (previsione degli spostamenti urbani, posizionamento ottimale di una nuova struttura).

Introduzione all'autocorrelazione spaziale: indici Join Count, Moran e Geary

Studi demografici territoriali in Italia:

- approccio atomistico (la classificazione del territorio rurale - urbano; lettura dei processi insediativi; aree di malessere demografico);
- approccio contestuale: definizione di area metropolitana in Italia; aree di attrazione; distanza funzionale, tempi medi di primo passaggio.

Introduzione all'uso dei GIS

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento dell'esame di Demografia

Metodi didattici

Blended e-Learning: 28 ore (4 crediti) di lezioni frontali in aula + 14 ore (2 crediti) di lavoro on line.

Modalità di verifica dell'apprendimento

Frequentanti: esame orale. L'esame verte sulla conoscenza e comprensione degli argomenti teorici affrontati nel corso delle lezioni e sulla discussione delle esercitazioni pratiche svolte in autonomia dagli studenti. L'esame ha l'obiettivo di valutare la competenza teorica e la capacità di applicare gli strumenti alla trattazione di argomenti concreti di ricerca.

Non frequentanti: esame orale. L'esame verte sulla conoscenza approfondita e sulla comprensione degli argomenti affrontati durante le lezioni e sulla discussione di problemi di ricerca applicata. L'esame ha l'obiettivo di valutare la competenza teorica e la capacità di trasferire tale conoscenza sul piano pratico.

Testi di riferimento

- Plane D.A. e Rogerson P.A., *The geographical analysis of population. With application to planning and business*, John Wiley & Sons, New York, 1994.
- Ebdon D., *Statistics in Geography – Second Edition*, Blackwell Publishing, 1985
- Siegel J S., Swanson D. A., *The Methods and Material of Demography – Second Edition*, Elsevier Academic Press, London, 2004.
- Golini A., Mussino A, Savioli M, *Il malessere demografico in Italia, Studi e Ricerche* , Il Mulino, 2001.

Dispense a cura del docente.

Periodo di erogazione dell'insegnamento

1° Semestre, 1° ciclo - periodo approssimativo: inizio Ottobre -
15 Giugno

Lingua di insegnamento

Italiano

2.21 Sistemi Informativi

Docente: Mezzanzanica Mario

Mail: mario.mezzanzanica@unimib.it

Ore: 66

Tipo di attività: Obbligatorio

Settore disciplinare: ING-INF/05

Obiettivi formativi

Il Corso intende creare le necessarie conoscenze, sotto il profilo tecnico e metodologico, che consentano un approccio corretto alla progettazione di un sistema informativo, quale risorsa strategica essenziale al raggiungimento degli obiettivi di un'organizzazione aziendale.

Contenuti sintetici

Architetture applicative e tecnologiche dei sistemi informativi
Le applicazioni informatiche e l'analisi del sistema informativo
Progettazione del sistema informativo per l'analisi dati e di supporto direzionale
Data quality
Sistemi informativi e social media
Introduzione alla Social Media Analytics

Big Data e tecniche di trattamento dei dati non strutturati

Programma esteso

Architetture applicative e tecnologiche dei sistemi informativi:
- Processi di elaborazione e basi dati - Architetture distribuite, client server, di rete, internet e World Wide Web Le applicazioni informatiche e l'analisi del sistema informativo: - Il portafoglio applicativo nelle aziende industriali e di servizi - CRM Progettazione del sistema informativo per l'analisi dati e di supporto direzionale: - Progetto dei processi e dei dati - BPR - Analisi delle attività e delle informazioni - Data - warehouse e data mining Data quality Sistemi informativi e social media - Evoluzione dei sistemi informativi aziendali - Social Media Marketing Introduzione alla Social Media Analytics: - Sentiment Analysis Big Data e tecniche di trattamento dei dati non strutturati
Information extraction

Prerequisiti

Sono richieste buone capacità di apprendimento, scrittura e comunicazione orale, generale conoscenza delle principali tecnologie ed applicazioni informatiche.

Metodi didattici

Il corso è erogato in italiano e prevede lezioni frontali in aula ed esercitazioni in laboratorio. Le lezioni frontali sono dedicate all'approfondimento degli argomenti teorici inerenti il corso. Le esercitazioni in laboratorio sono mirate all'utilizzo di un software

di trattamento dei dati non strutturati, in particolare per eseguire operazioni di Text Mining su dati scaricati da diverse fonti Web.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su una prova scritta e una eventuale prova orale su richiesta dello studente. La prova scritta si svolge al computer ed è composta da domande aperte e chiuse e risposta multipla su tutti gli argomenti del corso. In sede di valutazione viene considerata la capacità dello studente di rispondere a quesiti specifici facendo riferimento agli aspetti teorici e pratici (mediante esempi) connessi all'argomento richiesto. La prova scritta è comune sia per gli studenti frequentanti sia per i non frequentanti. L'eventuale prova orale è mirata ad accertare la conoscenza teorica dello studente sugli argomenti del corso. Potranno quindi essere valutate le capacità di ragionare e approfondire le tematiche proposte in sede di esame e il rigore metodologico del loro sviluppo.

Testi di riferimento

G. Bracchi, C. Francalanci, G. Motta. Sistemi informativi d'impresa. McGraw-Hill, 2010. V. Cosenza, "Social media ROI", Apogeo, 2012.

Nel corso delle lezioni sarà indicato dai docenti ulteriore materiale (slide, articoli...).

Periodo di erogazione dell'insegnamento

Il corso viene erogato nel secondo ciclo del primo semestre e nel primo ciclo del secondo semestre.

Lingua di insegnamento

Italiano

2.22 Statistica I

Docente: Chiodini Paola Maddalena

Mail: paola.chiodini@unimib.it

Ore: 47

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il corso di Statistica I, in quanto primo corso di statistica nell'ambito del CdL di Statistica e Gestione delle Informazioni, ha come obiettivo quello di introdurre gli studenti alle conoscenze ed alle competenze teorico-pratiche iniziali della statistica che costituiscono le basi del corso di laurea medesimo. Gli studenti alla fine del corso dovranno essere in grado di saper riconoscere la natura delle variabili statistiche sapendole estrarre da un data base. Dovranno inoltre saper rappresentare graficamente in modo opportuno e saper sintetizzare attraverso indicatori adeguati le informazioni che possono essere sia di tipo univariato che bivariato. Si richiederà che lo studente inizi ad esprimersi in modo chiaro e con proprietà di linguaggio sapendo altresì interagire anche con figure professionali non necessariamente preparate sulla disciplina statistica.

Contenuti sintetici

Costruzione e classificazione di dati statistici. Principali tecniche della statistica univariata e bivariata.

Programma esteso

- La statistica come scienza
- Popolazione e unità statistiche
- Variabili statistiche e scale di misura
- Rappresentazioni grafiche
- Distribuzioni di frequenza univariate
- Indici di posizione e medie
- Variabilità e sua misura
- Standardizzazione degli indici
- Studio della forma di una distribuzione (asimmetria e curtosi)
- Distribuzioni di frequenza bivariate
- Connessione e dipendenza in media
- Correlazione.

Prerequisiti

Nessun prerequisito formale richiesto.

Metodi didattici

Lezioni frontali di teoria ed esempi in aula. Esercitazioni in aula con svolgimento di alcuni esercizi alla lavagna. Inoltre l'inse-

gnamento è affiancato da attività di tutoraggio in cui si svolgono ulteriori attività pratiche e correzioni agli esercizi svolti a casa.

Modalità di verifica dell'apprendimento

È prevista una prova scritta con domande di teoria ed esercizi ed una prova orale.

Testi di riferimento

G. Leti, L. Cerbara, Elementi di statistica descrittiva, Il Mulino, Bologna 2009

L. Santamaria, Statistica descrittiva – Applicazioni economiche e aziendali, Vita e Pensiero, Milano 2006

Zanella, Elementi di statistica descrittiva, CUSL, Milano 2000

M. Zenga, Lezioni di statistica descrittiva, G. Giappichelli, Torino 2007

Periodo di erogazione dell'insegnamento

I Semestre, I ciclo

Lingua di insegnamento

Italiano

2.23 Statistica I - Complementi

Docente: Chiodini Paola Maddalena

Mail: paola.chiodini@unimib.it

Ore: 47

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il corso si propone quale naturale proseguimento del corso di Statistica I. Vengono ripresi alcuni concetti relativi alle distribuzioni bivariate approfondendo gli aspetti legati alla definizione e verifica della bontà di adattamento di modelli matematici, quali funzioni polinomiali o modelli linearizzabili, in un contesto descrittivo tali da permettere la definizione del legame funzionale esistente fra due (o più) caratteri. Per la stima dei coefficienti dei modelli si ricorre al metodo dei minimi quadrati.

Contenuti sintetici

Principali tecniche della statistica bivariata nell'ambito della regressione polinomiale.

Programma esteso

- Regressione polinomiale
- Metodo dei minimi quadrati

- Adattamento del modello ai dati
- Residui di interpolazione
- Indice di miglioramento
- Piano di regressione
- Coefficiente di correlazione parziale

Prerequisiti

Il corso di Statistica I è propedeutico al corso di Statistica I - Complementi.

Il corso richiede la conoscenza di strumenti di analisi matematica quali derivate.

Metodi didattici

Lezioni frontali di teoria ed esempi in aula. Esercitazioni in aula con svolgimento di alcuni esercizi alla lavagna.

Modalità di verifica dell'apprendimento

È prevista una prova scritta con domande di teoria ed esercizi ed una prova orale.

Testi di riferimento

G. Leti, L. Cerbara, Elementi di statistica descrittiva, Il Mulino, Bologna 2009

L. Santamaria, Statistica descrittiva – Applicazioni economiche e aziendali, Vita e Pensiero, Milano 2006

Zanella, Elementi di statistica descrittiva, CUSL, Milano 2000

M. Fraire, A. Rizzi, Esercizi di statistica, Carocci Editore, Urbino 2012

Periodo di erogazione dell'insegnamento

I Semestre, I ciclo

Lingua di insegnamento

Italiano

2.24 Statistica II

Docente: Berta Paolo

Mail: paolo.bera@unimib.it

Ore: 94

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il Corso intende fornire agli studenti un'adeguata padronanza delle metodologie di inferenza statistica per l'analisi di fenomeni univariati rappresentabili tramite modelli stocastici.

Alla fine del corso, lo studente sarà in grado di:

- 1) applicare le tecniche per la stima puntuale e intervallare dei parametri della distribuzione di una variabile casuale
- 2) costruire test statistici idonei a verificare ipotesi sulla distribuzione di una variabile casuale normale e individuare approssimazioni adeguate nel caso di variabili casuali qualsiasi
- 3) impostare alcuni piani di campionamento ricorrenti in molti contesti applicativi.

Contenuti sintetici

Il corso è suddiviso in macro argomenti che vanno dalla definizione delle variabili casuali all'inferenza basata sulla verosimiglianza. E' prevista una parte del corso finalizzata all'introduzione

ai più comuni piani di campionamento probabilistico nel contesto di popolazioni finite secondo l'approccio classico.

Programma esteso

Il corso si articola in macro argomenti:

La nozione di campione e lo spazio campionario. La stima puntuale. Proprietà degli stimatori: correttezza, consistenza, efficienza assoluta e relativa. Il teorema di Fréchet-Rao-Cramér. L'errore quadratico medio. Metodi di stima: il metodo della massima verosimiglianza e il metodo dei momenti.

Stima intervallare e metodi per la determinazione dell'intervallo di confidenza. La quantità pivotale. La verifica statistica delle ipotesi. I test di significatività. I principali test statistici: il test Z, il test T, il test chi-quadrato, il test F. Le basi della teoria di Neyman-Pearson. Errore di prima e di seconda specie. Il test più potente e il lemma di Neyman-Pearson. I test uniformemente più potenti. I test basati sul rapporto di verosimiglianza. Campionamento da popolazioni finite. Stima del totale, della media e della varianza di una variabile continua. Stima della frequenza relativa di una variabile binaria. Il campionamento casuale semplice. Il campionamento stratificato. Determinazione della numerosità campionaria.

Prerequisiti

Il corso prevede, quali corsi propedeutici, Statistica I, Calcolo delle Probabilità e Analisi Matematica I.

Metodi didattici

Il corso è erogato in italiano e prevede lezioni frontali ed esercitazioni in aula.

Le lezioni sono mirate all'approfondimento delle conoscenze teoriche dello studente sugli argomenti del corso. In questa sede viene dato ampio spazio alla formalizzazione e alla derivazione dei concetti probabilistici e matematici rilevanti per le tematiche considerate.

Le esercitazioni sono mirate a potenziare le capacità di problem solving dello studente. In questa sede è quindi dato spazio alla formalizzazione di problemi presenti in situazioni reali in termini di inferenza statistica, all'identificazione delle procedure idonee per la loro soluzione e alla discussione critica delle procedure utilizzate e dei risultati conseguiti.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su una prova scritta.

La prova scritta è mirata ad accertare le capacità di problem solving dello studente. Sarà quindi costituita da esercizi di tipo sia applicativo che teorico. In sede di valutazione viene considerata la capacità dello studente di: formalizzare il problema proposto in termini di inferenza statistica, identificare le procedure idonee alla sua soluzione e discutere in modo critico le procedure utilizzate e i risultati conseguiti.

Durante la prova scritta **non è ammesso** l'uso di testi o altro materiale con l'esclusione delle tavole della variabili. Durante la

prova non è ammesso l'uso del cellulare.

A chi ne faccia richiesta è consentito di sostenere la prova d'esame in forma orale. La prova orale è mirata ad accertare la conoscenza teorica dello studente sugli argomenti del corso. Saranno quindi valutate la capacità di formalizzare in termini statistico-probabilistici le tematiche proposte in sede di esame e il rigore metodologico del loro sviluppo.

Testi di riferimento

Cicchitelli G., D'Urso P., Minozzo M. 2018. Statistica: Principi e Metodi. Terza edizione. Pearson Italia, Milano-Torino

Ulteriore materiale (esercizi e dispense su argomenti specifici) sarà messo a disposizione sulla pagina e-learning del corso

Periodo di erogazione dell'insegnamento

Il corso viene erogato nel primo semestre.

Lingua di insegnamento

Italiano

2.25 Statistica Medica

Docente: Corrao Giovanni

Mail: giovanni.corrao@unimib.it

Ore: 44

Tipo di attività: Obbligatorio

Settore disciplinare: MED/01

Obiettivi formativi

Il corso si propone di fornire gli strumenti concettuali per capire le peculiarità del metodo statistico nell'area della ricerca biomedica. Alla fine del corso lo studente sarà essere in grado di i) impostare il calcolo di una misura di frequenza della malattia e dei suoi possibili esiti; ii) impostare il calcolo di una misura di associazione tra un'esposizione ed un esito; iii) applicare una procedura di standardizzazione diretta e indiretta; iv) riconoscere le differenze tra studi sperimentali ed osservazionali.

Contenuti sintetici

1. Ruolo delle scienze statistiche nella ricerca e nella pratica clinica
2. Frequenza della malattia e dei suoi possibili esiti
3. La ricerca scientifica tra sperimentazione ed osservazione
4. Studi clinici controllati e randomizzati
5. Studi epidemiologici osservazionali

Programma esteso

Ruolo delle scienze statistiche nella ricerca e nella pratica clinica

- Storia naturale della malattia
- Approccio clinico alla malattia
- Approccio di sanità pubblica alla malattia
- Approccio statistico alla malattia

Frequenza della malattia e dei suoi possibili esiti

- Casi incidenti e prevalenti
- Coorte fissa e popolazione dinamica
- Tassi e proporzioni
- Standardizzazione diretta e indiretta
- Analisi per coorti
- Mortalità nel periodo perinatale

La ricerca scientifica tra sperimentazione ed osservazione

- Concetto di causa
- Ricerca sperimentale
- Ricerca osservazionale

Studi clinici controllati e randomizzati

- Fasi della sperimentazione clinica
- Disegni entro pazienti
- Disegni tra pazienti

Studi epidemiologici osservazionali

- Studi di correlazione
- Studi di coorte
- Studi caso-controllo

Prerequisiti

Nessun prerequisito formale richiesto

Metodi didattici

Lezioni frontali ed esercitazioni

Modalità di verifica dell'apprendimento

L'esame consiste in una prova scritta (composta da esercizi pratici) e da una prova orale facoltativa (su richiesta del docente o dello studente).

Testi di riferimento

Sulla piattaforma e-learning del corso sarà disponibile il materiale didattico del corso (diapositive delle lezioni ed esercitazioni).

Periodo di erogazione dell'insegnamento

Il semestre, II ciclo (periodo approssimativo da maggio a metà giugno)

Lingua di insegnamento

Italiano

2.26 Statistica Sociale

Docente: Terzera Laura

Mail: laura.terzera@unimib.it

Ore: 63

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/04

Obiettivi formativi

Il corso si propone di fornire gli strumenti per l'analisi dei fenomeni sociali attraverso la presentazione delle modalità di conduzioni di indagini sul campo e costruzione di indicatori sintetici e di scale di atteggiamenti.

Contenuti sintetici

Strumenti per la ricerca sociale

- Questionario: struttura e caratteristiche
- Scale di atteggiamento
- Indicatori sociali

Programma esteso

Prima parte

- Indagini campionarie, implementazione del disegno complessivo .

- Misure e modalità di raccolta dei dati. Struttura di un questionario. Domande e categorie di risposte.
- Aspetti psicologici e fraseggio
- Domande delicate: risposte casualizzate tecnica di Warner e Simmons
- Effetti risposta; effetti d'ordine
- La tecnica delle scale: Scala di Bogardus, scalogramma di Guttman, Scala di Thurstone, scala di Likert, Differenziale semantico di Osgood, cenni altre tipologie di scale

Seconda parte

- Informazioni di base sugli indicatori sociali. Gli Indici di Sviluppo Umano e misure di Gap.
- Rassegna dei principali siti nazionali e internazionali.
- Simulazione di un percorso di ricerca, scelta di una area tematica: fonti statistiche analisi di trend
- Dai dati grezzi alla costruzione degli indicatori, gestione missing values, costruzione indicatori e definizione pesi
- Verifiche della adeguatezza della batteria degli indicatori
- Elementi per un'analisi della qualità dei dati e di un rapporto di ricerca

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento dell'esame di Statistica I

Metodi didattici

Lezioni frontali e pratiche. Le lezioni frontali affrontano da un punto di vista teorico il programma sopra citato, nelle lezioni pratiche attraverso l'utilizzo di Open-Data nazionali ed internazionali si realizzano i diversi passi di un percorso di ricerca: dall'acquisizione dei dati, la pulizia della database, costruzione indicatori e scale, analisi elementari dei risultati. Dati campionari verranno utilizzati per la costruzione di scale di atteggiamento. Verrà introdotta l'alfabetizzazione a SPSS

L'insegnamento è affiancato in parte da attività di tutoraggio nelle lezioni pratiche

Modalità di verifica dell'apprendimento

Frequentanti: La prova d'esame consiste in uno scritto parziale (cioè riguardante solo la prima parte del programma, il questionario) e contenente sia esercizi sia domande teoriche e due elaborati pratici (il cui contenuto verrà esposto a lezione) da consegnare. La prova orale è facoltativa, riguarda tutto il programma e vi possono accedere coloro che hanno ottenuto almeno la sufficienza nelle prove parziali (scritto+elaborati).

Non frequentanti: La prova d'esame consiste in uno scritto (esercizi e domande teoriche) e in un orale (domande teoriche)

Testi di riferimento

· Vanda Zammuner (1998), Tecniche dell'intervista e del questionario, il Mulino, Bologna

- Enrica Aureli Cutillo (2002) Lezioni di statistica sociale. Fonti, strumenti e metodi.Ed. CISU
- Handbook on Constructing Composite Indicators METHODOLOGY AND USER GUIDE, OECD, European Commission, 2008
<https://www.oecd.org/std/42495745.pdf>
- Letture integrative e documentazione di lavoro saranno distribuite durante il corso sulla piattaforma e-learning.

Periodo di erogazione dell'insegnamento

annualità, I e II semestre, II e III ciclo (periodo approssimativo da fine novembre a metà aprile)

Lingua di insegnamento

Italiano

2.27 Statistica Spaziale e Ambientale

Docente: Borgoni Riccardo

Mail: riccardo.borgoni@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Obiettivi formativi

Il corso intende fornire un'introduzione ai metodi statistici per l'analisi di fenomeni il cui valore varia nello spazio e ad alcune problematiche connesse con l'analisi statistica dei dati ambientali. Alla fine del corso lo studente ha acquisito la capacità di manipolare dati georeferenziati in termini della loro analisi tramite software per l'elaborazione statistica dei dati e ha maturato competenze su specifici aspetti connessi con i dati provenienti da survey ambientali.

Contenuti sintetici

Introduzione all'analisi dei dati spaziali; georeferenziazione; elementi di cartografia; metodi per l'interpolazione di superfici continue o relative ad un supporto spaziale discreto. Stimatori bayesiani empirici per tassi. Stima del valore di fondo naturale di una sostanza nel suolo o nell'acqua. Spline e modelli di regressione semiparametrica per l'analisi dei dati ambientali. Applicativi in ambiente R per l'analisi di dati spaziali e ambientali.

Programma esteso

a) Georeferenziazione. Sistemi di riferimento e proiezioni cartografiche. Funzionalità dei gis, dati raster e vettoriali. Lisciamen-
to di una mappa nel caso di dati continui e reticolari, metodi IWD,
poligoni di Voronoi, medie mobili spaziali, stimatori bayesiani em-
pirici per tassi.

b) Stima del valore di fondo nel suolo o nell'acqua

c) WLS, B-spline e i modelli additivi

d) Laboratorio in R.

Prerequisiti

Il corso non ha propedeuticità. Sono consigliati i contenuti del
corso di Analisi Statistica Multivariata, Statistica II e Informatica.

Metodi didattici

Lezioni frontali e sessioni di laboratorio informatico in ambien-
te R mirate all'analisi di dati contenuti in archivi amministrativi o
raccolti in survey ambientali.

Modalità di verifica dell'apprendimento

L'esame finale consiste di due parti: una prova in laboratorio
informatico e una prova orale.

Nella prova di laboratorio informatico, svolta in ambiente R,
allo studente sarà richiesto di eseguire un'elaborazione di dati am-

bientali e/o spaziali utilizzando codici e funzioni presentati e discussi durante le sessioni di laboratorio informatico del corso. La prova mira a verificare la capacità acquisita dallo studente nel rispondere a concrete esigenze che, nell'ambito della tematica dei dati spaziali e ambientali, possono emergere in contesto lavorativo o di ricerca. Saranno quindi valutate le capacità di analisi dei dati, di sintesi e interpretazione dei risultati, nonché la competenza acquisita nell'uso del software R.

Durante la prova **non è ammesso** l'uso di testi o altro materiale con l'esclusione dei codici che verranno messi a disposizione dal docente all'inizio della prova.

Durante la prova non è ammesso l'uso del cellulare.

La prova orale è mirata ad accertare la conoscenza teorica dello studente sugli argomenti del corso. Saranno quindi valutate la capacità di formalizzare in termini statistico-probabilistici le tematiche proposte in sede di esame e il rigore metodologico del loro sviluppo.

Il punteggio finale è costituito da una media dei voti ottenuti nelle due parti.

Testi di riferimento

Saranno messe a disposizione degli studenti dispense e codici software tramite la pagina elearning del corso.

Periodo di erogazione dell'insegnamento

Primo ciclo del secondo semestre.

Lingua di insegnamento

Italiano.

Capitolo 3

Courses

3.1 Algebra Lineare

Docente: Bandini Elena

Mail: elena.bandini@unimib.it

Ore: 42

Tipo di attività: Obbligatorio

Settore disciplinare: MAT/02

Learning objectives

This course aims to provide the foundations of linear algebra that will form the basis of the courses of Probability and Multivariate Statistical Analysis.

Contents

Vector spaces and linear maps; matrices and diagonalization procedure; determinants; similarity of matrices.

Detailed program

\mathbb{R}^n and its vector subspaces. Vector spaces. Systems of generators and bases. Dimension. Standard scalar product and Cauchy-Schwartz inequality. Norm and triangular inequality. Angles. Elements of analytic geometry. Linear maps. Rank plus nullity theorem. Study of linear maps by matrices and linear systems. Orthonormal bases. Orthogonal projections. Determinant and its computation by row operations. Invertible matrices. Eigenvalues and eigenvectors. Diagonalization. Orthogonal and symmetric matrices. Spectral Theorem.

Prerequisites

No prerequisites are required.

Teaching methods

Lectures.

Assessment methods

Written exam based on exercises and theoretical questions aimed at evaluating the problem-solving skills and the ability to apply the theoretical results in order to solve specific problems.

Oral exam (upon student's or teacher's request).

Textbooks and Reading Materials

T.M. Apostol, *Calcolo*, volume secondo (Geometria), Bollati Boringhieri, 2003.

Semester

II semester, III and IV cycle (approximately from March to June).

Teaching language

Italian.

3.2 Analisi Matematica I

Docente: Travaglini Giancarlo

Mail: giancarlo.travaglini@unimib.it

Ore: 63

Tipo di attività: Obbligatorio

Settore disciplinare: MAT/05

Learning objectives

This course mainly aims at providing a rigorous introduction to differential and integral calculus for functions of one variable.

Contents

Mathematical and everyday language, Studying a Math Book, Sequences and series, Differential calculus of one variable, Functions, Taylor series, Integral calculus of one variable, Integral functions and distribution functions.

Detailed program

Mathematical and everyday language. Quantifiers and conjunctions. Structure of a mathematical sentence. The language of sets. Negating a mathematical sentence. Contrapositives and converses. Unknowns in mathematical statements. Indices in sums and

in set theoretic operations. Studying a math book. Definitions and examples. Studying a proof: checking the steps, looking at a few examples, applying the argument to similar problems.

Real numbers: metric and arithmetic properties. Powers with real exponent. Equations and inequalities. Supremum. Sequences and limits. Monotone sequences. Indeterminate forms. The number "e". Computing some limits. Series. Geometric series.

Limits and continuity of functions. Composition of functions. Derivative of a function. Using differential calculus to draw the graph of a function. The mean value theorem. Higher order derivatives. Convexity. Taylor expansions and Taylor series. The exponential series.

Riemann integral. The fundamental theorem of Calculus. Improper integrals. Numerical series and improper integrals. The Gamma function. Integral functions and their graphs. Distribution functions.

Prerequisites

+ Algebra: inequalities (I and II degree, irrational, exponential, logarithmic). + Euclidean Geometry. + Analytic Geometry. + Trigonometry (trigonometric functions, equations and inequalities). + Elementary properties of integer and rational numbers.

Teaching methods

Lessons, tutoring, homeworks.

Assessment methods

Written and oral exam.
No midterm exam.

Textbooks and Reading Materials

M. Bramanti, C. Pagani, S. Salsa, *Analisi Matematica I*, Zanichelli.

M. Bramanti, G. Travaglini, *Matematica. Questione di Metodo*, Zanichelli.

M. Bramanti, *Precalculus*, Progetto Leonardo, Esculapio.

M. Bramanti, *Esercizi di Calcolo Infinitesimale e Algebra Lineare*, Seconda Edizione, Progetto Leonardo, Esculapio.

M. Boella, *Analisi matematica e algebra lineare*, vol.1, Pearson.

Notes, videos and hundreds of solved tasks at <http://elearning.unimib.it/enrol/index.php?id=19508>

Semester

First semester (from September to January).

Teaching language

Italian

3.3 Analisi Matematica II

Docente: Travaglini Giancarlo

Mail: giancarlo.travaglini@unimib.it

Ore: 42

Tipo di attività: Obbligatorio

Settore disciplinare: MAT/05

Learning objectives

This course mainly aims at providing a rigorous introduction to Fourier series and to the differential and integral calculus for functions of d variables.

Contents

Fourier series.

Differential calculus on R^d .

Integral calculus on R^d .

Detailed program

Fourier series, Benford's law.

Differential calculus on R^d .

Partial derivatives and differentiability.

Maxima and minima (with or without constraint), Lagrange multipliers.

Taylor expansion, Hessian matrix. Linear regression.

Convex functions
Integral calculus on R^d .
Changes of variables.
Monte Carlo method.
Improper integrals.
Integration of radial functions on R^d .

Prerequisites

Differential and integral calculus in one variable. Linear Algebra.

Teaching methods

Lessons, tutoring, homeworks.

Assessment methods

Written and oral exam.
No midterm exam.

Textbooks and Reading Materials

M. Bramanti, C. Pagani, S. Salsa, *Analisi Matematica 2*, Zanichelli.
M. Bramanti, *Esercizi di Calcolo Infinitesimale e Algebra Lineare*, Seconda Edizione, Progetto Leonardo, Esculapio.
M. Boella, *Analisi Matematica 2*, Pearson.

Notes and hundreds of solved tasks at <http://elearning.unimib.it/course/view.php?id=19499>

Semester

First semester (from September to November).

Teaching language

Italian

3.4 Analisi Statistica Multivariata

3.4.1 Modulo: Analisi Esplorativa

Docente: Solaro Nadia

Mail: solaro.nadia@unimib.it

Ore: 49

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

The Exploratory Analysis module introduces the main descriptive statistical methods addressed to study two or more variables jointly observed on a set of statistical units. These methods aim at exploring multidimensional data to detect underlying structures and reduce their dimensionality, however preserving the main observed features. From a practical point of view, data analysis is carried out through the R software (RStudio environment).

Contents

Introduction to multivariate statistical analysis. Quantitative, qualitative and mixed data matrices. Graphical representations of multidimensional data. Cluster Analysis: Hierarchical and non-hierarchical clustering methods. Principal component analysis. Exploratory factor analysis. Integrated use of exploratory multivariate methods. Applications to real data with software R (RStudio environment).

Detailed program

- Introduction to the multivariate statistical analysis: French and Anglo-Saxon schools, classification of multivariate analysis methods
- Quantitative, qualitative and mixed-type data matrices. Main syntheses and transformations. Data representation, individual space and variable space. Dissimilarities and distances between units, distances between variables
- Cluster Analysis: Hierarchical and non-hierarchical clustering methods, goodness of classification, applications to quantitative and qualitative variables
- Principal component analysis: Extraction of the principal components, stopping criteria, evaluation of the reproduced variability, interpretation of the principal components, applications
- Exploratory Factor Analysis: Factor model, study of correlations, factor extraction methods, factor rotation, factor scores, applications
- Integrated use of exploratory multivariate techniques
- Analyses of empirical cases with RStudio

Prerequisites

Passing of preliminary examinations of Calculus, Linear Algebra, Probability, Statistics I

Teaching methods

Theoretical lectures in the classroom and practical exercises in the statistical-informatics laboratory with R software (RStudio environment)

Assessment methods

The exam consists of a written test (total duration: 2 hours) with three questions (divided into several points) that deal with both the theoretical and applicative aspects of the methodologies covered in the course. A fourth question is optional and concerns programming with R software. The theoretical questions regard the methodological aspects of the topics covered in the course. The practical questions involve both numerical exercises and commenting on parts of R outputs.

The oral exam is optional and covers both theoretical and practical topics. Access to the oral test is subject to passing the written test with a mark of at least 18/30.

Given the abundance of teaching material uploaded on the e-learning platform of the course, no distinction is made between exams for attending students and exams for non-attending students.

Textbooks and Reading Materials

- Teaching material uploaded on the e-learning website of the course

- Frosini, B.V. (2014). Complementi di analisi statistica multivariata, EDUCatt, Milano
- Zani, S., Cerioli, A. (2007). Analisi dei dati e data mining per le decisioni aziendali, Giuffrè Editore, Milano
- Bolasco, S. (1999). Analisi multidimensionale dei dati: strategie e criteri di interpretazione, Carocci, Roma
- Dillon, W.R., Goldstein, M. (1984). Multivariate Analysis, J. Wiley, New York
- Everitt, B.S., Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R, Springer, Berlin

Semester

First semester, second period

Teaching language

Italian

3.4.2 Modulo: Modelli Statistici

Docente: Pennoni Fulvia

Mail: fulvia.pennoni@unimib.it

Ore: 57

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

The course aims to provide students with the technical background on the multiple linear regression model.

Some aspects related to computations using the matrix algebra are considered.

Theory and practical applications on real and simulated data are jointly explained to support the student with a deep practical knowledge.

During the course discussion on the techniques and results is encouraged.

Contents

During the course the following main issues are raised:

The student is introduced to the graphical examination of the data and to the use of linear total and partial correlation coefficients to inspect the linear relations among the variables.

The multiple regression model is introduced according to its basic assumptions. The ordinary least square estimation method

is explained and the main basic properties of the estimators are illustrated.

The bivariate and multivariate Gauss distributions are introduced and their properties are illustrated also through applicative examples. Predictions are considered as well as information criteria to evaluate the model and to make a selection of the explicative variables.

Model diagnostics tools for checking model assumptions and unusual observations are explained.

Detailed program

During the course the knowledge of the student based on univariate distributions is extended to include the bivariate and multivariate Gaussian distributions. The variance and covariance matrix, the correlation and partial correlation matrices are introduced as well as the contours of the Bivariate Gauss distribution. Random realizations are drawn from these distributions and they are illustrated by means of the scatterplots in two and three dimensions.

The multiple linear regression model is introduced also through the matrix notation. The deviance decomposition is explained. For the model with two explicative variables the estimation of the model parameters is carried out by considering the determinant's rule applied to the variance covariance matrix. The improvement related to the reduction of the residual variance from the linear to the multiple linear regression is considered.

The properties of the estimators are illustrated according with the classical assumptions on the model. Inference is introduced for each single parameter and for couples of parameters jointly.

Many diagnostic tools are proposed to evaluate the model's residuals and some criteria for model selection such as the Bayesian Information Criterion are introduced. The way to forecast a new observation is illustrated.

Some of the main extensions of the basic linear regression model presented during the course are the following: i) maximum likelihood estimation method; ii) transformation of the variables; iii) categorical covariates; iv) models with some orders of interactions between covariates; v) categorical response variable mainly a count variable.

Some amount of time is devoted to explain the theory by imparting some flavor of applications on real data collected on different fields. Applications will be developed within the statistical environment R, RStudio, RMarkdown and by the SAS software.

Prerequisites

The student need to have done with success the examination of the following courses: Statistics I, Mathematics, Linear Algebra and Probability

Teaching methods

Lectures are related to the theoretical aspects. All the lectures are held in the lab. During the lectures many practical examples

based on real and simulated multivariate data referred to different contexts (e.g. Biology, Astronomy, Economics etc...) are proposed to the students to be solved with R, RStudio, RMarkdown, and SAS. The student is also encouraged to develop the cooperative learning in order to interact each other to finalize the required steps of the analysis, to develop the analyses and to report in a written form the results by adding also critical comments.

Assessment methods

The exam is in the lab. The first part is without the learning material and it concerns three open questions on the theoretical part. The student should provide a written answer to the questions. In the second part, to be performed at the computer, the student has to show to be able to apply the multiple linear regression model as well as the other methods to real and simulated data by using R and the RStudio environment as well as to produce the document in RMarkdown or SAS. S/he has to provide explanations concerning the results from the point of view of the statistical thinking and of the applied context.

The first part weights $1/3$ of the total score.

The oral examination is not compulsory. It can be required by the student with a score at the written exam of at least $18/30$. The results of the written examination are provided in the elearning page and the student has to attend the fixed meeting to see the spreadsheet.

Textbooks and Reading Materials

Slides and all the supporting material which includes R scripts, SAS code, exercises, solutions and data are downloadable from the webpage of the course of the e-learning website of the university.

The main texts are the following:

Faraway, J. J. (2014). *Linear models in R*, Second Edition, Chapman & Hall, CRC Press.

Johnson, R. A., and Wichern, D. W. (2002). *Applied multivariate statistical analysis*, Pearson Education International, Prentice Hall.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*, New York, Springer.

Nolan, D., & Lang, D. T. (2015). *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*. Chapman & Hall, CRC Press.

SAS/STAT 12.1. *User's guide*, SAS Institute, 2012.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Semester

II Semester, III cycle: from February to April 2018. Four days a week.

Teaching language

The teaching language is Italian.

Erasmus students will be provided with the teaching material in English. They can carry out the exam in English.

3.5 Analisi dei Dati

Docente: Vittadini Giorgio

Mail: giorgio.vittadini@unimib.it

Ore: 47

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

The course illustrates the most popular data analysis techniques both for quantitative and qualitative data. Lectures will be alternated to lab activities in order to allow the students to learn statistical packages and apply to real-world data the techniques introduced in the lectures.

Contents

Canonical Correlation Analysis

Discriminant Analysis

Correspondence Analysis

Multidimensional Scaling

Detailed program

Introduction to data analysis

A Canonical Correlation Analysis

1 Purpose

- 2 How to derive the 1° canonical variable and subsequent canonical variates
- 3 Interpretation of the results
- 4 Examples
- 5 Examples in SAS
- B Discriminant Analysis
 - 1 Purpose
 - 2 Data
 - 3 How to derive the 1° discriminant variable and subsequent discriminant variables
 - 4 Interpretation of the results
 - 5 Alternative way to derive discriminating variables
 - 6 Examples
 - 7 Examples in SAS
- C Correspondence Analysis
 - 1 Purpose
 - 2 Row and column profiles
 - 3 Correspondence Analysis as a special case of Canonical Correlation and Principal Component Analysis
 - 4 Interpretation of the results: χ^2 decomposition
 - 5 Interpretation of other results
 - 6 Multiple Correspondence Analysis
 - 7 Examples
 - 8 Examples in SAS
- D Multidimensional scaling
 - 1 Metric Multidimensional Scaling with Euclidean distances
 - 2 Torgerson Theorem
 - 3 Interpretation of the results
 - 4 Examples

- 5 Examples in SAS
- 6 Multidimensional scaling with dissimilarity metrics
- 7 Multidimensional scaling with ordinal dissimilarity
- 8 Examples
- 9 Examples in SAS

Prerequisites

This course should be taken after passing the exam of Multivariate Statistical Analysis.

Teaching methods

The course consists of lectures and practical exercises in a lab with SAS software. The lectures will offer a theoretical background to the methods proposed, while the practical examples will be an application to real data.

Assessment methods

Written exam, lab test (to be taken at the time of the written exam) and oral exam. The written exam consists in two theoretical questions about the methods introduced in the course, while the lab test is an exercise with SAS software, similar to one of the examples proposed in the lab. During the oral exam the students will be questioned about their written exam.

Textbooks and Reading Materials

Slides and reading material edited by the professor.

Johnson R. A., Wichern D. W., Applied Multivariate Statistical Analysis, Pearson, 2007.

Morrison D.F., Multivariate Statistical Methods, McGraw-Hill, 2004 (4° ed.).

Semester

2° semester 2° period

Teaching language

Italian

3.6 Basi di Dati

Docente: Mezzanzanica Mario
Mail: mario.mezzanzanica@unimib.it
CFU: 6
Ore: 47
Tipo di attività: Obbligatorio
Settore disciplinare: ING-INF/05

Learning objectives

The course aims at providing competences to understand the model of a relational database for supporting decision making (Database and Data Warehouse) as a principal asset to achieve the objectives of a business organization, focusing on both technical and methodological aspects. The course also aims to provide technical skills for querying the database through the SQL language.

Contents

- Database Design
- Conceptual and Logical Design
- Normalization
- Databases for Decision Support
- Changes will be notified during lessons

Detailed program

- Database systems design
 - o Methodologies and models for designing database systems
 - o Information Systems life cycle
 - o Entity-Relationships Model
 - E-R Schema Documentation, Business rules, Documentation techniques
 - Conceptual and Logical Design
 - o Requirements analysis and specification
 - o Design strategies: top-down, bottom-up, inside-out, mixed strategy
 - o Quality of conceptual schemas
 - o CASE tools for designing database systems
 - Normalization
 - o Redundancies and inconsistencies
 - o Functional dependencies
 - o Boyce Codd normal form
 - o Database systems design and normalization: normal form tests, relationship and conceptual schema decomposition
 - Database systems for decision support
 - o Architectures and paradigms for data analysis
 - o Data warehouse architectures
 - o Data warehouse schema typologies: star and snowflake schemas
 - o Data analysis, interfaces for carrying out query, drill down, and roll up activities on data cubes
 - o Data warehouse implementation: bitmap and join indexes, view materialization
 - o Data mining: processes, issues, and perspectives
- Changes will be notified during lessons

Prerequisites

None

Teaching methods

The course will be provided by means of lessons, seminars, and laboratory sessions and homeworks.

Assessment methods

The verification method is based on both a written test and an oral exam.

The written test takes place at the computer and it consists of open and closed questions with multiple answers on all course topics.

The evaluation is focused on the student's ability to answer to specific questions by referring both to the theoretical and practical aspects (through examples) connected to the requested topic.

The written test is common for both attending students and non-attending students.

The oral exam is aimed at assessing the theoretical knowledge of the student on the topics of the course. The ability to reason and deepen the issues proposed during the examination and the methodological rigor of their development will be evaluated.

Textbooks and Reading Materials

Chianese, Moscato, Picariello e Sansone: Sistemi di basi di dati e applicazioni. Maggioli Editore, 2015.

Semester

I semester, II cycle

Teaching language

Italian

3.7 Calcolo delle Probabilità

Docente: Masiero Federica

Mail: federica.masiero@unimib.it

CFU: 6

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MAT/06

Learning objectives

The aim is to present in a rigorous way the fundamental concepts in the theory of probability. Multidimensional random variables and convergence of random variables (which are fundamental tools in Statistic courses) are treated in a systematic way.

Contents

- Revisions on the basic concepts in probability.
- Random variables and multidimensional random variables.
- Characteristic functions.
- Independent random variables.
- Convergence of random variables.
- Laws of Large Numbers and Central Limit Theorem.

Detailed program

- Revisions on the basic concepts in probability, probability measures and their properties.
- Definition of random variable, expectation, moments of order p , law of a random variable
- Multidimensional random variables, joint density, mean vector and covariance matrix
- Characteristic functions: definition, main properties, correspondence between laws and characteristic functions.
- Independent random variables.
- Convergence of random variables: almost sure, in mean of order p , in probability and in law.
- Laws of Large Numbers and Central Limit Theorem.
- Hints of stochastic processes.

Prerequisites

No exams are required as prerequisites. Nevertheless, this activity requires basic knowledge on calculus, mathematical analysis and linear algebra as the ones given in a three year degree course in Statistics.

Teaching methods

Lectures, aiming at giving the theoretical basis of Probability; during the lectures we will discuss exercises on the topics treated in the course.

Assessment methods

Written and oral exam, with exercises and a theoretical question, and oral, with theoretical questions; the oral exam is on the same day of the written exam.

Textbooks and Reading Materials

- P. Baldi, Calcolo delle Probabilità
- F. Caravenna, P. Dai Pra, Probabilità, Un'introduzione attraverso modelli e applicazioni
- A. F. Karr, Probability

Semester

First semester, first period.

Teaching language

Italian.

3.8 Complex Data Analysis

Docente: Fattore Marco

Mail: marco.fattore@unimib.it

Ore: 45

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Learning objectives

To make students capable to set and lead data science projects on complex data systems, integrating their statistical competencies with skills on project management and communication.

Particularly, the course focuses on projects addressing soft and "open" questions, where a precise specification of analytical goals lacks, differently from a "classical statistics setting", but more similarly to current business and institutional contexts. There, in fact, new data sources are the basis for new services, whose design and implementation requires creativity, managerial and communicative skills as well as methodological competencies.

Contents

The course illustrates the kind of activities that a data scientist performs in real organizations and the topic of how to manage data science projects. In practice, these topics are addressed by means of concrete projects to be managed by groups of students, with the supervision of the teacher.

Detailed program

The course is divided into two parts. In the first some lessons, under the form of seminars, will be held to touch upon some specific basic topics, In the second part, students will be assigned projects to be managed and concluded within the course.

PART I

1. Data science: similarities and differences with respect to "classical statistics"
2. The current technological and economic context: complex socio-economic processes, the need for new knowledge and innovative services.
3. Data complexity and new data sources: web, e-commerce, Internet of Things, Smartphones...
4. Examples of data science projects.
5. Basics of Project Management: specificities and criticalities of data science projects, data quality and technological choices.

PART II

Assignment of projects to students, setting of the project management activities and supervision by the teacher.

Prerequisites

There are no formal prerequisites but basic competencies in inferential statistics, data analysis, data mining and R programming are necessary

Teaching methods

Frontal lessons and project supervision

Assessment methods

The competence level will be assessed by:

1. An ongoing evaluation of the way students face the management of the project and its difficulties.
2. A final oral presentation of the project and its result, with a critical discussion.

This assessment method is motivated by the goal to put students into the setting of real business activities and to make their soft skills (e.g. organizational, communicative) and creativity emerge.

Textbooks and Reading Materials

Computers Ltd. What they really can't do, Harel D. =xfordUniversity Press, 2000

Documents and slides provided by the teacher

Web resources and TED Talks

Semester

II semester IV cycle

Teaching language

Italian

3.9 Data Mining e Statistica Computazionale

Docente: Lovaglio Pietro Giorgio

Mail: piergiorgio.lovaglio@unimib.it

Ore: 105

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

Computational statistics

The course aims at introducing software and complex procedures for modelling statistical models both from the theoretical and from the applicative point of view

Data mining

The course aims at introducing statistical models of DATA MINING both from the theoretical and from the applicative point of view.

The student at the end of the course should be able to understand, discern and propose complex models and algorithms, being able to assess the studied topics analyzing read dataset.

Contents

The course deals with complex modelling techniques and main problems and algorithm of Data Mining

Detailed program

Computational statistics

- (1) SAS language
- (2) Interpretation of complex linear Models (Anova, Ancova, GLM)
- (3) Robust methods (Bootstrap, Jackknife, Robust Regression, IRLS, WLS, nonparametric regression, loess smoothing and splines)
- (4) Step of robust model building
- (5) Influence, diagnostics model selection
- (6) Logistic Regression

Data mining

Principles of Data mining, robustness, over fitting and validation. Association rules, Statistical models: linear, discriminant analysis, logistic models, (polytomic and ordinal), Algorithms for the classification: (Naive Bayes, Nearest Neighbour, regression, neural network, Classification TREE, PLS, Bagging, Boosting and Random forest)

Prerequisites

Need to pass exam of Analisi statistica Multivariata

Teaching methods

Class lessons and computer lab

Assessment methods

WRITTEN EXAM: PROJECT WORK

Project work (also in group, to complete before the date of the oral exam) involving a data analysis (R or SAS) on a dataset chosen by the student to replicate arguments and analyses discussed during lab sessions.

Analyses of each module:

Computational statistics (sas base or R)

1 applied Complete work with quantitative target

(descriptive analysis, models, transformations, diagnostics, robust model)

1 more simple applied work with binary target (Fit two models: with all covariates and with model selection/check only collinearity and separation)

Data mining (sas Enterprise Miner or R)

1 applied work with binary target

(To do: descriptive analysis, classification, preprocessing, tuning of models, assessment, score of new data)

Web portals for the choice of the dataset:

<https://archive.ics.uci.edu/ml/datasets>

www.kaggle.com

ORAL EXAM

The oral exam deals with questions on statistical THEORY (see arguments) and on the comments of outputs of the project work to assess the comprehension of principal statistical tools and consequently the "modus operandi" of the conducted statistical analyses.

Textbooks and Reading Materials

Computational statistics

Gareth, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R

Chapter 3 (no section 3.5), Chapter 4, 6,7

Handouts on moodle

Data mining

Gareth, Witten, Hastie, Tibshirani, An Introduction to Statistical Learning with Applications in R

Chapter 2-3-4-5- 8

Handouts on moodle

Semester

I semester cycles I and II

Teaching language

ITA

3.10 Data Science e Modelli Statistici per il Trattamento dei Dati Non Strutturati

Docente: Fattore Marco

Mail: marco.fattore@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Learning objectives

To introduce students to modern tools for non-supervised classification and dimensionality reduction, with a particular focus on:

1. The conceptual unity underlying the problem and its connection of the statistical tools
2. The logic and mathematical structure underlying the algorithms
3. The differences among the algorithms and their consequences in data analysis
4. The intrinsic limits of statistical algorithms

In summary, the goal of the course is to provide students with state-of-the-art competencies on non-supervised classification tools, together with a deep comprehension of the structure of the statistical techniques and a criticism capability towards their use

Contents

The course introduces the problem of non-supervised classification and dimensionality reduction, shows its application to real context, provides basic mathematical results (mainly from linear algebra), illustrates the main statistical algorithms for linear and non-linear reduction, as well as for numerical, non-numerical and partially ordered data. The illustration of the statistical tools is complemented with and supported by examples of analysis and software coding.

Detailed program

1. The problem of non-supervised classification and dimensionality reduction: examples of data analysis from socio-economics and humanities
2. Elements of linear algebra: vector spaces, scalar products, orthogonal projections, matrix norms
3. Linear techniques
 - Singular Value Decomposition (SVD) and its link with Principal Component Analysis
 - Non-negative Matrix Factorization and its comparison to SVD.
 - The Johnson-Lindenstrauss Lemma: bounded distortion dimensionality reduction: Random Projections
 - Multidimensional Scaling.
4. Non-linear techniques
 - Data in differentiable manifolds: Isomap
 - Self-organizing map (SOM)

- Entropy, Kullback-Liebler divergence and dimensionality reduction: SNE e t-SNE
5. Partially ordered data
 - Ranking extraction from multidimensional datasets
 6. The limits of statistical algorithms: the no-free lunch theorem.

Prerequisites

There are no formal prerequisites, but basic competencies on linear algebra, descriptive statistics and data analysis are necessary

Teaching methods

Frontal lessons with exercises and simulations, using the R language, held by the teacher.

Assessment methods

Oral exam.

Such a choice is motivated by:

1. The content of the course, which is of a methodological nature.
2. The relevance for the students to become capable to argue and to organize their thought, being able to perform analogical connections among the different parts of the program.

3.10. DATA SCIENCE E MODELLI STATISTICI PER IL TRATTAMENTO DEI DATI NON STRUTTURATI

Textbooks and Reading Materials

Geometric Structure of High-Dimensional Data and Dimensionality Reduction, Wang J. - Springer 2012.

Methods of Multivariate Analysis, Rencher A. C., Wiley 2002

Introduction to Lattices and Order (second edition), Davey B.A., Priestley H. A., CUP 2002 (chapter 1).

Papers and notes provided by the teacher

Semester

I semester I cycle

Teaching language

Italian

3.11 Demografia

Docente: Blangiardo Gian Carlo

Mail: giancarlo.blangiardo@unimib.it

Ore: 63

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/04

Learning objectives

The course provides the knowledge to analyse qualitative and quantitative changes in a population and the determinants of these changes. The course focuses on the demographic and projection methods both at individual and family level. The students will acquire the ability to find and use demographic data, to analyse and present demographic topics.

Contents

- Demographic data
- Structure and size of a population
- Growth rates in demography
- Analysis of demographic events
- Population projections

Detailed program

- Introduction: the demography

Demographic data: national and international

Structure and size of a population

Growth rates in Demography. Demographic structure. Ageing

Population changes: measure and comparison. Age-specific rates and rates. Comparison between demographic rates.

Analysis of demographic events: methods, basic concepts, rates and probabilities.

Mortality (life table, infant mortality and general mortality); nuptiality (family formation and dissolution, households); fertility (intensity and characteristics, trends and comparison); migration (internal and international, measures and characteristics, the foreign-born population in Italy)

Population projections, synthetic and analytic (age and sex) method.

Demographic trends in Italy and elsewhere: evaluation and problems.

Prerequisites

none

Teaching methods

Lectures and practice labs.

Assessment methods

Written in order to evaluate the ability to treat and interpret statistical data concerning the population and the society and oral (optional)

Textbooks and Reading Materials

G. C. Blangiardo, Elementi di Demografia, Il Mulino, Bologna 1997 (ristampa 2006).

G.C. Blangiardo (a cura di) Sussidiarietà e crisi demografica. Rapporto sulla sussidiarietà 2016/2017. Fondazione Sussidiarietà. Milano 2017

Additional material will be provided during the course on the e-learning website

Semester

II semester, III and IV ciclo (approximatively from the End of February to mid-June)

Teaching language

Italian

3.12 Demografia Sociale (Mobilità e Migrazioni)

Docente: Terzera Laura

Mail: laura.terzera@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/04

Learning objectives

The course offers the tools to study the migratory flows, their characteristics and push and pull factors. Moreover, the course focuses on the key measures and techniques used in studying illegal people on our national territory.

Contents

- History of migrations/Theories of international migration
- Data sources
- Sampling designs
- Estimate of the size of immigrants
- The Italian case

Detailed program

- Mobility and migration definitions

- Migrations in modern societies
- International and internal emigration: the Italian case
- Phases, cycles and types
- Theories on migrations
- Data Sources
- Capture-recapture sampling
- Snowball sampling
- Housing units Sampling Technique
- Centers sampling Technique
- Randomized responses sampling
- Irregular foreigners estimating Methods

Prerequisites

Demography

Teaching methods

Lectures and laboratory (SPSS)

Assessment methods

Attending students: the exam consists in the discussion of an agreed paper and discussion of the topics covered during classes

Not attending students: the exam consists of an oral test in order to evaluate the methodological preparation and the knowledge of the phenomenon

Textbooks and Reading Materials

- C. Bonifazi, *L'Italia delle migrazioni*, Il Mulino, 2013
- M. Ambrosini, *Sociologia delle migrazioni*, Il Mulino, 2005
- International Migration Report 2017, UN, New York, 2018
<http://www.un.org/en/development/desa/population/migration/publications/mi>
Teaching material made available during the lessons on e-learning
(slides, reports, examples, etc)

Semester

I semester II cycle

Teaching language

Italian

3.13 Demografia Sociale (Paesi in Via di Sviluppo)

Docente: Farina Patrizia

Mail: patrizia.farina@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/04

Learning objectives

The course is particularly addressed to supply interpretative frameworks of differences in population dynamics in less developed countries. Since the course has an empirical approach, work groups are formed in order to analyze demographic topics through national and international materials and data.

Contents

- The demographic transition
- The epidemiological and fertility transitions
- Supply and demand for children
- Birth control Programmes
- Contraceptive prevalence and unmet need for contraception
- The Sustainable Development Goals SDGs

Detailed program

The course is designated to increase the knowledge of demographic dynamics comparing the poor and rich countries of the world. To this aim, geographical differences of demographic dynamics will be underlined. In particular, such issues are discussed through interpretative frameworks of differences on population dynamics in less and more developed countries. This approach is able to put forward the historical differences in demographic process that characterizes the areas of the world. Since the course has an empirical approach, work groups are formed in order to analyze demographic topics as mortality, fertility and population dynamics in an international perspective.

Prerequisites

Students must have overcome the exam of Demography

Teaching methods

Taught classes

Assessment methods

For attending students a written examination of contents of the classes, and an oral examination on the Report on Project agreed with Professor. For other students only oral examination on textbooks and reading materials

Textbooks and Reading Materials

Attending students : materials will be provided by professor on e-learning platform Otherwise students have to prepare the following texts:

M. Livi Bacci, Storia minima della popolazione del mondo, Il Mulino, Bologna ultima edizione

G. Micheli Demografie, Mc-Graw Hill, 2011, capitoli 1,4,5,6.

and one among:

Giovannini E. L'utopia sostenibile Laterza 2018

Angeli A, Salvini S. Popolazione mondiale e sviluppo sostenibile, il Mulino 2018

Semester

Second Semester, first period

Teaching language

Italian

3.14 Elementi di Biostatistica

Docente: Bagnardi Vincenzo

Mail: vincenzo.bagnardi@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MED/01

Learning objectives

The aim of the course is to teach how to design an experimental or an observational study in the biomedical field, how to choose the proper statistical method in analyzing data and how to interpret the results.

Contents

1. Introduction to the course
2. Analysis of continuous responses
3. Analysis of categorical responses
4. Analysis of time-to-event (survival) data
5. Insights on the methodology of clinical and epidemiological research

Detailed program

1. Introduction to the course

- 1.1 The steps of the biomedical research and the role of the biostatistician
- 2. Analysis of continuous responses**
 - 2.1 T-test and analysis of variance
 - 2.2 Assumptions
 - 2.3 Non-parametric tests
 - 2.4 Simple and multiple linear regression
- 3. Analysis of categorical responses**
 - 3.1 Analysis of contingency tables
 - 3.2 Simple and multiple logistic regression
 - 3.3. Dose-response relationship
- 4. Analysis of time-to-event (survival) data**
 - 4.1 Time-to-event data
 - 4.2 Non-parametric estimate of the survivor function (Kaplan-Meier method)
 - 4.3 Regression models for survival data: the Cox proportional hazard model
- 5. Insights on the methodology of clinical and epidemiological research**
 - 5.1 Random sampling and random assignment, bias, confounding, causal inference
 - 5.2 Power and sample size calculation in a clinical trial
 - 5.3 Explorative research
 - 5.4. Understanding risk: absolute risk and relative risk

Prerequisites

None

Teaching methods

Lectures

Computer lab with applications in SAS

Assessment methods

Written exam (only for attending students)

During the course, a dataset from a real biomedical study will be provided.

Data must be analyzed using all the methods considered suitable for achieving the aims of the study.

The results must be presented following the style of a research paper.

Oral exam

The oral exam will be based on the course textbook.

Textbooks and Reading Materials

Martin Bland – An Introduction to Medical Statistics – Oxford University Press

Semester

Semester I, Cycle I

Teaching language

Italian

3.15 Epidemiologia

Docente: Zambon Antonella

Mail: antonella.zambon@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MED/01

Learning objectives

The course aims to provide the basic concepts and the tools for the planning and the statistical interpretation of an epidemiological study. At the end of the course students should be able to properly set the level of an epidemiological study, to orient oneself between different observational designs and to provide a statistical contribution to the writing of a research report. For some topics are planned practical exercises conducted with language SAS.

Contents

- Overview of observational study designs
- Measures of occurrence of disease, association and potential impact
- Introduction to bias and methods to control it

Detailed program

- Concept of cause in epidemiology

Observational methods in epidemiology

Ecological studies; ecological fallacy, geographic and temporal correlation studies

Analytical studies; cohort studies (retrospective, follow-up, latency time; case-control studies (retrospective, selection of cases and controls); other studies (case-cohort and case-control nested in the cohort)

Measures of occurrence of disease. Measures of association. Measures of potential impact

Validity and accuracy of the estimates: selection bias, misclassification bias (differential and non-differential) and confounding

Exact and approximate methods for interval estimation of epidemiological measures

Introduction to meta-analysis

Introduction to pharmacoepidemiology

Prerequisites

For this course it's necessary to pass the following exams: Statistica I and Medical statistics

Teaching methods

Lectures and computer lab classes

Assessment methods

Final test mode for attending students

The final test will consist of two moments. In the first moment (basically the same for all) the working groups will present a short seminar of the practical activity carried out organized as a scientific article: introduction to the problem, materials and methods used, main results, discussion and conclusions. Each member of the group will have about 5 minutes available to comment the slides produced. The vote attributed to the seminar will summarize the students' organizational and communication skills as well as the clarity of exposition and the correctness of the statistical methodologies used and the conclusions drawn from the analyzes. At a later stage, each student will hold an oral exam in which the knowledge of the topics covered in the course will be checked (with less emphasis on those already dealt with in the student's seminar) as well as the ability to communicate with an appropriate technical language and the ability to reason in particular scenarios proposed by the teacher. Even the oral will produce a vote. The final grade awarded to the student will be an average of the seminary and oral examination vote provided both are sufficient.

Final test mode for non-attending students

The final exam will be identical to that of the attending students for the oral exam. The seminar will be replaced by an individual in-depth study (paper) on a topic chosen by the student among those dealt with in the course. The paper will have a theoretical and / or practical cut. The essay must be sent to the teacher at the same time as the oral exam. The final grade awarded to the student will be an average of the exam paper and oral exam, provided both are sufficient.

Textbooks and Reading Materials**Students attending the course**

Slides from <http://elearning.unimib.it/>. Other material will be provided by the teacher

Students non attending the course

Slides from <http://elearning.unimib.it/>. Other material will be provided by the teacher

Semester

I semester, II cycle (about from November to January).

Teaching language

The language of the course is the Italian. The work for the seminar (bibliographic research, research of data form national or international sites, etc) and the online self-evaluations are in English language

3.16 Informatica

Docente: Pescini Dario

Mail: dario.pescini@unimib.it

Ore: 68

Tipo di attività: Obbligatorio

Settore disciplinare: ING-INF/05

Learning objectives

The course will introduce basic concepts on the architecture of computers, hardware, software, operating systems, software development and relational databases.

The course will focus on the development of queries to access data and of programs to manipulate the information.

On completion of this course, students will be able to solve problems using a programming language and computing tools.

Contents

- Information elaboration
- Hardware/Software
- Operating Systems
- Databases and relational paradigm
- SQL language:
 - data extraction from diverse tables
 - aggregation operators
 - mathematical operators

- Simple and on aggregate values conditions
- nested queries
- Programming: the python language
 - Variables, Conditional Instructions, Loops
 - Complex Data Structures
 - Files
 - Functions and Procedures

Detailed program

- Elaboration of the information
- Sketches of Hardware/Software
- Operating Systems
- Command Line Interface e Graphic User Interface
- Sketch of the File System
- The Unix File System and the Widows File System
- Programming paradigms
- Databases:
 - client/server paradigm
 - centralized data management
 - the relational paradigm
 - primary and foreign keys
 - sketches of boolean algebra
- SQL language (Data Definition and Data Manipulation Language):
 - Anatomy of an SQL query
 - Aggregation operators
 - Clause Group By

- Nested queries
- Introduction to algorithms
- From algorithms to programs
- Programming languages (interpreter/compiler)
- Programming in Python
 - Structure of a source file
 - Instructions
 - Variables
 - Boolean logic
 - Design and execution environments for Python
 - Conditional instructions
 - Input Output
 - Functions
 - Actual and formal parameters
 - Modules, Packages and Packages Manager
 - Complex data structures (tuples, lists, dictionaries)
 - Complex data structures and iterators
 - File
 - Working with textual files

Prerequisites

None

Teaching methods

Frontal lecture

Blended E-Learning (videos, tests to fix the contents, auto-evaluation tests)

Hands-on session in a computer science laboratory

Final test simulation (to introduce the student to the test platform and help him to gain knowledge about his/her preparation level)

Assessment methods

The examination consists of a written part held in a computer lab.

Facultative Optional oral exam (on request of the teacher or of the students). The oral exam can both increase or decrease the overall evaluation.

Textbooks and Reading Materials

- All the material presented during the lectures will be published on this web site.
- Books
 - Atzeni, P., Ceri, S., Paraboschi, S., & Torlone, R. *Basi di dati*. McGraw-Hill. qua <https://www.mheducation.it/9788838694455-italy-basi-di-dati-5ed-con-connect>.
 - T. Gaddis, *Introduzione a Python*. Pearson. More details here http://www.pearson.it/opera/pearson/0-6270-introduzione_a_python.
 - A. Downey, J. Elkner, C. Meyers. “Pensare da informatico, Imparare con Python”, is the italian version of “How

to Think Like a Computer Scientist”, Green Tea Press, Wellesley, Massachusetts. The EBook can be easily retrieved from the web. Some links: pdf format http://www.python.it/doc/Howtothink/HowToThink_ITA.pdf, html format <http://www.python.it/doc/Howtothink/Howtothink-html-it/index.htm>.

Semester

I semester

Teaching language

Italian

3.17 Laboratorio di Informatica

Docente: Boselli Roberto
Mail: roberto.boselli@unimib.it
Ore: 24
Tipo di attività: Obbligatorio
Settore disciplinare: NN

Learning objectives

The course provides the basics of software programming, in order to allow the student to acquire the necessary knowledge and skills to use automatic information processing tools. The proposed topics will be contextualized through examples expressed in a programming language, that is Python.

Contents

The course is based on the following topics: Programming tools
Structured programming
Object oriented programming
Introduction to scientific computing

Detailed program

The course program is as follows: Summary of the fundamentals of structured programming using the Python programming language: variables, conditional statements, cycles, complex data

structures, files. Processing of text encodings in computer science and with Python. Object-oriented programming: objects, classes and methods in Python. Exception handling in Python.

Prerequisites

No prerequisite.

Teaching methods

The course includes lectures in the classroom and laboratory exercises. The lectures are dedicated to the study of the theoretical topics related to the course. The lab exercises are aimed at solving problems related to programming. In particular, students must learn to program in Python, identifying the most suitable solutions for the exercises presented by the teacher. The solutions are then commented and discussed at the end of each exercise. Finally, tutoring activities are envisaged to support the student who has to face particular difficulties or fill gaps on specific topics among those concerning the course.

Assessment methods

The verification method is based on a written test and a possible oral exam. The written test takes place at the computer and is aimed at assessing the student's ability in programming. It will therefore consist of exercises of both applicative and theoretical type. The evaluation is focused on the student's ability to solve the

proposed problem in terms of Python programming, to identify the appropriate procedures for its solution and to critically discuss the procedures used and the results achieved. Students who pass the written test are admitted to a possible oral exam. Students learn the outcome of the written test through the elearning page of the course itself. The oral exam must necessarily be sustained in the exam appeal in which the written test was held. The oral exam is aimed at assessing the theoretical knowledge of the student on the topics of the course. The ability to formalize the issues proposed during the examination and the methodological rigor of their development will be evaluated. The course does not provide a vote but an approval.

Textbooks and Reading Materials

A. Downey, J. Elkner, C. Meyers, “Pensare da informatico, Imparare con Python”, (italian translation of “How to Think Like a Computer Scientist”), Green Tea Press, Wellesley, Massachusetts, 2002. Further material (exercises and handouts on specific topics) is available on the elearning page of the course. T. Gaddis, *Introduzione a Python*, Pearson, 2016.

Semester

1st cycle, 2nd semester (approximate period from February to April)

Teaching language

Italian

3.18 Modelli Lineari Generalizzati in Epidemiologia e Medicina

Docente: Bellocco Rino

Mail: rino.bellocco@unimib.it

Ore: 45

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: MED/01

Learning objectives

The course aim is to introduce a wide class of statistical models, known as Generalized Linear Models, pinpointing theoretical principles and highlighting their practical aspects and applications in epidemiology and medicine

Contents

- 1) Review of basic principles of probability and inference
- 2) The Linear regression model
- 3) Generalized linear regression model
- 4) The Logistic regression model
- 5) The Ordinal logistic regression model
- 6) The Poisson regression model

Detailed program

Prerequisites

Teaching methods

The material will be presented in class with formal lectures plus applied work using Stata, one of the most used statistical software in biostatistics and epidemiology, with the twofold goal of teaching the software as well as how to interpret the models introduced in class.

Assessment methods

The written exam, based partially on Stata output, will allow to assess the learning outcomes of the course in terms of 1) model formulation 2) model application 3) interpretation of the parameters 4) statistical inference 5) diagnostic and model fitting) choice of best model

An oral discussion will serve to clear the exam questions; the final score can be increased or decreased afterwards.

Textbooks and Reading Materials

Dobson, A.J., An introduction to generalized linear models, 3rd Edition. Chapman & Hall/CRC.B.

Material (notes, scientific articles provided by the teacher).

Semester

II Semester, II term

Teaching language

Italian with English material

3.19 Piano degli Esperimenti

Docente: Chiodini Paola Maddalena

Mail: paola.chiodini@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Learning objectives

The course aims to provide the conceptual basis and tools for both the construction of basic sample designs in the finite populations. The student must know how to plan, analyze and interpret data of an experimental design.

At the end of the course the student must be able to know how to orientate basilarly in the identification of sample units necessary for carrying out an experimental design.

Contents

Definition of a sample plan in the presence of finite populations. Analysis of data deriving from an experimentation.

Detailed program

- Sampling from finite populations
- Simple random sample

- Introduction to proportion estimate
- Stratified sampling
- Cluster sampling
- Introduction to panel sampling
- Fully randomized design (one factor)
- One way ANOVA; two or more ways ANOVA
- 2^k factorial design
- Randomized block design.

Prerequisites

No formal prerequisites required

Teaching methods

Lectures of theory and examples in the classroom.

Assessment methods

The exam consists of a written test that includes questions of theory and exercises and an oral exam.

Textbooks and Reading Materials

For sampling techniques:

Frosini B.V., Montinaro M., Nicolini G., *Il campionamento da popolazioni finite*, UTET, 1999 ; Cochran W.G., *Sampling Techniques*, J. Wiley, New York, 1977.

For design of experiments:

Cochran W.G., Cox M.G., *Experimental Designs*, II ed. Wiley, New York, 1992

Montgomery, D.C., *Progettazione e analisi degli esperimenti*, McGraw-Hill, Milano, 2005

Semester

II Semester, III cycle

Teaching language

Italian

3.20 Popolazione, Territorio e Società I

Docente: Rimoldi Stefania Maria Lorenza

Mail: stefania.rimoldi@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/04

Learning objectives

The course aims to complete the theoretical and practical skills on demographic analyses, with particular attention to specific problems pertaining to spatial aspects.

Contents

- Tools for spatial analysis in Demography.
- Tools for the classification of the territory.
- Introduction to Spatial Autocorrelation.
- Italian contribution to the spatial analysis of the population.
- Introduction to GIS.

Detailed program

Tools for spatial analysis in Demography:

- distribution and concentration;
- accessibility;
- composition;

- geographical association;
- migration;
- diversity and segregation.

Tools for the classification of the territory:

- factorial ecology;
- cluster analysis (homogeneous areas);
- infrastructures planning (urban mobility, optimal location).

Introduction to Spatial Autocorrelation: Joint Count, Moran and Geary indices.

Italian contribution to the spatial analysis of the population:

- atomistic approach (rural-urban dichotomy; settling patterns identification; demographic malaise);
- contextual approach (the metropolitan areas; gravitational areas; functional distances and meantime at first passage).

Introduction to GIS.

Prerequisites

The exam must be preceded by Demography

Teaching methods

Blended e-Learning: 28 hours (4 credits) of lessons in classroom + 14 hours (2 credits) of on line work.

Assessment methods

Attendant: oral examination. The exam consists in evaluating the student's knowledge and comprehension of the subjects presented during the class and in discussing the student's reports on the practical works assigned. The object is to evaluate the student's theoretical knowledge and his ability to apply the methods to practical research themes.

Not attendant: oral examination. The exam consists in evaluating the student's knowledge and comprehension of the subjects presented during the class, also by discussing on practical research problems. The object is to evaluate the student's theoretical knowledge and his ability to apply the methods to practical research themes.

Textbooks and Reading Materials

- Plane D.A. e Rogerson P.A., *The geographical analysis of population. With application to planning and business*, John Wiley & Sons, New York, 1994. Ebdon D., *Statistics in Geography – Second Edition*, Blackwell Publishing, 1985
 - Siegel J S., Swanson D. A., *The Methods and Material of Demography – Second Edition*, Elsevier Academic Press, London, 2004.
 - Golini A., Mussino A, Savioli M, *Il malessere demografico in Italia, Studi e Ricerche* , Il Mulino, 2001.
- Lecture notes by the teacher.

Semester

1° Semester, 1st session - approximately October- November15th

Teaching language

Italian

3.21 Sistemi Informativi

Docente: Mezzanzanica Mario

Mail: mario.mezzanzanica@unimib.it

Ore: 66

Tipo di attività: Obbligatorio

Settore disciplinare: ING-INF/05

Learning objectives

The course aims to create the necessary knowledge, under the technical and methodological point of view, to enable a correct approach to the design of an information system, as a key strategic resource to achieve the goals of a business organization.

Contents

Application architectures and technological architectures of information systems
Digital applications and analysis of information system
Design of information system for data analysis and decision support system
Data quality
Information systems and social media
Introduction to Social Media Analytics
Big Data and techniques for processing unstructured data

Detailed program

Application architectures and technological architectures of information systems: - Database management processes - Di-

distributed architectures, client server, network, Internet and World Wide Web Digital applications and analysis of information system:

- The application portfolio in the industrial and service companies - CRM Design of information system for data analysis and decision support system: - Design of processes and data modelling - BPR - Activities and information analysis Data quality Information systems and social media: - Evolution of enterprise information systems - Social Media Marketing Introduction to Social Media Analytics: - Sentiment Analysis Big Data and techniques for processing unstructured data:

Information extraction

Prerequisites

Fairly good skills in learning, writing and speaking are expected, together with a general knowledge about the main technologies and applications of Computer Science.

Teaching methods

The course is delivered in Italian and includes lectures in the classroom and laboratory exercises. The lectures are dedicated to the study of the theoretical topics related to the course. The lab exercises are aimed at using a tool to analyze unstructured data, in particular to perform Text Mining techniques on data scraped from the web.

Assessment methods

The verification method is based on a written test and a possible oral exam at the student's request. The written test takes place at the computer and it consists of open and closed questions with multiple answers on all course topics. The evaluation is focused on the student's ability to answer to specific questions by referring both to the theoretical and practical aspects (through examples) connected to the requested topic. The written test is common for both attending students and non-attending students. The possible oral exam is aimed at assessing the theoretical knowledge of the student on the topics of the course. The ability to reason and deepen the issues proposed during the examination and the methodological rigor of their development could be evaluated.

Textbooks and Reading Materials

G. Bracchi, C. Francalanci, G. Motta. Sistemi informativi d'impresa. McGraw-Hill, 2010. V. Cosenza, "Social media ROI", Apogeo, 2012. Further material (slides and papers on specific topics) is available on the elearning page of the course.

Semester

The course is delivered in the second cycle of the first semester and in the first cycle of the second semester.

Teaching language

Italian

3.22 Statistica I

Docente: Chiodini Paola Maddalena

Mail: paola.chiodini@unimib.it

Ore: 47

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

The Statistica I course is the first course of statistics in the bachelor of Statistica e Gestione delle Informazioni. It has as its objective to introduce students to the theoretical knowledge and practical skills of statistics that constitute the basis of the bachelor itself. Students at the end of the course must be able to recognize the nature of the statistical variables, knowing how to extract them from a database. They must also be able to represent them graphically in an appropriate manner and should be able to synthesize information that can be both univariate and bivariate through appropriate indicators. It will be required that the student begin to express himself clearly and with proper language, also knowing how to interact with professional figures not necessarily prepared on the statistical discipline.

Contents

Collection and classification of statistical data. Main tools of univariate and bivariate descriptive statistics.

Detailed program

- Statistics as a science
- Populations and statistical units
- Statistical variables and measurement scales
- Graphical representations
- Univariate frequency distributions
- Means
- Variability measures
- Standardized indices
- Asymmetry and skewness
- Bivariate distributions
- Stochastic independence and mean dependence
- Correlation.

Prerequisites

No formal prerequisites required.

Teaching methods

Lectures of theory and examples in the classroom. Exercises in the classroom with some exercises on the blackboard. Moreover, the teaching is flanked by tutoring activities in which further practical activities and corrections of exercises done at home are carried.

Assessment methods

The exam consists of a written test that includes questions of theory and exercises and an oral exam.

Textbooks and Reading Materials

G. Leti, L. Cerbara, Elementi di statistica descrittiva, Il Mulino, Bologna 2009

L. Santamaria, Statistica descrittiva – Applicazioni economiche e aziendali, Vita e Pensiero, Milano 2006

Zanella, Elementi di statistica descrittiva, CUSL, Milano 2000

M. Zenga, Lezioni di statistica descrittiva, G. Giappichelli, Torino 2007

Semester

I Semester, I cycle

Teaching language

Italian

3.23 Statistica I - Complementi

Docente: Chiodini Paola Maddalena

Mail: paola.chiodini@unimib.it

Ore: 47

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

The course is proposed as a natural continuation of the Statistics I course. Some concepts related to bivariate distributions are taken up. Aspects related to the definition and verification of the goodness of adaptation of mathematical models, such as polynomial functions or linearizable models, in a descriptive context that allow the definition of the functional link existing between two (or more) characters are presented. To estimate the coefficients of the models, the least squares method is used.

Contents

Main techniques of bivariate statistics in the field of polynomial regression

Detailed program

- Polynomial regression

- Least squares method
- Data model adaptation
- Residuals of interpolation
- Improvement index
- Multidimensional regression
- Partial correlation coefficient

Prerequisites

The course of Statistics is a prerequisite for the course of Statistics I - Complements.

The knowledge of concepts of mathematical analysis, such as derivatives, is requested.

Teaching methods

Lectures of theory and examples in the classroom. Exercises in the classroom with some exercises on the blackboard.

Assessment methods

The exam consists of a written test that includes questions of theory and exercises and an oral exam.

Textbooks and Reading Materials

G. Leti, L. Cerbara, Elementi di statistica descrittiva, Il Mulino, Bologna 2009

L. Santamaria, Statistica descrittiva – Applicazioni economiche e aziendali, Vita e Pensiero, Milano 2006

Zanella, Elementi di statistica descrittiva, CUSL, Milano 2000

M. Fraire, A. Rizzi, Esercizi di statistica, Carocci Editore, Urbino 2012

Semester

I Semester, I cycle

Teaching language

Italian

3.24 Statistica II

Docente: Berta Paolo

Mail: paolo.bera@unimib.it

Ore: 94

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/01

Learning objectives

The course aims to provide students with an adequate command of the methodologies of statistical inference for the analysis of univariate phenomena that can be represented by stochastic models.

At the end of the course, the student will be able to:

- 1) apply the techniques for the precise estimation and interval of the parameters of the distribution of a random variable
- 2) construct statistical tests to verify assumptions about the distribution of a normal random variable and identify appropriate approximations in the case of any random variable
- 3) set up some recurrent sampling plans in many application contexts.

Contents

The course is divided into macro topics ranging from the definition of random variables to inference based on likelihood. A part

of the course is designed to introduce the most common probabilistic sampling plans in the context of finite populations according to the classical approach.

Detailed program

The course is divided into macro topics:

The notion of sample and the sample space. Punctual estimation. Properties of estimators: correctness, consistency, absolute and relative efficiency. The theorem of Fréchet-Rao-Cramér. The average square error. Estimation methods: the method of maximum likelihood and the method of moments.

Estimate interval and methods for determining the confidence interval. The pivotal quantity. Statistical verification of hypotheses. Significance tests. The main statistical tests: the Z test, the T test, the chi-square test, the F test. The basis of Neyman-Pearson's theory. Error of first and second species. The most powerful test is Neyman-Pearson's lemma. The most uniformly powerful tests. The tests based on the relationship of verisimilitude. Sampling from finite populations. Estimation of the total, average and variance of a continuous variable. Estimate of the relative frequency of a binary variable. Simple random sampling. The stratified sampling. Determination of the sample size.

Prerequisites

The course includes, as preparatory courses, Statistics I, Probability Calculation and Mathematical Analysis I.

Teaching methods

The course is delivered in Italian and includes frontal lessons and classroom exercises.

The lessons are aimed at deepening the theoretical knowledge of the student on the topics of the course. Here, ample space is given to the formalization and derivation of probabilistic and mathematical concepts relevant to the topics considered.

The exercises are aimed at enhancing the student's problem-solving skills. Here, space is given to the formalization of problems present in real situations in terms of statistical inference, to the identification of appropriate procedures for their solution and to the critical discussion of the procedures used and the results achieved.

Assessment methods

The assessment is based on a written test.

The written test is aimed at ascertaining the student's problem-solving skills. It will therefore consist of exercises of both applicative and theoretical type. During the assessment, the student's ability to: formalize the proposed problem in terms of statistical inference, identify the appropriate procedures for its solution and discuss critically the procedures used and the results achieved is considered.

During the written test, the use of texts or other materials is not allowed, with the exception of the variable tables. The use of mobile phones is not allowed during the test.

Anyone who so requests is allowed to take the examination in oral form. The oral test is aimed at ascertaining the student's theoretical knowledge of the course topics. The ability to formalize in statistical-probabilistic terms the themes proposed during the examination and the methodological rigour of their development will then be evaluated.

Textbooks and Reading Materials

Cicchitelli G., D'Urso P., Minozzo M. 2018. Statistica: Principi e Metodi. Terza edizione. Pearson Italia, Milano-Torino

Further material (exercises and handouts on specific topics) will be available on the e-learning page of the course

Semester

The course is delivered during the first semester.

Teaching language

Italian

3.25 Statistica Medica

Docente: Corrao Giovanni

Mail: giovanni.corrao@unimib.it

Ore: 44

Tipo di attività: Obbligatorio

Settore disciplinare: MED/01

Learning objectives

The aim of the course is to provide the fundamental tools to understand the main characteristics of the statistical method in the biomedical research area. At the end of the course, the student will be able to calculate measures of frequency of a disease and the association with the outcomes, and to apply a procedure of direct/indirect standardization. Moreover, the student will recognize the main differences between experimental and observational studies.

Contents

1. The role of statistical sciences in the scientific research
2. The frequency of the disease and of its outcomes
3. The scientific research between experiment and observation
4. Randomized clinical trials
5. Observational studies

Detailed program

The role of statistical sciences in the scientific research

- Natural history of the disease
- Clinical approach to disease
- Public health approach to disease
- Statistical approach to disease

The frequency of the disease and of its outcomes

- Incident and prevalent cases
- Fixed cohort and dynamic population
- Rates and proportions
- Direct/indirect standardization
- Perinatal mortality

The scientific research between experiment and observation

- The concept of causality
- Experimental research
- Observational research

Randomized clinical trials

- The steps of the experimental research
- Within-subjects designs
- Between-subjects designs

Observational studies

- Correlation studies
- Cohort studies
- Case-control studies

Prerequisites

No prerequisites needed

Teaching methods

Lectures/Tutorials

Assessment methods

The exam consists of a written exam (practical exercises) and of an oral exam (if requested by either the teacher or the student).

Textbooks and Reading Materials

The teaching material will be available on the e-learning platform (slides of the lectures, exercises).

Semester

II semester, II cycle (approximately from May to mid June)

Teaching language

Italian

3.26 Statistica Sociale

Docente: Terzera Laura

Mail: laura.terzera@unimib.it

Ore: 63

Tipo di attività: Obbligatorio

Settore disciplinare: SECS-S/04

Learning objectives

The aim is to give the tools for the analysis of social phenomena. Implementation of a survey, attitude scales and the basics of social indicators

Contents

Tools for social research

- Questionnaire: structure and characteristics
- Attitude scales
- Social indicators

Detailed program

First part:

- Opinion survey, data collection. Questionnaire design.
- Survey questions and response alternatives.
- Respondents and responses: Psychological and cognitive processes, phrasing

- Randomized response technique (RRT): Warner and Simmons methods
 - Question order effects and response order effects.
 - Attitude scales: Bogardus, Guttman, Thurstone, Likert, Osgood
- Second part:
- Human Development Indexes and Gap Indicators.
 - Official international and national data base available through Internet.
 - Illustrative scheme of the applied research: choice of a social phenomenon, sources ; analysis, missing values analysis
 - Raw data and construction of social indicators; check of the adequacy of the indicators; drafting a progress report.

Prerequisites

Statistics I

Teaching methods

Frontal and practical lectures. The first faces the theoretical aspects, the second shows , through the use of national and international Open-Data, the different steps of a research: from the acquisition of data, the cleaning of the database, construction of indicators and scales, elementary analysis of results. Software SPSS

Assessment methods

Attending students: The exam consists of a partial written (concerning questionnaire and attitude scales) and two papers (concerning the part of the indicators), the oral exam is optional.

Not attending students: The exam consists of a written test (exercises and theoretical part) and an oral test (theoretical part)

Textbooks and Reading Materials

- Vanda Zammuner (1998), Tecniche dell'intervista e del questionario, il Mulino, Bologna
- Enrica Aureli Cutillo (2002) Lezioni di statistica sociale. Fonti, strumenti e metodi.Ed. CISU
- Handbook on Constructing Composite Indicators METHODOLOGY AND USER GUIDE, OECD, European Commission, 2008
<https://www.oecd.org/std/42495745.pdf>
Teaching material made available during the lessons on e-learning

Semester

II and III cycle (annual)

Teaching language

Italian

3.27 Statistica Spaziale e Ambientale

Docente: Borgoni Riccardo

Mail: riccardo.borgoni@unimib.it

Ore: 42

Tipo di attività: Obbligatorio a scelta

Settore disciplinare: SECS-S/01

Learning objectives

The course aims at providing students with a set of methodologies to deal with spatial and environmental data.

Contents

Geocoding. An introduction to cartography. Explorative spatial data analysis; spatial interpolation. Semiparametric statistical models for environmental data analysis. Estimation of the background level. Computer sessions using the R software for the analysis of environmental spatial data.

Detailed program

- a) Geocoding and an introduction to cartography. GIS. Raster and vector data. Map smoothing for continuous and lattice data
- b) Estimation of the background level
- c) WLS, B-spline and additive models
- d) Laboratory sessions in R.

Prerequisites

Elements of Multivariate Statistics, Inference and Informatics.

Teaching methods

Class lessons and lab sessions in R.

Assessment methods

Lab assesment and oral examination .

The overall mark is obtained by averaging the marks obtained in each part.

Textbooks and Reading Materials

Reading material and software codes will be made available through the eLearning web page of the course.

Semester

First term of the second semester.

Teaching language

Italian.

Indice

1	Introduzione	3
1.0.1	Informazioni generali	13
2	Corsi	33
2.1	Algebra Lineare	33
2.2	Analisi Matematica I	36
2.3	Analisi Matematica II	40
2.4	Analisi Statistica Multivariata	43
2.4.1	Modulo: Analisi Esplorativa	43
2.4.2	Modulo: Modelli Statistici	47
2.5	Analisi dei Dati	53
2.6	Basi di Dati	57
2.7	Calcolo delle Probabilità	61
2.8	Complex Data Analysis	64
2.9	Data Mining e Statistica Computazionale	68
2.10	Data Science e Modelli Statistici per il Trattamento dei Dati Non Strutturati	73

2.11	Demografia	77
2.12	Demografia Sociale (Mobilità e Migrazioni)	81
2.13	Demografia Sociale (Paesi in Via di Sviluppo)	85
2.14	Elementi di Biostatistica	88
2.15	Epidemiologia	92
2.16	Informatica	97
2.17	Laboratorio di Informatica	102
2.18	Modelli Lineari Generalizzati in Epidemiologia e Medicina	106
2.19	Piano degli Esperimenti	110
2.20	Popolazione, Territorio e Società I	113
2.21	Sistemi Informativi	117
2.22	Statistica I	121
2.23	Statistica I - Complementi	124
2.24	Statistica II	127
2.25	Statistica Medica	131
2.26	Statistica Sociale	134
2.27	Statistica Spaziale e Ambientale	138
3	Courses	143
3.1	Algebra Lineare	143
3.2	Analisi Matematica I	146
3.3	Analisi Matematica II	149
3.4	Analisi Statistica Multivariata	152
3.4.1	Modulo: Analisi Esplorativa	152
3.4.2	Modulo: Modelli Statistici	156
3.5	Analisi dei Dati	161
3.6	Basi di Dati	165
3.7	Calcolo delle Probabilità	169

3.8	Complex Data Analysis	172
3.9	Data Mining e Statistica Computazionale	176
3.10	Data Science e Modelli Statistici per il Trattamento dei Dati Non Strutturati	180
3.11	Demografia	184
3.12	Demografia Sociale (Mobilità e Migrazioni)	187
3.13	Demografia Sociale (Paesi in Via di Sviluppo)	190
3.14	Elementi di Biostatistica	193
3.15	Epidemiologia	197
3.16	Informatica	201
3.17	Laboratorio di Informatica	206
3.18	Modelli Lineari Generalizzati in Epidemiologia e Me- dicina	210
3.19	Piano degli Esperimenti	213
3.20	Popolazione, Territorio e Società I	216
3.21	Sistemi Informativi	220
3.22	Statistica I	224
3.23	Statistica I - Complementi	227
3.24	Statistica II	230
3.25	Statistica Medica	234
3.26	Statistica Sociale	237
3.27	Statistica Spaziale e Ambientale	240