

Corso di Laurea Magistrale in

Biostatistica

Anno accademico 2021-2022

1	Inquadramento generale	3
1.1	Obiettivo formativo.....	3
1.2	Insegnamenti.....	4
1.3	Risultati attesi del processo formativo	5
1.4	Per cosa e da chi sono richieste competenze biostatistiche.....	6
2	Ammissione	8
3	Articolazione del percorso formativo	10
3.1	Insegnamenti comuni.....	10
3.2	Insegnamenti curricolari	10
3.3	Insegnamenti a scelta	11
3.4	Altre attività formative	12
3.5	Prova finale.....	12
4	Frequenza.....	13
5	Propedeuticità.....	14
6	Piani di studio	15
7	Orientamento e tutorato	16
8	Scansione delle attività formative e appelli di esame.....	17
9	Programmi dei singoli insegnamenti	18
9.1	Analisi e Modelli Demografici.....	18
9.2	Big Data Analytics.....	21
9.3	Bioinformatica	24
9.4	Elementi di Biologia	26
9.5	Elementi di Medicina per la Ricerca Clinica.....	28
9.5.1	Il Linguaggio Medico	28
9.5.2	Diagnosi e Terapia in Medicina Interna.....	30

9.5.3	Evidenza e Ricerca Clinica	32
9.6	Farmacoepidemiologia.....	35
9.7	Inferenza Statistica	38
9.8	Introduzione ai Modelli Statistici	40
9.9	Introduzione all'Inferenza in Biostatistica	43
9.10	Laboratorio R per la Biostatistica	46
9.11	Machine Learning	48
9.12	Metodi e Modelli Biostatistici per la Ricerca Clinica	50
9.13	Metodologia della Ricerca Clinica ed Epidemiologica (blended)	52
9.14	Modelli Statistici Applicati alle Sperimentazioni Cliniche.....	55
9.15	Modelli Statistici e Inferenza Bayesiana	57
9.15.1	Inferenza Bayesiana	57
9.15.2	Modelli Statistici II.....	61
9.16	Modelli Statistici I	64
9.16.1	Modello Lineare Generalizzato	64
9.16.2	Modelli Lineari per Dati Categoriali	66
9.17	Modelli Statistici per la Genetica.....	69
9.18	Popolazione, Disuguaglianze e Benessere	71
9.19	Statistical Models in Epidemiology	73
9.20	Valutazione Statistica dei Sistemi Sanitari	76
9.21	Programmazione SAS per la Biostatistica	78

Inquadramento generale

Il Corso appartiene alla Classe delle Lauree Magistrali in Scienze Statistiche (LM-82). Ha una durata di due anni. Il conseguimento del titolo comporta l'acquisizione di 120 crediti formativi universitari (CFU).

I CFU vengono acquisiti attraverso:

- il superamento di 12 esami (cfr. par. 3.1, 3.2 e 3.3) ai quali nel complesso sono associati 108 CFU (di cui 18 a scelta)
- lo svolgimento delle cosiddette altre attività formative (cfr. par. 3.4) alle quali sono associati 2 CFU
- il superamento della prova finale (cfr. par. 3.5) alla quale sono associati 10 CFU.

Il corso di studio è ad accesso libero. L'accesso prevede la verifica del possesso dei requisiti curriculari che verrà effettuato in occasione del colloquio con il **coordinatore del Corso** (cfr. par. 2). Il colloquio, oltre a verificare i requisiti curriculari, accerterà la motivazione dello studente.

La lingua ufficiale del corso è l'italiano.

Il titolo consente l'accesso a Master di secondo livello, Dottorati di Ricerca attivati presso l'Università degli Studi di Milano-Bicocca o presso altri Atenei secondo le modalità stabilite nei rispettivi regolamenti.

Il laureato magistrale in Biostatistica ha la possibilità di iscriversi all'albo professionale degli Attuari, previo superamento dell'Esame di Stato.

Il percorso formativo di questo corso di studio garantisce automaticamente allo studente l'acquisizione del *SAS® Graduate joint certificate program curriculum in Advanced Data Analysis in Biostatistics*.

Il 70.4% degli immatricolati si laurea in corso, contro il 59.8% degli studenti delle lauree magistrali non telematiche italiane (Fonte: dati ANVUR 2017/2018 Anagrafe Nazionale Studenti). Il tasso di occupazione dei laureati a tre anni dal conseguimento del titolo è del 100% (fonte: XII indagine AlmaLaurea, anno di riferimento 2019). Si considerano come occupati, secondo la definizione dell'ISTAT, tutti coloro che dichiarano di svolgere un'attività, anche di formazione, purché retribuita).

1.1 Obiettivo formativo

Fornire una preparazione qualificata allo svolgimento di funzioni di progettazione, gestione, analisi, interpretazione statistica e valutazione di studi sperimentali, indagini osservazionali e sistemi di monitoraggio e sorveglianza nei campi della salute umana e animale e nelle sue articolazioni disciplinari della biologia, biotecnologie, studi di popolazione, veterinaria, medicina preventiva, clinica e riabilitativa, scienze socio-economiche e ambientali.

1.2 Insegnamenti

Il corso è strutturato in insegnamenti classificati in tre aree.

Statistica e Data Science

Gli insegnamenti appartenenti a quest'area consentiranno allo studente di acquisire le conoscenze su calcolo delle probabilità, inferenza statistica (approccio frequentista e bayesiano), modelli statistici e demografici, tecniche avanzate di *machine learning* e *big data analytics*, e bioinformatica:

- Introduzione ai modelli statistici
- Introduzione all'inferenza in biostatistica
- Analisi e modelli demografici
- Big data analytics*
- Bioinformatica
- Inferenza statistica
- Machine learning*
- Modelli statistici I (integrato)
 - Modello lineare generalizzato
 - Modelli lineari per dati categoriali
- Modelli statistici e inferenza bayesiana (integrato)
 - Modelli statistici II
 - Inferenza bayesiana
- Popolazione, disuguaglianze e benessere
- Valutazione statistica dei sistemi sanitari

Biostatistica

Gli insegnamenti appartenenti a quest'area consentiranno allo studente di indirizzare le conoscenze ottenute nell'area Statistica e Data Science all'acquisizione di competenze nella pianificazione, analisi e interpretazione di studi osservazionali e sperimentali:

- Metodologia della ricerca clinica ed epidemiologica
- Modelli statistici applicati alle sperimentazioni cliniche
- Metodi e modelli biostatistici per la ricerca clinica
- Statistical models in epidemiology*
- Farmacoepidemiologia
- Modelli statistici per la genetica
- Laboratorio "R" per la biostatistica
- Programmazione SAS per la biostatistica

Biomedicina

Gli insegnamenti appartenenti a quest'area forniranno allo studente le conoscenze di base per interpretare correttamente il quesito biologico/clinico insito in un progetto di ricerca, anche attraverso la documentazione bibliografica e il lavoro in team con biologi e medici:

- Elementi di biologia
- Elementi di medicina per la ricerca clinica (integrato)

Il linguaggio medico
Evidenza e ricerca clinica
Diagnosi e terapia in medicina interna

1.3 Risultati attesi del processo formativo

Conoscenza e capacità di comprensione (knowledge and understanding)

Gli insegnamenti dell'area Statistica e Data Science devono fornire al laureato in Biostatistica:

- solide basi matematiche e statistiche
- approfondita conoscenza delle tecniche statistiche che trovano particolare applicazione nel contesto della biostatistica e della statistica applicata all'ambiente fisico e sociale

Gli insegnamenti dell'area Biostatistica devono fornire al laureato in Biostatistica:

- buona conoscenza degli strumenti logico-concettuali e metodologici della ricerca sperimentale e osservazionale

Gli insegnamenti dell'area Biomedica devono fornire al laureato in Biostatistica:

- particolare sensibilità al linguaggio e un adeguato livello di conoscenza del contesto biomedico al quale la biostatistica è applicata, che lo mette in grado di collaborare con esperti in discipline biologiche, mediche, sociali ed ambientali
- buona conoscenza dei principi etici relativi all'ambito in cui si opera

Capacità di applicare conoscenza e comprensione (applying knowledge and understanding)

Grazie agli insegnamenti dell'area Statistica e Data Science, il laureato in Biostatistica deve essere in grado di:

- effettuare l'analisi statistica dei dati attraverso la costruzione di modelli atti a spiegare i fenomeni oggetto di studio
- riportare i risultati dell'analisi in modo chiaro e non tendenzioso, perché siano direttamente fruibili anche da soggetti non esperti della disciplina statistica

Grazie agli insegnamenti dell'area Biostatistica, il laureato in Biostatistica deve essere in grado di:

- partecipare alla stesura dei protocolli di ricerca fornendo le proprie competenze in diversi ambiti: definizione e scelta del disegno dello studio e della numerosità campionaria tenendo conto del livello di significatività e della potenza dello studio rispetto al contesto in cui sta operando, ii) scelta dei criteri di selezione delle unità statistiche da includere nello studio, iii) messa a punto degli strumenti di raccolta e di misura, dei metodi statistici da impiegare, delle procedure per la gestione dei dati mancanti, per la segnalazione di qualunque deviazione dal piano statistico originale e per il controllo e l'assicurazione della qualità dei dati,

- partecipare alla stesura dei rapporti intermedi e finali di ricerca fornendo le proprie competenze nella i) presentazione e interpretazione statistica dei risultati, ii) valutazione critica dell'evidenza scientifica generata dallo studio, iii) collaborazione al processo decisionale.
- Con tali funzioni il laureato è coinvolto dalla fase di pianificazione degli studi, siano essi sperimentali od osservazionali, a quella di diffusione dei risultati, quale membro stabile di un team dedicato di esperti.

Grazie agli insegnamenti dell'area Biomedica, il laureato in Biostatistica deve essere in grado di:

- lavorare con altre figure professionali recependo il problema biologico, clinico, sociale o ambientale che il gruppo di lavoro intende affrontare e concretizzandolo in una ipotesi statistica da saggiare mediante un opportuno disegno campionario.

Per ognuna delle aree sopra descritte il laureato in Biostatistica deve essere inoltre in grado di aggiornarsi attraverso la:

- letteratura scientifica più qualificata relativamente ai metodi statistici, alle tecniche di indagine e al contesto applicativo nel quale opera
- partecipazione a corsi, congressi/seminari e workshop

1.4 Per cosa e da chi sono richieste competenze biostatistiche

Il laureato in Biostatistica è in grado di esercitare funzioni ed attività coerentemente con gli obiettivi formativi ed i risultati di apprendimento attesi corrispondenti agli sbocchi occupazionali e professionali nei quali siano richieste conoscenze teoriche e abilità pratiche più avanzate per la progettazione, gestione, analisi e interpretazione statistica di studi sperimentali, indagini osservazionali e sistemi di monitoraggio e sorveglianza.

Funzioni

Il laureato magistrale in Biostatistica è coinvolto nella:

- pianificazione degli studi, siano essi sperimentali od osservazionali;
- gestione del dato dalla sua raccolta su supporto cartaceo o elettronico, al controllo della sua qualità;
- analisi statistica dei dati attraverso la costruzione di modelli atti a spiegare i fenomeni oggetto di studio anche con l'impiego di strumenti di calcolo automatico avanzati e di algoritmi sofisticati;
- stesura dei rapporti di ricerca e in particolare della descrizione del disegno dello studio e dei metodi statistici utilizzati e della presentazione e interpretazione critica dei risultati statistici al fine mettere in luce l'evidenza scientifica generata dallo studio;
- collaborazione al successivo processo decisionale.

Competenze

Per poter svolgere tali funzioni il laureato in Biostatistica deve possedere:

- conoscenze biomediche di base per interpretare correttamente il quesito biologico/clinico insito in un progetto di ricerca, anche attraverso la documentazione bibliografica e il lavoro in team con altre figure professionali;
- conoscenze di base e avanzate degli elementi di calcolo delle probabilità, della statistica inferenziale, della modellizzazione dei fenomeni;
- conoscenze dei più utilizzati pacchetti statistici e competenze nell'analisi e nella progettazione di sistemi per la gestione di grandi banche dati.

Sbocchi

La Laurea magistrale in Biostatistica consente l'inserimento nei seguenti ambiti lavorativi:

- istituti di ricerca biologica, biotecnologica, clinica ed epidemiologica,
- aziende ospedaliere,
- istituti di ricovero e cura a carattere scientifico,
- aziende sanitarie locali,
- agenzie sanitarie regionali,
- osservatori epidemiologici,
- registri di patologia,
- aziende farmaceutiche e CRO (Contract Research Organization),
- agenzie regionali per la protezione ambientale,
- aziende di servizi
- istituti di ricerca in ambito socio-economico-demografico

2 Ammissione

Sono ammessi al Corso gli studenti in possesso di titolo di laurea triennale o equipollente, previo colloquio con il **coordinatore del Corso**, per l'accertamento della motivazione e dei requisiti minimi di accesso.

È considerato requisito quantitativo di accesso irrinunciabile la conoscenza/abilità maturata dallo studente nella carriera universitaria precedente nelle aree di competenza del corso, documentata da un numero di esami corrispondenti a un congruo numero di CFU distribuiti in almeno due delle sei aree indicate nel seguente prospetto:

- Area informatica (Contenuti: uno o più pacchetti statistici, Basi di dati; SSD di riferimento: INF/01, ING-INF/05 e affini)
- Area matematica (Contenuti: Analisi matematica, Algebra lineare; SSD di riferimento: MAT/01, MAT/02, MAT/03, MAT/04, MAT/05, MAT/06, MAT/07, MAT/08, MAT/09, SECS-S/06 e affini)
- Area statistica metodologica (Contenuti: Calcolo delle probabilità, Statistica descrittiva, Statistica inferenziale, Statistica multivariata, Piano degli esperimenti; SSD di riferimento: MAT/06, SECS-S/01, SECS-S/02 e affini)
- Area statistica applicata (Contenuti: Statistica medica, Epidemiologia, Psicometria, Demografia, Statistica sociale; SSD di riferimento: MED/01, MED/42, M-PSI/03, SECS-S/04, SECS-S/05 e affini)
- Area biologica (Contenuti: Elementi di scienze della vita (biochimica, fisiologia, genetica, farmacologia e affini; SSD di riferimento: CHIM/08, CHIM/09, BIO/09 – BIO/15, BIO/18, MED/03, MED/04, AGR/07, VET/02, VET/07 e affini)
- Area medica (Contenuti: Basi patologiche delle malattie, ragionamento clinico, medicina preventiva; SSD di riferimento: MED/05, MED/09, MED/42, MED/44, MED/45, MED/50, VET/01, VET/03, VET/08, VET/10, M-PSI/01 e affini)
- Area ambientale (Contenuti: Ecologia e scienze ambientali; SSD di riferimento: CHIM/12, GEO/01, GEO/12, FIS/07, BIO/03, BIO/07, ICAR/03 e affini)

Si trovano normalmente nella condizione di soddisfare tali requisiti le seguenti categorie di laureati:

- nelle classi di laurea di primo livello definiti dal DM 3 novembre 1999 n. 509 in Biotecnologie (classe 1) Scienze biologiche (12), Ingegneria (classi 8-10), Economia (classe 17, 19 e 28), Scienze e tecnologie chimiche (21), Scienze e tecniche farmaceutiche (24), Scienze e tecnologie fisiche (25), Scienze e tecnologie informatiche (26), Scienze e tecnologie per l'ambiente e la natura (27), Scienze matematiche (32), Scienze delle attività motorie e sportive (33), Scienze e tecniche psicologiche (34), Scienze statistiche (37) o nelle corrispondenti classi delle lauree acquisite con l'ordinamento previgente
- nelle classi di laurea a ciclo unico definiti dal DM 3 novembre 1999 n. 509 in Farmacia e farmacia industriale (14/S), Medicina e chirurgia (classe 46/S) Medicina veterinaria (47/S), o nelle corrispondenti classi delle lauree acquisite con l'ordinamento previgente
- nelle classi delle lauree universitarie delle professioni sanitarie (classi 1-4 definiti dal DM 3 novembre 1999 n. 509 e classi SNT1, SNT2, SNT3 e SNT4 definiti dal DM 16 marzo 2007).

Altre situazioni sopra non previste verranno di volta in volta valutate del **Coordinatore del Corso**.

3 Articolazione del percorso formativo

Per conseguire il titolo di laurea magistrale in Biostatistica lo studente è tenuto ad acquisire 120 CFU di cui:

- 60 CFU con il superamento di esami curricolari (cfr. par. 3.1)
- 30 CFU con il superamento di esami curricolari (cfr. par. 3.2)
- 18 CFU con il superamento di esami relativi a insegnamenti a scelta (cfr. par. 3.3)
- 2 CFU con le cosiddette altre attività formative (cfr. par. 3.4)
- 10 CFU il superamento della prova finale (cfr. par. 3.5)

3.1 Insegnamenti comuni

Il percorso formativo è articolato in insegnamenti comuni, curricolari e a scelta.

Gli insegnamenti comuni a tutti i percorsi qui di seguito riportati corrispondono a **60 CFU**:

	SSD	CFU
Inferenza statistica	SECS-S/01	6
Modelli statistici I (integrato)	SECS-S/01	12
Modelli statistici e inferenza bayesiana (integrato)	SECS-S/01	12
Metodologia della ricerca clinica ed epidemiologica	MED/01	12
Statistical models in epidemiology	MED/01	12
Programmazione SAS per la biostatistica	MED/01	6

3.2 Insegnamenti curricolari

Sulla base delle competenze acquisite nella carriera accademica pregressa, e in parte degli orientamenti / interessi personali espressi dallo studente, in occasione del colloquio di ammissione (cfr. par. 2) il coordinatore del corso stabilisce e comunica allo studente lo specifico percorso curricolare al quale lo studente stesso è tenuto ad attenersi nella predisposizione del piano di studi.

I percorsi curricolari previsti sono qui di seguito descritti.

I primi due (“Statistica e data science” e “Biostatistica per statistici”) sono dedicati ai laureati che nella carriera pregressa hanno acquisito competenze nelle aree informatica, matematica e statistica metodologica (cfr. par. 2), normalmente i laureati nella classe in Scienze statistiche (classe 37), o nelle corrispondenti classi delle lauree acquisite con l’ordinamento previgente. Questi studenti possono scegliere tra i seguenti due percorsi:

Statistica e data science, caratterizzato da un orientamento maggiormente indirizzato agli insegnamenti dell’area di Statistica e data Science (cfr. par. 1.2)

Biostatistica per statistici, caratterizzato da un orientamento maggiormente indirizzato agli insegnamenti dell’area di Biostatistica e di Biomedicina (cfr. par. 1.2)

Il terzo percorso, dedicato agli studenti con competenze diverse da quelle descritte nel punto precedente, è denominato:

Biostatistica per non statistici, caratterizzato da alcuni insegnamenti tesi al recupero delle competenze statistiche di base non acquisite nella carriera pregressa

Gli insegnamenti curriculari qui di seguito riportati corrispondono a **30 CFU**:

	SSD	CFU	Statistica e data science	Biostatistica per statistici e affini	Biostatistica per non statistici
Introduzione ai modelli statistici	SECS-S/01	6			X
Introduzione all'inferenza in biostatistica	MED/01	6			X
<i>Machine learning</i>	SECS-S/01	6	X	X	
Analisi e modelli demografici	SECS-S/04	6	X		
<i>Big data analytics</i>	ING-INF/05	12	X		
Laboratorio "R" per la biostatistica	MED/01	6	X		
Modelli statistici applicati alle sperimentazioni cliniche	MED/01	6		X	X
Elementi di biologia	BIO/10	6		X	
Elementi di medicina per la ricerca clinica	MED/09	12		X	X

3.3 Insegnamenti a scelta

Al fine di completare i 108 CFU dedicati al superamento degli esami di profitto, gli studenti devono scegliere ulteriori **18 CFU**. La scelta può riguardare gli insegnamenti curriculari elencati nella tabella del par. 3.2 non già ricompresi tra quelli previsti per lo specifico percorso e/o altri insegnamenti del Corso di studi qui di seguito riportati:

	SSD	CFU
Valutazione statistica dei sistemi sanitari	SECS-S/01	6
Popolazione disuguaglianze e benessere	SECS-S/04	6
Farmacoepidemiologia	MED/01	6
Modelli statistici per la genetica	MED/01	6
Metodi e modelli biostatistici per la ricerca clinica	MED/01	6

3.4 Altre attività formative

Sono volte ad acquisire ulteriori conoscenze utili sia per l'inserimento nel mondo del lavoro che per agevolare le scelte professionali mediante la conoscenza diretta del settore lavorativo cui il titolo di studio può dare accesso. Tali attività sono svolte dallo studente mediante la frequenza di seminari/workshop predisposti a questo fine dal Dipartimento di Statistica e Metodi Quantitativi, da altri Enti universitari o extrauniversitari, o la partecipazione ad attività di ricerca presso aziende pubbliche o private.

Tutti gli studenti devono acquisire 2 CFU nell'ambito di queste attività.

3.5 Prova finale

Per il conseguimento della laurea magistrale è richiesta l'elaborazione di una tesi di laurea originale.

La tesi consiste in un elaborato scritto, contenente i contributi e/o i risultati conseguiti dallo studente mediante un'attività di ricerca teorica e/o sperimentale riguardante tematiche coerenti con gli obiettivi formativi del Corso. In particolare, lo studente dovrà dimostrare di saper inquadrare criticamente la letteratura di riferimento più recente, nonché di essere in grado di utilizzare gli strumenti più appropriati per l'analisi teorica e/o l'investigazione empirica dell'argomento oggetto della tesi e di essere in grado di discutere criticamente i risultati.

Le attività di preparazione della tesi possono essere svolte presso:

- un istituto, ente o centro di ricerca nazionale o estero con il tutoraggio di un ricercatore dell'istituto e la supervisione di un docente del Dipartimento di Statistica e Metodi Quantitativi o di altri Dipartimenti dell'università di Milano-Bicocca,
- il Dipartimento di Statistica e Metodi Quantitativi dell'università di Milano-Bicocca con il tutoraggio di un docente o di un ricercatore del Dipartimento stesso.

La prova finale consiste nella discussione della tesi in seduta pubblica, di fronte a una Commissione composta da professori e/o ricercatori del Dipartimento di Statistica e Metodi Quantitativi ed eventualmente di altri Dipartimenti o Università. La commissione esprime la valutazione finale in centodecimi, con eventuale lode, tenendo conto sia dello svolgimento della prova finale, sia dell'intera carriera universitaria dello studente, secondo quanto stabilito dal Regolamento didattico di Ateneo.

La trasformazione in centodecimi dei voti conseguiti nelle attività formative che danno origine a votazione in trentesimi comporterà una media ponderata rispetto ai relativi CFU acquisiti. L'eventuale attribuzione della lode, è subordinata alla valutazione unanime della Commissione.

La tesi di laurea magistrale può essere redatta e discussa in lingua inglese.

Alla preparazione dell'elaborato e alla discussione dello stesso (ovvero alla prova finale) sono attribuiti **10 CFU**.

4 Frequenza

Non è previsto alcun obbligo di frequenza.

5 Propedeuticità

Non è prevista alcuna propedeuticità.

6 Piani di studio

Il piano di studio è l'insieme delle attività formative comuni, caratterizzanti, a scelta e altre attività formative in coerenza con il Regolamento didattico del corso di studio.

Lo studente deve presentare un proprio piano di studio con l'indicazione delle attività caratterizzanti e a scelta.

Il piano di studio è approvato dal Coordinatore del Corso di Laurea magistrale in Biostatistica.

Se lo studente, oltre agli esami obbligatori, inserisce gli insegnamenti a scelta rimanenti direttamente afferenti al corso (cfr. par. 3.3) il piano si considera automaticamente accettato.

Le modalità e le scadenze di presentazione del piano sono definite dall'Ateneo.

Il diritto dello studente di sostenere prove di verifica relative a una attività formativa è subordinato alla presenza dell'attività stessa nell'ultimo piano di studio approvato.

Per quanto non previsto si rinvia al Regolamento d'Ateneo per gli studenti.

7 Orientamento e tutorato

Le attività di supporto e pianificazione del piano di studi sono svolte da

- Prof. Giovanni Corrao, coordinatore del Corso (giovanni.corrao@unimib.it)
- Prof.ssa Stefania Rimoldi, (stefania.rimoldi@unimib.it)
- Prof. Marco Fattore (marco.fattore@unimib.it)

8 Scansione delle attività formative e appelli di esame

L'organizzazione della didattica è su base semestrale. Ciascun semestre è diviso in due cicli. Ciascun ciclo comprende, di norma, oltre alle settimane dedicate alle lezioni, una settimana di recupero e un periodo dedicato agli appelli d'esame degli insegnamenti che si sono conclusi alla fine del ciclo. Ciascun insegnamento prevede 6 appelli d'esame ogni anno accademico così suddivisi:

- una sessione a novembre (per tutti i corsi del secondo semestre oppure del primo semestre che terminano nel primo ciclo),
- due sessioni nel periodo gennaio-febbraio,
- una sessione ad aprile (per tutti i corsi del primo semestre oppure del secondo semestre che terminano nel terzo ciclo) e
- tre sessioni nei mesi di giugno, luglio e settembre.

Il calendario didattico e l'orario delle lezioni sono pubblicati sul sito del CdL.

Nel caso l'emergenza sanitaria dovuta alla pandemia da Covid-19 dovesse comportare limitazioni alla presenza in Ateneo, i docenti di ogni singolo corso comunicheranno, tramite la piattaforma e-learning, le modalità di svolgimento a distanza delle lezioni e degli appelli di esame.

9 Programmi dei singoli insegnamenti

9.1 Analisi e Modelli Demografici

<i>Codice insegnamento:</i>	F8203B007M
<i>Docente/i:</i>	Elisa Barbiano Di Belgiojoso
<i>Settore scientifico disciplinare:</i>	SECS-S/04
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, III Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	42

Obiettivi formativi

Il corso intende offrire gli strumenti per valutare ed interpretare le manifestazioni dei fenomeni demografici e sociali e delle trasformazioni quantitative e strutturali di una popolazione. La prima parte del corso si focalizzerà sullo studio delle principali tendenze demografiche attraverso modelli di scomposizione su dati aggregati e cross-sectional. La seconda parte si incentrerà sulle conseguenze dei cambiamenti demografici sociali in atto utilizzando tecniche di previsione. La terza parte del corso si concentrerà sullo studio di dati demografici longitudinali attraverso l'analisi delle sequenze. Alla fine del corso gli studenti disporranno di strumenti per un'analisi critica della realtà demografica e delle sue evoluzioni.

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

- interpretazione delle trasformazioni della popolazione
- scomposizione tra effetto tassi ed effetto sostituzione
- analisi delle sequenze

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

- identificare le cause delle trasformazioni demografiche
- prospettare evoluzioni future per i fenomeni demografici
- analizzare dati longitudinali attraverso la definizione di sequenze

Contenuti sintetici

1. Evoluzione nel tempo di fenomeni demografici e sociali, confronto nel tempo. Intensità ed effetto struttura.
2. Scenari demografici futuri: previsioni della popolazione e previsioni derivate.
3. Analisi delle sequenze applicata in ambito demografico.

Programma esteso

1. Evoluzione nel tempo di fenomeni demografici.
 - a) Indagini ripetute e indagini panel confronto tra i due disegni campionari, vantaggi svantaggi e possibili tecniche di analisi dati
 - b) Misure sintetiche: tassi generici e tassi specifici, relazione tra tasso generico e tassi specifici

- c) Confronti tra misure sintetiche: problematiche, distinzione effetto struttura e intensità. Scomposizione di Kitagawa per tassi generici
- d) Presentazione delle principali problematiche demografiche italiane
- e) Laboratorio 1: applicazioni a dati reali per lo studio dell'evoluzione demografica e delle componenti che determinano i cambiamenti.
- 2. Scenari demografici futuri: previsioni della popolazione e previsioni derivate.
 - a) Modelli di previsione della popolazione
 - b) Previsioni derivate
 - c) Tassi fissi e tassi variabili
 - d) Contributo dei tassi e della struttura della popolazione alla definizione degli scenari futuri
 - e) Laboratorio 2: applicazione a dati reali per la formulazione di scenari futuri
- 3. Analisi delle sequenze per lo studio di fenomeni demografici:
 - a) dati longitudinali (struttura)
 - b) definizione di sequenza
 - c) metodi di confronto tra sequenze
 - d) distanza tra le sequenze
 - e) algoritmo optimal matching
 - f) costi
 - g) esempi in ambito demografico e biostatistico
 - h) Laboratorio 3: applicazioni al processo di ricongiungimento familiare della popolazione straniera.

Prerequisiti

Nessun prerequisito formale

Metodi didattici

Lezioni frontali teoriche e laboratori con applicazione a dati e problemi reali. I laboratori sono finalizzati allo sviluppo di un progetto individuale su un tema concordato con la docente. Per i laboratori saranno utilizzati STATA ed Excel.

Modalità di verifica dell'apprendimento

Frequentanti: consegna di un elaborato scritto svolto individualmente sul terzo laboratorio e presentazione orale dei risultati dei primi due laboratori. Tale modalità di verifica dell'apprendimento permette di verificare le capacità di analisi degli studenti attraverso l'applicazione di modelli demografici a dati reali, l'interpretazione dei risultati ottenuti e la discussione di questi alla luce delle teorie viste a lezione. Inoltre permette di valutare le capacità di stesura di un breve rapporto di ricerca (elaborato) e le capacità espositive nella presentazione dei risultati del progetto.

Non frequentanti: una prova orale. Tale prova verterà sui testi e il materiale indicato nella sezione “testi di riferimento” e accerterà l'apprendimento sia degli aspetti teorici (metodi e modelli) che di interpretazione dei risultati prodotti (non viene richiesto di produrre i risultati bensì di saperli interpretare). Durante la prova orale sarà inoltre accertata la conoscenza delle principali teorie rispetto all'evoluzione passata dei fenomeni demografici e alle tendenze in atto e prospettive future. Non sono previste prove in itinere.

Testi di riferimento

Per i frequentanti:

Testi di riferimento:

- Stefani Scherer (2013). Analisi dei dati longitudinali, Il Mulino. Capitoli 1,2 e 5.
- Firebaugh (1997). Analyzing repeated surveys, SAGE University Paper (e-book sul sito biblioteca)

Materiale aggiuntivo:

- articoli selezionati sui temi trattati a lezione

I non frequentanti per preparare l'esame devono utilizzare i seguenti testi e materiali aggiuntivi:

Testi di riferimento:

- Stefani Scherer (2013). Analisi dei dati longitudinali, Il Mulino. Capitoli 1,2 e 5.
- Golini A. (2019). Italiani poca gente. Il paese ai tempi del malessere demografico. Luiss
- Golini A. e Rosina A. (a cura di) (2011). Il secolo degli anziani. Come cambierà l'Italia. Il Mulino. Introduzione e capitoli 1,2, 3, 7 e 8.

Materiale aggiuntivo

- Articoli selezionati sui temi trattati a lezione (indicati nella sezione non frequentanti sulla pagina del corso).
- Le lezioni videoregistrate disponibili sulla piattaforma e-learning

Lingua di insegnamento

Italiano

9.2 Big Data Analytics

<i>Codice insegnamento:</i>	F8203B041
<i>Docente/i:</i>	Fabio Mercurio
<i>Settore scientifico disciplinare:</i>	INF/01
<i>CFU:</i>	12
<i>Periodo:</i>	II Semestre, III / IV Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	94

Obiettivi formativi

Il corso intende fornire le competenze (metodologiche e tecnici) e gli strumenti per la comprensione e la realizzazione di soluzioni per il processamento di Big Data (strutturati e non), attraverso l'uso degli algoritmi e tool di AI per l'estrazione e rappresentazione della conoscenza da dati reali. Inoltre, il corso intende fornire gli strumenti tecnici per la modellazione e realizzazione di data model in accordo con il paradigma NoSQL, focalizzando principalmente sui graph-database e database NoSQL. Infine, si forniranno competenze tecniche e metodologiche relativamente ad algoritmi di explainable AI per la comprensione di algoritmi di machine learning black box

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

- Architetture big data e tool principali
- SQL e NoSQL
- eXplainable AI
- text processing e text mining (cenni)
- big data quality
- cloud computing

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

- modellare dati per scalabilità orizzontale e verticale
- comprendere le basi delle principali architetture big data
- modellare dati in paradigmi noSQL (graph e documentale)
- pre-processare testo con le librerie python (stato dell'arte)
- comprendere le tecniche per l'analisi della qualità per dataset di grandi dimensioni
- comprendere gli elementi alla base del cloud computing (iaas,paas,saas)
- comprendere le tecniche per spiegare algoritmi black-box (interpretable machine learning)

Contenuti sintetici

- Introduction to AI and Big Data Analytics
- Getting knowledge from data
- Modelling and Querying the Resulting knowledge through NoSQL
- Explainable AI (global and local interpretation models)

Programma esteso

Introduction to AI and Big Data Analytics

1. Goal and rationale of AI. The relation between Big Data and AI
2. The value of knowledge – digital economy and data-driven decision making
3. The big data architecture (state of the art)
4. Cloud Computing

Getting knowledge from data

1. Word Embedding (Word2Vec, Doc2Vec, GLOVE, FastText, BERT)
2. Evaluate word embedding models (intrinsic vs extrinsic evaluation)
3. Topic Modelling through Python

Modelling and Querying the Resulting knowledge through NoSQL

1. introduction to NoSQL data stores
2. graph-databases and graph-traversal query languages (Cypher)
3. Document Databases (mongoDB)
4. Big Data Quality

Explainable AI (global and local interpretation models)

1. Introduction to XAI, local/global interpretation models. model agnostic-specific algorithms
2. XAI techniques as in the state of the art (eg. LIME, SHAP, etc)

Prerequisiti

- Python
- SQL

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 (o un progetto) ed una eventuale prova orale.

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/progettuale 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

In fase di valutazione. Per aggiornamenti consultare il syllabus del corso sulla piattaforma e-learning e didattica online dell'Ateneo, disponibile nella seconda metà del 2021.

Lingua di insegnamento

Italiano

9.3 Bioinformatica

<i>Codice insegnamento:</i>	F8203B018M
<i>Docente/i:</i>	Gianluca Della Vedova, Dario Pescini
<i>Settore scientifico disciplinare:</i>	INF/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, III Ciclo
<i>Tipo di attività:</i>	Opzionale
<i>Ore:</i>	52

Obiettivi formativi

L'obiettivo principale consiste nel fornire una introduzione ai dati di natura biologica, alle tecniche algoritmiche di analisi e alle tecniche di simulazione.

Tutti gli argomenti vengono illustrati usando il linguaggio Python.

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

- Notebook Jupyter
- Gestione di dati in formato tabellare tramite la libreria Pandas
- Gestione di dati biologici tramite la libreria Biopython
- Formati di file in bioinformatica (FASTA, GTF)

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

- Scrivere programmi in Python per la gestione e l'analisi di dati biologici

Contenuti sintetici

Introduzione a Python

Introduzione a Pandas

Biopython

Numpy e Matplotlib

Introduzione alla systems biology

Programma esteso

Introduzione a Python: sintassi, liste, array, dizionari. Notebook. Moduli e librerie

Numpy

Introduzione a Python: lettura file, espressioni regolari

Introduzione a Pandas: DataFrame e lettura csv

Pandas: funzioni di riepilogo, gestione tabelle

Biopython

Dal rilevatore alla stringa

Matplotlib

Introduzione alla systems biology

Modellazione stocastica e deterministica

Importanza dei parametri e Ottimizzazione combinatoria
Constraint based modeling

Prerequisiti

Nessuno

Metodi didattici

Lezioni frontali in laboratorio, attività di laboratorio.

Modalità di verifica dell'apprendimento

Prova orale che consiste nella discussione di un progetto che può essere svolto in piccoli gruppi. La modalità d'esame non cambia durante l'emergenza.

La realizzazione del progetto permette di verificare se gli studenti sono stati in grado di trasformare le conoscenze apprese nel corso in competenze.

La valutazione è basata sulla capacità di realizzare le analisi richieste con un codice che sia corretto, semplice e manutenibile.

Non sono previste prove in itinere. La modalità di esame è la stessa per tutti gli studenti.

Testi di riferimento

- [Pensare in Python](#)
- [Python Data Science Handbook](#)

Lingua di insegnamento

Italiano

9.4 Elementi di Biologia

<i>Codice insegnamento:</i>	F8203B005M
<i>Docente/i:</i>	Maria Elena Regonesi
<i>Settore scientifico disciplinare:</i>	BIO/10
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, I e II ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	42

Obiettivi formativi

Il corso intende fornire le conoscenze di base di biologia e la capacità di comprendere i meccanismi alla base dei principali processi biologici per poter affrontare e discutere argomenti di natura biologica in termini statistici.

Contenuti sintetici

Il corso è suddiviso in 3 macroargomenti (biochimica, genetica e farmacologia) e fornisce le conoscenze essenziali di chimica generale e organica, una trattazione delle principali classi di molecole biologiche, con particolare enfasi su proteine e acidi nucleici. Vengono anche impartite conoscenze elementari circa l'organizzazione strutturale delle cellule e un quadro generale del metabolismo. Il materiale genetico. Basi dell'ereditarietà. Genetica quantitativa. Genetica di popolazioni. Microarray. Farmacocinetica. Farmacodinamica. Farmacotossicologia.

Programma esteso

Elementi di biochimica: teoria del legame chimico (legame ionico e covalente); le principali classi di composti organici; interazioni non covalenti e loro ruolo nella struttura delle molecole biologiche; concetto di concentrazione; l'equilibrio chimico; il pH e il suo ruolo nei processi biologici; soluzioni tampone; principali classi di composti di interesse biologico e loro ruoli strutturali e funzionali; carboidrati; lipidi; nucleotidi e acidi nucleici; struttura di DNA e RNA; struttura degli aminoacidi presenti nelle proteine; il legame peptidico; i diversi livelli organizzativi delle proteine (struttura primaria, secondaria, terziaria e quaternaria); ruoli funzionali delle proteine, alcuni esempi di proteine di rilevante interesse biologico; l'organizzazione strutturale dei sistemi viventi; differenze nella struttura e nell'organizzazione generale di cellule procariotiche ed eucariotiche; struttura e ciclo vitale dei virus; il ruolo centrale dell'ATP nel metabolismo energetico; il metabolismo ossidativo; vitamine e ormoni.

Elementi di genetica: basi fisiche dell'ereditarietà (cromosomi, mitosi, meiosi); replicazione e trascrizione del DNA; genetica Mendeliana; genetica quantitativa; genetica di popolazioni (equilibrio di Hardy-Weinberg, polimorfismo delle popolazioni naturali, variazioni delle frequenze geniche); principi generali dei microarray.

Elementi di farmacologia: definizione di farmaco, agonisti, antagonisti, agonisti inversi e parziali; farmacocinetica (assorbimento, distribuzione ed eliminazione dei farmaci); farmacodinamica (modalità d'azione di un farmaco, esempi di categorie di farmaci, determinazione delle curve dose-risposta). Sistema nervoso (potenziale d'azione, neurotrasmettitori, sinapsi, anatomia funzionale). Sistema muscolare (scheletrico, cardiaco e liscio. La contrazione muscolare). Il rene. Il sangue. Il sistema immunitario.

Prerequisiti

Nessuno

Metodi didattici

Lezione frontale

Modalità di verifica dell'apprendimento

Il corso è suddiviso in 3 macroargomenti (biochimica, genetica e farmacologia) che prevedono 3 prove distinte, una orale di biochimica, uno scritto di genetica e una orale di farmacologia. La prova scritta di genetica prevede domande aperte per verificare l'acquisizione dei fondamenti teorici della materia e due esercizi pratici. Le prove saranno le stesse per frequentanti e non frequentanti. Non sono previste prove in itinere.

Testi di riferimento

Il materiale didattico sarà fornito all'inizio del corso.

Lingua di insegnamento

Italiano

9.5 Elementi di Medicina per la Ricerca Clinica

9.5.1 Il Linguaggio Medico

<i>Codice insegnamento:</i>	F8203B031
<i>Docente/i:</i>	Michele Bombelli
<i>Settore scientifico disciplinare:</i>	MED/09
<i>CFU:</i>	1
<i>Periodo:</i>	I Semestre, I-II Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	7

Obiettivi formativi

L'obiettivo del corso consiste nel preparare gli studenti alla più produttiva collaborazione tra il biostatistico e il medico. A tal fine, è necessario soffermarsi sulla progressiva suddivisione della medicina specialistica in decine di discipline e sulla sempre più complessa articolazione tra anamnesi, esami di laboratorio, esami strumentali ed alternative terapeutiche. L'analisi di lavori scientifici della letteratura medica e lo studio di alcune tra le più rilevanti patologie, dall'eziologia alla prognosi, sono i principali strumenti per introdurre alla comprensione dell'attività medica.

Contenuti sintetici

Elementi di Semeiotica: rilevazione dei sintomi e dei segni e collegamento con le alterazioni anatomiche, patologiche e metaboliche che li determinano. Processo di formulazione della diagnosi.

Valore degli esami clinici nel processo diagnostico: esemplificazione di iter diagnostici di alcune importanti patologie di medicina interna (tromboembolismo venoso, insufficienza cardiaca, diagnosi di forme secondarie di ipertensione arteriosa).

Collaborazione fra medici e statistici in studi di popolazione: esemplificazione di studio epidemiologico sulla misurazione dei valori di pressione arteriosa nella popolazione generale, studio della loro correlazione dei valori di pressione arteriosa ottenuti con metodi diversi con altre variabili rilevanti nella determinazione del rischio cardiovascolare e studio del loro valore prognostico.

Programma esteso

- Le domande del paziente al medico in relazione alla malattia.
- Raccolta diretta dei dati clinici
 - 1) anamnesi familiare, fisiologica, patologica remota e patologica prossima
 - 2) descrizione dettagliata dell'esame obiettivo
- Integrazione dei dati clinici diretti con gli esiti dei test laboratoristici e radiologici nella formulazione della diagnosi clinica.
- Interpretazione dei test clinici: sensibilità, specificità, valori predittivi.
- Esempi di percorsi diagnostici in importanti patologie internistiche di diversa natura:
 - 1) Il tromboembolismo venoso
 - 2) Forme secondarie di ipertensione arteriosa
 - 3) Insufficienza cardiaca

- Presentazione ed interpretazione dei dati di uno studio epidemiologico sulla popolazione generale inerente ai valori di pressione arteriosa misurati con differenti metodiche, loro correlazione con altre variabili nella definizione del rischio cardiovascolare globale e studio del loro valore prognostico attraverso i dati di morbilità e mortalità generale e cardiovascolare.
- Esempi di differente rilevanza clinica della variabilità di parametri biologici utilizzati nei processi diagnostici.

Prerequisiti

Nessun prerequisito formale.

Metodi didattici

Lezione frontale

Modalità di verifica dell'apprendimento

Esame orale. Questa modalità di esame è quella che meglio consente di verificare l'acquisizione da parte dello studente della capacità di intendere la natura dei quesiti diagnostici e/o terapeutici che il medico deve affrontare, essendo ciò di basilare importanza per il corretto utilizzo della metodologia statistica in relazione alle differenti tipologie di problemi clinici e per l'ottimizzazione dell'interazione fra medico e biostatistico.

Testi di riferimento

Il materiale è costituito da diapositive e videoregistrazioni disponibili sulla piattaforma elearning.

Lingua di insegnamento

Italiana

9.5.2 Diagnosi e Terapia in Medicina Interna

<i>Codice insegnamento:</i>	F8203B031
<i>Docente/i:</i>	Michele Bombelli, Hernan Francisco Polo Friz
<i>Settore scientifico disciplinare:</i>	MED/09
<i>CFU:</i>	4
<i>Periodo:</i>	I Semestre / I e II Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	28

Obiettivi formativi

L'obiettivo del corso consiste nel preparare gli studenti alla più produttiva collaborazione tra il biostatistico e il medico. A tal fine, è necessario soffermarsi sulla progressiva suddivisione della medicina specialistica in decine di discipline e sulla sempre più complessa articolazione tra anamnesi, esami di laboratorio, esami strumentali ed alternative terapeutiche. L'analisi di lavori scientifici della letteratura medica e lo studio di alcune tra le più rilevanti patologie, dall'eziologia alla prognosi, sono i principali strumenti per introdurre alla comprensione dell'attività medica

Contenuti sintetici

Fisiologia dei principali organi del corpo umano (concetti generali) e patologia generale (premessa del ragionamento medico)

Specialità mediche (suddivisione delle attività clinica in specialità mediche e chirurgiche) e attività del medico (diagnosi, terapia, prevenzione, riabilitazione e ricerca)

Componenti dell'anamnesi. Anamnesi familiare, lavorativa, chirurgica, farmacologica, stili di vita

Anamnesi sintomatologica e semeiotica. Il linguaggio dei sintomi e dei segni

Terapia medica e terapia chirurgica. Farmaci, procedure strumentali, interventi. Compliance

Principali malattie internistiche (etiologia, patogenesi, sintomatologia, semeiotica, esami di laboratorio e strumentali e cenni di terapia di alcune malattie: malattie cardiovascolari, broncopneumopatie croniche)

La salute nel mondo (malattie comunicabili e non comunicabili)

Programma esteso

Cardiopatía ischemica: definizione, fisiopatologia, clinica, diagnosi, terapia.

Iperensione arteriosa: definizione, fisiopatologia, clinica, diagnosi, terapia.

Insufficienza cardiaca: definizione, fisiopatologia, clinica, diagnosi, terapia.

Ictus: definizione, fisiopatologia, clinica, diagnosi, terapia.

BPCO: definizione, fisiopatologia, clinica, diagnosi, terapia.

Artrite Reumatoide: definizione, fisiopatologia, clinica, diagnosi, terapia

Embolia polmonare: definizione, fisiopatologia, clinica, diagnosi, terapia

Prerequisiti

Nessuno

Metodi didattici

Lezione frontale

Modalità di verifica dell'apprendimento

Esame orale. Questa modalità di esame è quella che meglio consente di verificare l'acquisizione da parte dello studente della capacità di intendere la natura dei quesiti diagnostici e/o terapeutici che il medico deve affrontare, essendo ciò di basilare importanza per il corretto utilizzo della metodologia statistica in relazione alle differenti tipologie di problemi clinici e per l'ottimizzazione dell'interazione fra medico e biostatistico.

Lingua di insegnamento

Italiano

9.5.3 Evidenza e Ricerca Clinica

<i>Codice insegnamento:</i>	F8203B031
<i>Docente/i:</i>	Sarino Aricò
<i>Settore scientifico disciplinare:</i>	MED/09
<i>CFU:</i>	7
<i>Periodo:</i>	I Semestre, I e II Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	49

Obiettivi formativi

Rendere lo studente in grado di interagire con il medico nell'elaborazione di studi clinici. A tal fine, è essenziale che lo studente conosca i processi decisionali del medico, sia nella pratica clinica che nella gestione della letteratura medica.

Contenuti sintetici

Si inizia con la rassegna storica dell'iter che ha portato la medicina e la chirurgia a frazionarsi in oltre 50 discipline specialistiche. Si analizzano i determinanti di questo processo che è iniziato a metà del secolo scorso ed oggi trova compimento in una suddivisione in campi ristretti di ogni specialità medica o chirurgica. Si studiano poi le varie fasi del lavoro del medico in dettaglio.

Si parte dall'anamnesi, si esaminano le potenzialità dell'esame obiettivo, si passano in rassegna gli esami di laboratorio e quelli strumentali. Si tratta poi del processo terapeutico e della riabilitazione e dei metodi per stabilire la prognosi.

Ogni studente è incaricato di esaminare la letteratura medica a riguardo di un argomento che viene scelto congiuntamente. Tale argomento viene esposto in dettaglio con un'analisi critica della letteratura.

Si studiano le caratteristiche del lavoro scientifico in medicina: introduzione, stato dell'arte, bisogni conoscitivi, materiali e metodi, criteri di inclusione e di esclusione, risultati, tabulazioni, discussione, proposta di nuovi studi, conflitti di interesse, bibliografia.

Programma esteso

Verranno trattati i seguenti argomenti:

- Le specialità mediche
- Le specialità chirurgiche
- L'anamnesi
- L'anamnesi chirurgica
- L'anamnesi farmacologica
- L'anamnesi familiare
- L'anamnesi lavorativa
- Le allergie
- I sintomi.
- L'esame obiettivo
- Gli esami di laboratorio

- Gli esami radiologici
- Gli esami endoscopici
- Gli altri esami strumentali
- Le terapie mediche
- Le terapie chirurgiche
- La prognosi
- La riabilitazione
- La letteratura medica
- L'introduzione
- I materiali e metodi
- I risultati
- Le tabulazioni
- La discussione
- Il conflitto di interessi
- I bisogni conoscitivi
- La bibliografia
- La gastroenterologia
- Malattie dell'esofago
- Malattie dello stomaco e del duodeno
- Malattie del piccolo intestino
- Malattie del colon
- Il microbiota intestinale
- I farmaci in gastroenterologia
- L'epatologia
- Epatopatie virali
- L'epatite cronica
- La cirrosi epatica
- Lo scompenso epatico
- L'epatocarcinoma
- Le malattie del pancreas
- L'alcol quale fattore di rischio
- Disturbi da uso di alcol
- Alcol e patologie organiche
- Alcol quale carcinogeno
- Alcolismo

Vari aspetti della letteratura medica verranno analizzati congiuntamente con gli studenti.

Prerequisiti

Nessun requisito formale richiesto.

Metodi didattici

Lezioni frontali con discussione continua con gli studenti che dovranno predisporre un seminario su un argomento scelto da loro.

Modalità di verifica dell'apprendimento

Modalità di verifica dell'apprendimento: Esame orale. Questa modalità di esame è quella che meglio consente di verificare l'acquisizione da parte dello studente della capacità di intendere la natura dei quesiti diagnostici e/o terapeutici che il medico deve affrontare, essendo ciò di basilare importanza per il corretto utilizzo della metodologia statistica in relazione alle differenti tipologie di problemi clinici e per l'ottimizzazione dell'interazione fra medico e biostatistico.

Testi di riferimento

Materiale specifico verrà fornito durante il corso

Lingua di insegnamento

Italiano

La letteratura medica viene letta e relazionata in inglese

9.6 Farmacoepidemiologia

<i>Codice insegnamento:</i>	F8203B019
<i>Docente/i:</i>	Giovanni Corrao, Anna Cantarutti
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, IV Ciclo
<i>Tipo di attività:</i>	Opzionale
<i>Ore:</i>	45

Obiettivi formativi

Il corso si propone di introdurre lo studente alla conoscenza dei principali disegni di studio e tecniche statistiche utilizzati nel campo della farmacoepidemiologia.

Conoscenza e comprensione

Questo Corso di Studio fornirà conoscenze e capacità di comprensione relativamente a:

- Processo di registrazione di un nuovo farmaco, principali tecniche di farmacovigilanza e studi *post-marketing* di fase IV per la valutazione dell'efficacia del farmaco dopo la sua immissione al commercio
- Principali disegni di studi osservazionali in ambito farmacoepidemiologico, e utilizzo degli archivi amministrativi sanitari
- Principali misure e indicatori utilizzati per valutare la farmacoutilizzazione
- Principali fonti di confondimento misurato (*detection and time related bias*) e/o misclassificazione dell'esposizione, insieme alle opportune tecniche per controllare tale confondimento (restrizione, appaiamento, stratificazione, etc.)
- Principali fonti di confondimento non misurato; e relative tecniche statistiche (approccio *rule-out*, etc.) e disegni di studio (*case-crossover*, etc.) per il controllo di tali fonti di confondimento

Capacità di applicare conoscenza e comprensione

Alla fine del Corso di Studio gli studenti saranno in grado di:

- Effettuare criticamente una ricerca nella letteratura scientifica
- Scegliere l'opportuno disegno di studio in funzione del quesito clinico/farmacoepidemiologico cui si vuole rispondere (impostare il protocollo di uno studio clinico sperimentale od osservazionale considerando le potenziali fonti di confondimento (misurato e non) e le tecniche/disegni da utilizzare per tenerle sotto controllo); oltre che interpretarne i risultati

Contenuti sintetici

- Trials clinici e studi osservazionali
- Uso degli archivi elettronici
- Aspetti legati al disegno dello studio
- Fonti di distorsione
- Approfondimenti: Indicatori di farmacoutilizzazione, Misclassificazione dell'esposizione, Confondimento misurato e non misurato

Programma esteso

1. Introduzione:

- Perché un approccio epidemiologico allo studio del rapporto tra uomo e farmaco?
- Insufficienza di trial clinici preregistrativi
- Insufficienza del sistema di sorveglianza basato sulle segnalazioni spontanee
- Inadeguatezza dei sistemi di monitoraggio attivo
- Il modello (farmaco)epidemiologico di riferimento
- Profili di farmacoutilizzazione e di beneficio-rischio
- Uso degli archivi elettronici in farmacoepidemiologia
- Database clinici vs. database amministrativi

2. Indicatori di farmacoutilizzazione

- Definizione di compliance, persistenza e aderenza

3. Farmacovigilanza

4. Detection bias

5. Misclassificazione dell'esposizione

6. Time related bias

- Immortal time bias
- Immeasurable time bias
- Time-window bias

7. Tecniche per il controllo del confondimento

- Riepilogo sulle tecniche per il controllo dei confondenti misurati (restrizione, appaiamento, stratificazione, regressione multivariate, propensity score)

8. Tecniche per il controllo del confondimento non misurato

- Approccio Rule-out
- Monte-Carlo sensitivity analysis
- Propensity score calibration

9. Disegni per il controllo del confondimento non misurato (disegni case-only): (i) case-crossover, (ii) case time control, (iii) Self-controlled case-series

Prerequisiti

Nessun prerequisito

Metodi didattici

Il corso prevede lezioni frontali e alcune ore in laboratorio per lo svolgimento del lavoro a gruppi utile al fine della prova finale.

Modalità di verifica dell'apprendimento

L'esame finale di verifica dell'apprendimento presenta delle distinzioni tra studenti frequentanti e studenti non frequentanti:

- **Studenti frequentanti.** L'esame consiste in (a) un lavoro di gruppo in cui verrà approfondito uno degli argomenti svolti a lezione tramite la lettura e l'analisi di un articolo scientifico che sarà

presentato al docente; e (b) una prova orale finale composta da domande aperte per verificare l'apprendimento degli argomenti del corso

- **Studenti non frequentanti.** L'esame consiste in una prova orale finale durante la quale gli studenti verranno invitati a ragionare insieme al docente sui principali argomenti trattati durante il corso, con l'intento di verificare se hanno appreso i concetti illustrati e raggiunto una adeguata capacità di maneggiare e possedere tali concetti.

Testi di riferimento

Giovanni Corrao. Real World Evidence Buone Pratiche della ricerca basata sull'osservazione del mondo reale. Il Pensiero Scientifico Editore, 2019.

Tutto il materiale necessario al superamento della prova verrà caricato online nella piattaforma e-learning.

Periodo di erogazione dell'insegnamento

Lingua di insegnamento

Italiano

9.7 Inferenza Statistica

<i>Codice insegnamento:</i>	F8203B003
<i>Docente/i:</i>	Claudio Giovanni Borroni
<i>Settore scientifico disciplinare:</i>	SECS-S/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, III Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	42

Obiettivi formativi

L'insegnamento si propone di presentare concetti e metodi avanzati dell'inferenza statistica, con particolare riguardo a problemi di stima e di verifica d'ipotesi. Allo studente verranno forniti gli strumenti per l'analisi e l'elaborazione dei dati campionari, con la duplice finalità di fornire elementi di supporto alle decisioni e di permettere l'applicazione di modelli complessi. Lo studente acquisirà inoltre gli strumenti per la comprensione di metodologie inferenziali più specifiche, sviluppate in altri insegnamenti del percorso universitario.

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- Quali sono le principali metodologie di stima dei parametri di una distribuzione o di un modello e quali sono i loro punti di forza e di debolezza.
- Quali problematiche emergono dall'applicazione pratica di tali metodologie; in particolare, quale ruolo è rappresentato dalle assunzioni su cui si basano.
- Quali problematiche emergono quando tali metodologie vengono applicate per produrre decisioni.

Alla fine dell'insegnamento gli studenti saranno in grado di:

- Individuare la tecnica inferenziale più adatta ad un certo contesto o ad un certo modello. In particolare, gli studenti sapranno distinguere tra tecniche basate unicamente sul concetto di verosimiglianza e tecniche che incorporano le opinioni iniziali (cosiddette tecniche bayesiane).
- Comprendere il significato e i limiti dei risultati ottenuti da terze parti mediante l'impiego di specifiche tecniche inferenziali.
- Implementare le tecniche inferenziali conosciute in contesti non banali da un punto di vista analitico, eventualmente ricorrendo a metodi di approssimazione numerica o a procedure di simulazione.
- Affrontare problemi inferenziali in cui le assunzioni classiche sono violate, eventualmente ricorrendo a tecniche non basate su particolari distribuzioni (cosiddette tecniche non parametriche).
- Comprendere gli argomenti sviluppati in altri insegnamenti del percorso universitario focalizzati sull'uso di modelli statistici, sulla statistica bayesiana e sulla statistica computazionale.

Contenuti sintetici

Stima puntuale: criteri di ottimalità degli stimatori, principali metodi di stima. Stima intervallare. Verifica di ipotesi: criteri di ottimalità dei test, metodi per ricercare test. Metodi nonparametrici. Applicazioni in SAS.

Programma esteso

Campionamento e distribuzioni campionarie. Campionamento da popolazioni Normali. Convergenza di sequenze di variabili aleatorie. Teoremi limite e loro applicazioni. Simulazione di distribuzioni campionarie mediante il metodo Monte Carlo. Principi di riduzione delle informazioni campionarie: sufficienza, sufficienza minimale, ancillarità, completezza. Funzione di verosimiglianza e principio di verosimiglianza. Stima puntuale e metodi per cercare stimatori: metodo dei momenti, metodo della massima verosimiglianza, stimatori Bayesiani, algoritmo EM. Ottimalità degli stimatori: disuguaglianza di Cramér-Rao, teorema di Rao-Blackwell, teorema di Lehman-Scheffé. Stima intervallare: quantità pivotali, intervalli di credibilità Bayesiani, probabilità di copertura. Verifica di ipotesi: test basati sul rapporto di verosimiglianze, test Bayesiani, test unione-intersezione, ottimalità dei test. Valutazioni asintotiche e robustezza: consistenza, metodo Bootstrap, intervalli basati su grandi campioni, test asintotici basati sul rapporto di verosimiglianze. Procedure SAS per i principali test parametrici e nonparametrici.

Prerequisiti

Nessun prerequisito formale. E' comunque consigliato, a chi proviene da una formazione pregressa non quantitativa, seguire le lezioni di "Introduzione all'inferenza statistica".

Metodi didattici

L'insegnamento prevede lezioni frontali ed esercitazioni.

Modalità di verifica dell'apprendimento

L'esame prevede una prova scritta, comprendente esercizi e domande di teoria, e una prova orale. Accedono all'orale gli studenti che abbiano riportato almeno 18 trentesimi nella prova scritta.

Le domande di teoria consentono di verificare la conoscenza e la comprensione delle tecniche inferenziali trattate. Gli esercizi consentono di verificare la capacità di applicare tali tecniche ad uno specifico problema e la capacità di interpretare i risultati ottenuti. Inoltre, sia le domande di teoria sia gli esercizi consentono di verificare la capacità di esprimersi con un linguaggio tecnico adeguato.

Non esiste distinzione formale tra le modalità d'esame per gli studenti frequentanti e non frequentanti. Tuttavia, durante il periodo di erogazione del corso, gli studenti possono svolgere degli assignment, ossia dei lavori da svolgersi autonomamente a casa. La partecipazione agli assignment è libera, ossia non prevede l'obbligo di frequenza e la gestione avviene tramite la piattaforma e-learning. Gli assignment comprendono esercizi, domande di approfondimento sulla teoria e applicazioni tramite il software SAS. Gli studenti che risolvono correttamente gli assignment potranno sostituire una parte (circa un terzo) della prova scritta.

Testi di riferimento

Indicazioni sui testi di riferimento verranno fornite all'inizio delle lezioni.

Lingua di insegnamento

Italiano

9.8 Introduzione ai Modelli Statistici

<i>Codice insegnamento:</i>	F8203B029
<i>Docente/i:</i>	Nadia Solaro
<i>Settore scientifico disciplinare:</i>	SECS-S/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, I Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	42

Obiettivi formativi

Il corso si propone di richiamare gli aspetti metodologici e applicativi di base dei modelli statistici secondo i punti di vista descrittivo e inferenziale. Argomenti principali sono il modello di regressione lineare sia semplice che multipla per dati di popolazione e per dati campionari, e la loro analisi con il software SAS. Il corso si rivolge ai laureati con formazione pregressa non quantitativa o non assimilabile a quella tipicamente acquisita nelle lauree triennali in Statistica.

Conoscenza e comprensione.

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- Principali metodi per lo studio del legame di dipendenza di una variabile quantitativa da una o più variabili quantitative, insieme eventualmente a variabili qualitative, e per la costruzione della relativa modellistica lineare in ambito sia descrittivo sia inferenziale
- Applicazione dei principali metodi di analisi bivariata mediante esercizi numerici svolti con la calcolatrice (ossia, senza l'ausilio del software statistico)
- Logica e funzionamento alla base del software SAS e suo utilizzo nell'ambito dell'analisi bivariata e della modellistica di regressione con le relative rappresentazioni grafiche
- Lettura e interpretazione degli output delle analisi prodotte con SAS.

Capacità di applicare conoscenza e comprensione.

Alla fine dell'insegnamento gli studenti saranno in grado di:

- Applicare i metodi di base dell'analisi bivariata, stabilire nel modo più opportuno il ruolo da assegnare alle variabili (ossia se dipendenti o indipendenti), e costruire i modelli di regressione più adeguati in base agli scopi delle analisi e alla natura dei dati a disposizione
- Selezionare un modello di regressione che sia al contempo parsimonioso ma di buon adattamento in base a criteri statistici descrittivi o inferenziali
- Interpretare i risultati delle analisi in modo critico e individuare gli eventuali margini di miglioramento nella costruzione delle modellistiche di base applicate in ottica sia interpretativo-descrittiva, sia predittiva
- Importare in SAS file di dati esterni di varia provenienza e formato e utilizzare in modo autonomo la sintassi di base di SAS.

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

Contenuti sintetici

Dipendenze in media e lineare, interpolazione fra punti e per punti, funzione di regressione. Regressione lineare semplice e multipla per dati di popolazione e per dati campionari. Selezione del modello di regressione in base a criteri descrittivi e inferenziali. Cenni al modello di regressione logistico binomiale. Applicazioni a dati reali con il software SAS.

Programma esteso

- Richiami sulle forme principali di dipendenza su dati di popolazione quantitativi: dipendenze in media e lineare. Interpolazione fra punti e per punti. Funzione di regressione. Criterio di accostamento dei minimi quadrati
- Regressione lineare: retta dei minimi quadrati, bontà di adattamento, estensione a più variabili esplicative, correlazioni multipla e parziale, regressori qualitativi
- Modello di regressione lineare semplice e multipla per dati campionari: distribuzione normale multivariata e sue proprietà, specificazione del modello, ipotesi, metodi di stima dei minimi quadrati e di massima verosimiglianza, intervalli di confidenza, verifica di ipotesi
- Selezione del modello di regressione lineare multipla: criteri descrittivi e inferenziali
- Cenni al modello di regressione logistico binomiale: specificazione, forma lineare e moltiplicativa, odds e odds ratio, stima dei parametri, intervalli di confidenza, verifica di ipotesi, bontà di adattamento
- Analisi di casi empirici con SAS

Prerequisiti

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

Metodi didattici

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

Modalità di verifica dell'apprendimento

L'esame di Introduzione ai Modelli Statistici consiste in una prova orale che ha quale oggetto gli argomenti sia teorici sia applicativi trattati durante il corso.

Le domande teoriche riguardano gli aspetti metodologici degli argomenti trattati e consentono di verificare le conoscenze teoriche acquisite in merito alla logica e agli aspetti di base dell'analisi affrontata su dati di popolazione o su dati campionari, dell'analisi bivariata, della specificazione formale dei modelli di regressione, dei problemi legati alla determinazione o alla stima dei parametri e alla valutazione della bontà di adattamento del modello ai dati, e della conduzione dell'inferenza statistica in un'ottica di modellistica. Permettono inoltre di verificare la capacità di utilizzare in autonomia il linguaggio simbolico-formale statistico, di fornire in modo appropriato le definizioni e di dimostrare analiticamente i principali risultati teorici.

Le domande applicative riguardano la trasposizione nella pratica della teoria, e consentono di verificare sia la capacità di leggere e interpretare i risultati delle analisi ottenute con il software SAS, sia le conoscenze acquisite in merito alla loro realizzazione con le procedure di SAS.

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

Testi di riferimento

- Materiale didattico della docente pubblicato sul sito e-learning del corso (ad accesso riservato con password)
- Delwiche, L.D., Slaughter, S.J. (2012), *The Little SAS Book: A Primer*, SAS Institute
- Freund, R. J., Wilson, W. J., and Sa, P. (2006), *Regression Analysis: Statistical Modeling of a Response Variable*, 2nd edition, Academic Press
- Johnston, J. (1993), *Econometrica*, 3rd edition, Franco Angeli, Milano
- Littell, R. C., Freund, R. J., and Spector, P. C. (2002), *SAS for Linear Models*, 4th Edition, Cary, NC: SAS Institute Inc.
- Piccolo, D. (2010), *Statistica*, Il Mulino, Bologna
- Spencer N. (2004), *SAS Programming - The One-Day Course*, 1st Edition, Chapman and Hall/CRC, New York
- Zelterman, D. (2010), *Applied Linear Models with SAS*, Cambridge University Press, New York
- Zenga, M. (2014), *Lezioni di Statistica Descrittiva*, Giappichelli, Torino

Lingua di insegnamento

Italiano

9.9 Introduzione all'Inferenza in Biostatistica

<i>Codice insegnamento:</i>	F8203B039
<i>Docente/i:</i>	Rino Bellocco
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, I Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	45

Obiettivi formativi

Obiettivo dell'insegnamento è fornire allo studente i concetti di base dell'inferenza statistica applicata all'ambito biomedico.

I concetti statistici saranno proposti sia da un punto di vista formale, che utilizzando un approccio intuitivo basato sulle simulazioni.

Ogni concetto proposto sarà illustrato con applicazioni ed esemplificazioni di ambito biomedico.

Esercizi pratici da svolgere individualmente o in gruppo saranno proposti durante il corso.

Il corso si rivolge a laureati di formazione non statistica.

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- concetto di probabilità e di distribuzione di probabilità;
- concetto di stima campionaria e di variabilità di una stima;
- concetto di verifica statistica di un'ipotesi e di test statistico;
- concetto di correlazione e regressione

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti avranno gli strumenti per poter seguire gli insegnamenti del corso di laurea in biostatistica, insegnamenti in cui saranno presentati i modelli statistici di base e avanzati per l'analisi di studi sperimentali e osservazionali in ambito biomedico.

Contenuti sintetici

- Richiamo agli elementi di statistica descrittiva
- Probabilità, distribuzioni e variabili casuali
- Metodi di simulazione
- Stima campionaria e variabilità di una stima
- Test di significatività
- Introduzione ai modelli di regressione lineare

Programma esteso

1. ***Richiamo agli elementi di statistica descrittiva***
 - 1.1 Tipi di dati
 - 1.2 Distribuzioni di frequenze
 - 1.3 Mediane e quantili
 - 1.4 La media

- 1.5 Varianza e deviazione standard
- 1.6 Applicazioni all'ambito biomedico
- 2. *Probabilità, distribuzioni e variabili casuali***
- 2.1 Proprietà fondamentali
- 2.2 Distribuzioni di probabilità e variabili casuali
- 2.3 La distribuzione bernoulliana e binomiale
- 2.4 La distribuzione di Poisson
- 2.5 La distribuzione normale
- 2.6 La distribuzione esponenziale
- 2.7 Applicazioni all'ambito biomedico
- 3. *Metodi di simulazione***
- 3.1 Importanza delle simulazioni in ambito statistico
- 3.2 Simulazione di una variabile casuale discreta
- 3.3 Simulazione di una variabile casuale continua
- 4. *Stima campionaria e variabilità di una stima***
- 4.1 Distribuzioni campionarie
- 4.2 Errore standard della media campionaria
- 4.3 Teorema del limite centrale
- 4.4 Intervalli di confidenza
- 4.5 Confronto tra due medie
- 4.6 Confronto tra due proporzioni
- 4.7 Applicazioni all'ambito biomedico
- 5. *Test di significatività***
- 5.1 Verificare un'ipotesi statistica
- 5.2 Principi dei test di significatività
- 5.3 Livelli di significatività e tipi di errore
- 5.4 Test statistici per il confronto tra due medie
- 5.5 Test statistici per il confronto tra due proporzioni
- 5.6 Test multipli
- 5.7 Applicazioni all'ambito biomedico
- 6. *Il modello di regressione lineare semplice***
- 6.1 Ipotesi del modello
- 5.2 Interpretazione e stima dei parametri
- 5.3 Inferenza statistica sui parametri
- 5.4 Goodness of fit del modello
- 5.5 Analisi dei residui

Prerequisiti

Nessuno

Metodi didattici

Lezioni

Esercitazioni al computer con applicazioni in Stata

Modalità di verifica dell'apprendimento

Prova scritta e orale

Ci saranno sia una prova scritta che un orale, secondo le linee guida che verranno date all'inizio del corso dai docenti. L'esame scritto sarà seguito da una valutazione orale basata sugli argomenti del programma.

L'esame scritto e orale consentirà di valutare le conoscenze dello studente sull'inferenza statistica applicata al campo biomedico.

Testi di riferimento

1. Principles of Biostatistics, M Pagano and K. Gauvreau. Second Edition, CRC press.
2. An introduction to Medical Statistics, M Bland. Fourth Edition. Oxford.
3. Fundamentals of Biostatistics, Bernard Rosner. Thomson. Sixth Edition.

Lingua di insegnamento

Il materiale di insegnamento sarà in inglese, le lezioni tuttavia saranno svolte in Italiano

9.10 Laboratorio R per la Biostatistica

<i>Codice insegnamento:</i>	F8203B044
<i>Docente/i:</i>	Antonella Zambon, Anna Cantarutti
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, I Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	45

Obiettivi formativi

Il corso si propone di fornire le basi teoriche e le conoscenze di R utili alla gestione e analisi di dati raccolti mediante un disegno epidemiologico sperimentale o osservazionale. Verranno inoltre forniti esempi di utilizzo di tecniche di machine learning in questo contesto.

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- la gestione dei dati con R
- l'uso di diversi modelli e tecniche biostatistiche con R
- l'applicazione di tecniche di machine learning di classificazione e di selezione delle variabili negli studi sperimentali ed osservazionali

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- gestire con il linguaggio R il database di partenza per una buona sintesi e visualizzazione dei dati
- analizzare con il linguaggio R i dati provenienti da uno studio sperimentale o osservazionale
- applicare tecniche di classificazione e di selezione delle variabili nella costruzione dei modelli statistici

Contenuti sintetici

Gestione dei dati con R

Metodi statistici per l'analisi dei principali disegni sperimentali e osservazionali in R

Metodi di machine learning in R per la classificazione e la selezione delle variabili

Programma esteso

Gestione dei dati con R: i) Introduzione al linguaggio R, ii) metodi per la pulizia dei dati, iii) metodi per la creazione di report e la visualizzazione dei dati

Metodi statistici per l'analisi dei principali disegni sperimentali e osservazionali in R: i) analisi di uno studio trasversale, ii) analisi di uno studio di coorte, iii) analisi di uno studio caso-controllo, iv) analisi di uno studio sperimentale

Metodi di machine learning in R per la: i) verifica della capacità discriminatoria di un test diagnostico, ii) la selezione delle variabili in un modello multivariato

Prerequisiti

Nessuna propedeuticità formale. Si richiede però la conoscenza dei contenuti degli insegnamenti di Modelli statistici I

Metodi didattici

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

Modalità di verifica dell'apprendimento

Modalità prova finale

Lavoro di gruppo su una analisi di dati reali da consegnare prima della prova orale che consiste nell'esposizione del lavoro svolto. Tale modalità d'esame consentirà di valutare nei candidati la capacità di elaborazione critica dei metodi applicati e dei risultati ottenuti.

Non sono previste prove in itinere.

La modalità d'esame è la stessa per frequentanti e non frequentanti

Testi di riferimento

Slide delle lezioni e materiale integrativo disponibili sulla piattaforma della didattica online <http://elearning.unimib.it/>.

Lingua di insegnamento

L'insegnamento è completamente erogato in italiano. La maggior parte del materiale di riferimento è in inglese.

9.11 Machine Learning

<i>Codice insegnamento:</i>	F8203B040
<i>Docente/i:</i>	Stefano Peluso, Mirko Cesarini
<i>Settore scientifico disciplinare:</i>	SECS-S/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, III Ciclo
<i>Tipo di attività:</i>	Obbligatorio a scelta
<i>Ore:</i>	44

Obiettivi formativi

Fornire sia le conoscenze sia le capacità di utilizzare algoritmi di Machine Learning e di applicarli alla soluzione di problemi. Lo studente acquisirà padronanza delle metodologie di machine learning per affrontare problemi di classificazione e regressione. acquisirà la capacità critica necessaria per valutare quali problemi possono essere risolti con tecniche di machine learning.

Conoscenza e comprensione. Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a: algoritmi di machine learning, categorie di problemi risolvibili con algoritmi di machine learning, principali metodologie applicabili e criticità a cui prestare attenzione.

Capacità di applicare conoscenza e comprensione. Alla fine dell'insegnamento gli studenti saranno in grado di: applicare algoritmi di machine learning per affrontare problemi di classificazione, regressione ed in generale per estrarre informazioni dai dati. Saranno in grado di svolgere sia dei test di fattibilità sia di realizzare soluzioni complete.

Contenuti sintetici

- Statistical methods for machine learning
- Beyond linear models
- Feature Engineering and Machine Learning Algorithms Tuning
- Artificial Neural Networks and Deep Learning

Programma esteso

- Statistical methods for machine learning
- Supervised and unsupervised learning
- Recall to regression analysis
- Classification analysis
- Cross validation and bootstrap
- Model selection and regularization
- Beyond linear models
- Tree-based methods
- Support vector machines
- Feature Engineering and Machine Learning Algorithms Tuning
- Feature Engineering and Selection (Bag of Words, Embeddings, Tensors, ...)

- Data Observability and Model existence issues
- Hyperparameters optimization (Grid-Search, Random-Search, Advanced Search methodologies)
- Artificial Neural Networks and Deep Learning
- Artificial Neural Networks (ANNs) and Feed Forward Neural Network introduction
- Training Algorithm: Gradient Descent, Optimization Methodology
- Deep learning and Artificial Neural Networks types (Fully Connected networks, Feed Forward networks, Convolutional networks, Recurrent networks, ...)
- Industrial applications and open research issues

Sulla base delle conoscenze iniziali degli studenti i docenti decideranno quali parti del programma approfondire.

Prerequisiti

Linear Algebra, Aver frequentato un corso di statistica descrittiva e inferenziale, Aver frequentato un corso di programmazione.

Desiderabili, ma non necessarie: conoscenze del linguaggio Python e del linguaggio R

Metodi didattici

Lezione in laboratorio informatico

Modalità di verifica dell'apprendimento

Esame orale. Potranno essere concordati dei progetti sostitutivi di parte dell'esame orale.

Testi di riferimento

Saranno comunicati dal docente all'inizio della lezione.

Gareth James, Daniela Wittens, Trevor Hastie and Robert Tibshirani (2013). An Introduction to Statistical Learning. Springer. Available at <http://www-bcf.usc.edu/~gareth/ISL/>

C.M. Bishop (2006), Pattern Recognition and Machine Learning. Springer (New York)

Lingua di insegnamento

Inglese

9.12 Metodi e Modelli Biostatistici per la Ricerca Clinica

<i>Codice insegnamento:</i>	F8203B034
<i>Docente/i:</i>	Antonella Zambon
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, IV Ciclo
<i>Tipo di attività:</i>	Opzionale
<i>Ore:</i>	42

Obiettivi formativi

Il corso si propone di fornire le basi teoriche e le conoscenze informatiche necessarie per l'analisi di dati raccolti mediante un disegno sperimentale o osservazionale, e l'interpretazione dei risultati, con particolare attenzione ai disegni adattativi e ai modelli predittivi. Tutti gli argomenti sono completati da esercitazioni pratiche condotte in ambiente SAS.

L'insegnamento consente allo studente di acquisire solide basi nell'applicazione della statistica al contesto lavorativo biostatistico.

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- l'uso dei diversi disegni sperimentali e i principi degli studi predittivi
- l'applicazione delle tecniche statistiche dei modelli misti agli studi sperimentali ed osservazionali

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- analizzare i dati provenienti da uno studio sperimentale o osservazionale con i modelli misti
- interpretare sinteticamente e con linguaggio tecnico le misure ottenute
- scrivere un protocollo di ricerca di uno studio sperimentale

Contenuti sintetici

Metodi statistici per l'analisi dei principali disegni sperimentali e di quelli adattivi

Metodi statistici per l'analisi dei modelli predittivi

Programma esteso

Introduzione teorica e pratica (in ambiente SAS) ai modelli statistici per l'analisi dei principali disegni degli esperimenti (completamente casualizzato, a blocchi, fattoriale, crossover)

Disegno e analisi degli studi randomizzato d'intervento (randomizzazione a cluster)

Introduzione agli studi adattativi

Introduzione ai modelli predittivi

Calibrazione e discriminazione nei modelli predittivi

Prerequisiti

Nessuna propedeuticità formale. Si richiede però la conoscenza dei contenuti degli insegnamenti di Modelli statistici I

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

Scritto con orale facoltativo. La prova scritta è divisa in due parti (prova di laboratorio e tesina). Nella prova di laboratorio ogni studente dovrà analizzare un piccolo dataset di dati reali relativi a uno studio osservazionale o sperimentale e riportare i principali passaggi delle analisi e dei risultati commentandoli in funzione del quesito clinico richiesto nel tema d'esame. Il software da utilizzare è SAS. Il prodotto dell'analisi (report dei risultati e commenti) vengono salvati in un file word e stampati per consegnarli al docente che verificherà la correttezza dell'impostazione e la valutazione dei risultati ottenuti. L'attività in laboratorio durerà circa un'ora e mezza. Il voto attribuito alla prova di laboratorio sintetizzerà le capacità organizzativa di un lavoro di analisi oltre alla correttezza delle metodologie statistiche utilizzate e delle conclusioni. Nella tesina ciascuno studente approfondirà, da un punto di vista teorico, un argomento a scelta tra quelli presentati nel corso. Anche la tesina produrrà un voto. Il voto finale attribuito allo studente sarà una media del voto della prova di laboratorio e della tesina purché entrambe siano sufficienti.

Modalità prova finale per non frequentanti

La prova finale sarà identica a quella dei frequentanti.

Testi di riferimento

Davies CS. Statistical Methods for the Analysis of Repeated Measurements. Springer, 2002

Chang M. Adaptive design theory and implementation using SAS and R. Chapman and Hall, 2008

Harrell F. Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer, 2015

Slide delle lezioni e materiale integrativo disponibili sulla piattaforma della didattica online <http://elearning.unimib.it/>.

Il materiale è identico per frequentanti e non frequentanti

Lingua di insegnamento

L'insegnamento è completamente erogato in italiano. La maggior parte del materiale su cui gli studenti dovranno lavorare per la preparazione della tesina (ricerca bibliografica di articoli scientifici e testi) sono in inglese. La tesina può essere scritta in inglese.

9.13 Metodologia della Ricerca Clinica ed Epidemiologica (blended)

<i>Codice insegnamento:</i>	F8203B001
<i>Docente/i:</i>	Giovanni Corrao, Matteo Franchi, Annalisa Biffi, Federico Rea
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	12
<i>Periodo:</i>	I Semestre (Ciclo I-II) e II Semestre (Ciclo III-IV)
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	90

Obiettivi formativi

Il corso si propone di introdurre le fonti di incertezza del processo clinico, dalla formulazione della diagnosi alla scelta della terapia e/o dell'intervento di prevenzione; di permettere allo studente di esaminare criticamente le misure proposte in letteratura sulla validità della diagnosi, sul disaccordo clinico, sulla frequenza della malattia e dei suoi possibili esiti, sulla associazione tra determinanti e rischio di malattia, sulla efficacia e sull'impatto degli interventi terapeutici e preventivi; di approfondire gli aspetti legati al disegno dello studio mettendo lo studente nella condizione di saper leggere e indagare criticamente la letteratura scientifica di riferimento e di essere in grado di redigere il protocollo di uno studio clinico sperimentale od osservazionale, così come di una meta-analisi.

Conoscenza e comprensione

Questo Corso di Studio fornirà conoscenze e capacità di comprensione relativamente a:

- Riconoscere le fonti di incertezza del processo clinico (dalla formulazione della diagnosi alla scelta della terapia e/o intervento preventivo)
- Principali misure o indicatori di sintesi riguardanti la validità della diagnosi, il disaccordo clinico, la frequenza della malattia e dei suoi possibili esiti
- Errori casuali e sistematici delle misure cliniche
- Misure di associazione tra determinanti e rischio di malattia, di efficacia e di impatto clinico di interventi terapeutici e preventivi
- Principali disegni di studi clinici sperimentali od osservazionali

Capacità di applicare conoscenza e comprensione

Alla fine del Corso di Studio gli studenti saranno in grado di:

- Effettuare criticamente una ricerca nella letteratura scientifica
- Interpretare le misure di associazione tra determinanti e rischio di malattia, di efficacia ed impatto clinico di interventi (terapeutici e/o preventivi), di validità della diagnosi e di disaccordo clinico
- Scegliere l'opportuno disegno di studio in funzione del quesito clinico cui si vuole rispondere, ed impostare il protocollo di uno studio clinico sperimentale od osservazionale, così come di una meta-analisi

Contenuti sintetici

1. Incertezza del processo diagnostico
2. Frequenza degli eventi clinici
3. Errori casuali e sistematici delle misure diagnostiche

4. Studi osservazionali e sperimentali
5. Meta-analisi

Programma esteso

1. Misure di incertezza del processo diagnostico

Incetezza del processo diagnostico (Introduzione al processo diagnostico; Accuratezza e precisione delle misure cliniche; Validità e riproducibilità del giudizio clinico).

Caratteristiche operative di un test diagnostico (Validità di un test diagnostico; Test espressi con una scala continua e la curva ROC; Scelta del test).

Predittività di un test diagnostico (Valore predittivo; Valore predittivo positivo e prevalenza; Valore predittivo positivo e ragionamento clinico).

Test multipli (Combinazione di test; Test in serie; Test in parallelo).

Accordo tra osservatori (Un richiamo ai concetti di validità e riproducibilità; La misura della concordanza; Quanto frequentemente si verifica il disaccordo tra clinici?).

2. Misure di frequenza degli eventi clinici e dei loro determinanti

Frequenza degli eventi clinici (I concetti di incidenza e prevalenza; Le misure di insorgenza (incidenza)).

Determinanti degli eventi clinici (Introduzione; Misure di associazione; Misure di impatto).

Efficacia delle azioni terapeutiche (Introduzione; Forza dell'efficacia; Bilancio tra benefici e rischi del trattamento).

3. Errori casuali e sistematici delle misure cliniche

Precisione delle misure (Un richiamo ai concetti di precisione e validità; Precisione, variabilità casuale e intervalli di confidenza; Misure descrittive; Misure di associazione; Riepilogo).

Validità delle misure (Un richiamo al concetto di validità; Distorsione da selezione; Distorsione da osservazione).

Il ruolo del confondimento (Introduzione; Definizione di confondente; Metodi di controllo del confondimento).

4. Studi osservazionali e sperimentali

La ricerca clinica tra sperimentazione e osservazione (Gerarchia delle evidenze; Studi non controllati; Studi controllati con controlli storici; Studi controllati con controlli concorrenti: l'epidemiologia analitica; Sperimentazioni cliniche controllate e randomizzate; Meta-analisi di sperimentazioni cliniche controllate e randomizzate).

Studi di coorte (Razionale e finalità; Aspetti legati alla pianificazione dello studio; Aspetti legati all'analisi dei dati; Aspetti legati alla comunicazione dei risultati; Punti di forza e debolezze del disegno osservazionale di coorte).

Studi caso-controllo (Definizione; Il rationale degli studi caso-controllo; La misura dell'associazione negli studi caso-controllo; Aspetti legati alla pianificazione dello studio; Aspetti legati all'analisi dei dati; Aspetti legati alla comunicazione dei risultati; Punti di forza e debolezze del disegno osservazionale caso-controllo).

Sperimentazioni cliniche (Razionale delle sperimentazioni cliniche; Cenni storici sulla sperimentazione clinica dei farmaci; Le fasi della sperimentazione clinica di un farmaco; Aspetti etici delle sperimentazioni cliniche; Aspetti legati alla pianificazione dello studio; Aspetti legati all'analisi dei dati; Aspetti legati alla comunicazione dei risultati; Punti di forza e debolezze del metodo sperimentale).

5. Revisioni sistematiche e modelli meta-analitici

Introduzione alle revisioni sistematiche e alle meta-analisi, dalla raccolta dei dati alle stime meta-analitiche utilizzando modelli ad effetti fissi o modelli ad effetti casuali. Cenni sui principali metodi per il controllo

delle distorsioni. Linee Guida: dalla ricerca sistematica della letteratura alla formulazione delle raccomandazioni adottate nella pratica clinica.

Prerequisiti

Nessun prerequisito formale richiesto.

Metodi didattici

E-learning/Lezioni frontali.

La maggior parte del corso di svolgerà secondo la metodologia e-learning. Sulla pagina dedicata al corso (sulla apposita piattaforma online di Ateneo) verrà regolarmente caricato dal docente del corso il materiale didattico (dispense, letture e materiale di approfondimento a seconda degli argomenti trattati). Al termine di ognuno dei 5 moduli in cui il corso è suddiviso, è prevista una lezione frontale in cui verranno ripresi e chiariti in aula i principali concetti introdotti nel modulo stesso, con l'obiettivo di chiarire eventuali dubbi ed approfondire alcuni argomenti su richiesta degli studenti.

Modalità di verifica dell'apprendimento

L'esame consiste in una prova orale finale durante la quale gli studenti verranno invitati a ragionare insieme al docente sui principali argomenti trattati durante il corso. Le domande avranno l'obiettivo di verificare la conoscenza della componente teorica del corso, oltre che di valutare la capacità di scelta, di calcolo e commento delle misure sintetiche presentate durante il corso, oltre che dei principali disegni di studio. Inoltre, le domande consentiranno di verificare anche la capacità dello studente di esprimersi con un adeguato linguaggio tecnico.

Non sono previsti distinzioni, in fase di esame, per studenti frequentanti e studenti non frequentanti

Inoltre, al termine di ognuno dei moduli in cui il corso è articolato, e dopo la lezione frontale riassuntiva di fine modulo, verranno svolte prove *in itinere* al fine di verificare l'apprendimento dello studente: tali prove saranno composte da alcune domande di teoria e da esercizi numerici. Verrà caricato sulla pagina e-learning del corso un assegnamento (composto da esercizi e da domande teoriche, a seconda degli argomenti trattati nel relativo modulo) che gli studenti dovranno svolgere e poi caricare sulla piattaforma. Ogni assegnamento verrà valutato con un punteggio (da 1 a 5) e servirà al docente per monitorare la frequenza di partecipazione al corso e per verificare l'apprendimento degli studenti.

Tuttavia, la partecipazione alla prova orale finale non è vincolata dagli assegnamenti in itinere.

Testi di riferimento

Data la vastità degli argomenti trattati, non è previsto un testo di riferimento specifico.

Per ogni argomento è disponibile sulla piattaforma e-learning il materiale didattico della parte del corso erogato in e-learning (testi scritti, esercizi di autoverifica, programmi di calcolo, e articoli scientifici).

Lingua di insegnamento

Italiano

9.14 Modelli Statistici Applicati alle Sperimentazioni Cliniche

<i>Codice insegnamento:</i>	F8203B036
<i>Docente/i:</i>	Vincenzo Bagnardi
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, I Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	42

Obiettivi formativi

Obiettivo dell'insegnamento è approfondire gli aspetti statistici legati alla pianificazione e all'analisi di uno studio clinico.

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- Implementazione di simulazioni con il linguaggio di programmazione SAS;
- Pianificazione della dimensione campionaria e della potenza statistica di uno studio clinico;
- Strategia di analisi statistica di uno studio clinico nel caso di outcome di sopravvivenza, eventi ripetuti, outcome continui misurati sia all'inizio che alla fine dello studio, dati correlati;
- Le fasi di una sperimentazione clinica.

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- Impostare in autonomia delle simulazioni nel linguaggio di programmazione SAS;
- Riconoscere gli elementi fondamentali necessari per il calcolo della dimensione campionaria e della potenza statistica di uno studio clinico;
- Analizzare studi clinici con outcome di sopravvivenza, eventi ripetuti, outcome misurati prima e dopo la fine dello studio, dati correlati.

L'insegnamento consente allo studente di acquisire solide basi nella pianificazione e analisi di uno studio clinico con il supporto del software di programmazione SAS.

Contenuti sintetici

- Simulazioni di dati in SAS
- Approfondimenti sulla regressione di Poisson e Negativa Binomiale per l'analisi di eventi ripetuti
- Dimensione campionaria e potenza statistica di uno studio clinico: un approccio basato sulle simulazioni
- Introduzione alle fasi della sperimentazione clinica
- Strategie per l'analisi di studi con outcome continuo misurato sia all'inizio che alla fine dello studio
- Analisi di dati correlati

Programma esteso

1. Simulazioni di dati in SAS

- 1.1 Importanza delle simulazioni come strumento di lavoro per il biostatistico
- 1.2 Elementi e tecniche di base per la simulazione dei dati
- 1.3 Utilizzo delle simulazioni per valutare distribuzioni campionarie, validità di tecniche statistiche e proprietà di un disegno statistico

- 2. Modelli per l'analisi di eventi ripetuti ed eventi singoli in uno studio clinico**
 - 2.1 Modello di Poisson e modello negativo-binomiale per l'analisi di eventi ripetuti
 - 2.2 Modello di Poisson e modello esponenziale per l'analisi di eventi singoli
 - 2.3 Modello di Poisson e analisi della sopravvivenza
- 3. Dimensione campionaria e potenza statistica di uno studio clinico: un approccio basato sulle simulazioni**
- 4. Elementi di metodologia della ricerca clinica**
 - 4.1 Le fasi della sperimentazione clinica: modelli statistici per gli studi di fase I, II e III
 - 4.2 Studio di superiorità e studio di non-inferiorità
 - 4.3 Studi adattativi
- 5. Strategie per l'analisi di studi con outcome continuo misurato sia all'inizio che alla fine dello studio**
- 6. Analisi di dati correlati**
 - 6.1 Introduzione all'analisi dei dati correlati
 - 6.2 Modelli lineari a effetti misti per l'analisi di risposte continue correlate
 - 6.3 Modelli lineari generalizzati a effetti misti per l'analisi di risposte binarie correlate

Prerequisiti

Nessuno

Metodi didattici

Lezioni

Esercitazioni al computer con applicazioni in SAS

Modalità di verifica dell'apprendimento

Prova scritta - project work (solo per frequentanti)

Durante il corso saranno proposti alcuni problemi che dovranno essere affrontati e risolti con uno studio di simulazione. Le simulazioni saranno impostate in aula da gruppi di lavoro formati da due persone. Ogni studente poi, individualmente, dovrà approfondire uno dei problemi proposti e dovrà scrivere una relazione in cui presenterà i risultati dello studio seguendo lo stile di un lavoro originale di ricerca.

Non sono previste prove di esame in itinere.

Prova orale

Valutazione sui contenuti delle slide del corso, e dei capitoli di libro e articoli che saranno indicati e messi a disposizione durante il corso.

La prova orale consente di verificare le conoscenze dei principali metodi legati alla pianificazione e all'analisi di uno studio clinico.

Testi di riferimento

Tutto il materiale didattico (estratti di libri, articoli, codice SAS, dataset) distribuito agli studenti durante il corso

Lingua di insegnamento

Italiano

9.15 Modelli Statistici e Inferenza Bayesiana

9.15.1 Inferenza Bayesiana

<i>Codice insegnamento:</i>	F8203B011
<i>Docente/i:</i>	Fulvia Pennoni
<i>Settore scientifico disciplinare:</i>	MAT/06
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, II Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	42

Obiettivi formativi

Il corso fornisce le conoscenze dei principi di base per l'inferenza statistica in ambito Bayesiano. Il ragionamento Bayesiano viene presentato in modo integrato con l'approccio tradizionale dell'inferenza statistica.

Conoscenza e comprensione

Questo insegnamento permette allo studente di apprendere:

- la regola di Bayes e l'utilizzo della probabilità per aggiornare l'informazione fornita dai dati osservati;
- gli elementi fondamentali dell'inferenza Bayesiana: il calcolo delle distribuzioni a priori, della verosimiglianza e della distribuzione a posteriori;
- il metodo Monte Carlo per la simulazione della distribuzione a posteriori;
- il calcolo della distribuzione predittiva;
- gli algoritmi Markov Chain Monte Carlo: Metropolis-Hastings e Gibbs sampler;
- il modello di regressione lineare multipla ed il modello di regressione logistica multipla in termini Bayesiani;
- i modelli di Markov per dati longitudinali.

Capacità di applicare conoscenza e comprensione

Questo insegnamento permette allo studente:

- applicare i metodi di inferenza Bayesiana a casi di studio rilevanti nell'ambito della biostatistica, dell'epidemiologia, della medicina, della biologia, dell'ambiente, della genetica e la salute pubblica;
- applicare i modelli statistici utilizzando dati ripetuti nel tempo per le stesse unità;
- applicare metodi di classificazione basati su modelli statistici;
- sviluppare del codice in ambiente R e SAS;
- Creare report riproducibili come strumento di presentazione dei risultati delle analisi.

Le lezioni teoriche sono affiancate da esercitazioni pratiche su dati reali e simulati in cui si utilizza sia l'ambiente R, Rstudio e Rmarkdown che il software SAS. In tal modo lo studente impara ad utilizzare due diversi linguaggi di programmazione.

L'insegnamento fornisce i concetti principali dell'inferenza Bayesiana, un metodo statistico essenziale nell'ambito teorico e dell'analisi dei dati per i contesti lavorativi di sbocco (biostatistico/statistico/demografico e affini) degli studenti del corso di laurea in Biostatistica. Il corso risulta indispensabile per il successivo percorso universitario.

Contenuti sintetici

Introduzione all'inferenza Bayesiana e alla regola di Bayes.

Metodi di specificazione del modello e delle distribuzioni a priori.

Determinazione della distribuzione a posteriori con metodi esatti, famiglie coniugate: Gaussiana, Poisson-Gamma, Beta-Binomiale, Multinomiale-Dirichlet.

Inferenza Bayesiana non parametrica.

Metodi di sintesi della distribuzione a posteriori, intervalli di credibilità e intervalli con la massima densità a posteriori.

Introduzione ai processi stocastici di Markov e proprietà. Modello passeggiata casuale.

Modello di transizione per dati longitudinali.

Modello di Markov a variabili latenti per dati longitudinali con covariate.

Metodi Markov Chain Monte Carlo: Algoritmo Metropolis-Hastings e campionamento Gibbs.

Ambiente R e Rstudio, utilizzando principalmente le seguenti librerie: probBayes, learnBayes, LMest. RMarkdown attraverso la libreria knitr per integrare codice e output. Software SAS: proc MCMC.

Programma esteso

Il corso comprende un'introduzione all'inferenza Bayesiana e il confronto con l'inferenza classica. Viene ripresa la regola di Bayes e la regola delle probabilità totali attraverso l'esempio del Bayes'billard. Vengono sviluppati gli aspetti di specificazione delle distribuzioni a priori, la stima esatta delle distribuzioni a posteriori e l'interpretazione dei modelli Bayesiani. Viene introdotto il modello Beta-Binomiale ed illustrato anche l'approccio Bayesiano non parametrico. Enfasi viene posta anche sulla distribuzione predittiva. Vengono illustrate le caratteristiche di scelta e di determinazione delle distribuzioni a priori: esempi e convenienza della famiglia coniugata. Si considera anche la scelta delle distribuzioni a priori non informative e la nozione di scambiabilità è illustrata attraverso il teorema di rappresentazione di De Finetti.

Vengono trattati i metodi di sintesi della distribuzione a posteriori: intervalli di credibilità, intervalli con la massima densità a posteriori. Famiglia coniugate: Beta-Binomiale e Gaussiana, modello Poisson-Gamma. Introduzione alla distribuzione multinomiale e di Dirichlet. La teoria viene affiancata da svariati esempi di applicazione dei modelli Bayesiani nell'ambito della biostatistica attraverso dati reali e simulati riguardanti l'epidemiologia, la farmaco epidemiologia, la medicina e la biologia oltre che l'ecologia e le scienze ambientali.

Vengono introdotti i processi stocastici Markoviani le proprietà e le caratteristiche delle catene di Markov. Si mostra il processo passeggiata casuale e anche attraverso la simulazione delle sue traiettorie. Viene introdotto il modello di transizione per dati longitudinali, ed il modello latente di Markov. Vengono illustrati anche da un punto di vista computazionale gli algoritmi di stima maggiormente utilizzati nell'ambito del metodo Markov Chain Monte Carlo (MCMC): l'algoritmo Metropolis-Hastings e l'algoritmo Gibbs sampling. Vengono discusse diverse misure che permettono la valutazione diagnostica della loro convergenza.

La teoria è affiancata da numerose applicazioni a dati reali e simulati riguardanti gli ambiti della biostatistica in modo da facilitare anche lo sviluppo della conoscenza della semantica in ambiente R e del software SAS. Gli esempi sono svolti in Rstudio con l'ausilio di RMarkdown. Lo studente durante le esercitazioni, è incoraggiato, anche tramite l'apprendimento cooperativo, ad elaborare documenti riproducibili e a sviluppare il commento ai risultati delle analisi in modo critico. Nelle ultime settimane viene spiegato l'utilizzo delle procedure SAS per la stima Bayesiana dei modelli di regressione lineare a logistica.

Prerequisiti

Si consiglia di riprendere le nozioni impartite nei seguenti corsi: Statistica, Probabilità e Inferenza Statistica, Modelli Statistici II.

Metodi didattici

Sono previste lezioni frontali riguardanti la parte teorica sui concetti di base dell'inferenza Bayesiana e dei modelli di Markov per dati longitudinali. Le lezioni di teoria sono affiancate da esercitazioni pratiche che permettono allo studente di sviluppare l'aspetto della scienza dei dati. Le lezioni si tengono presso il laboratorio informatico. Settimanalmente vengono assegnati esercizi di riepilogo da svolgere con dati reali o simulati relativi alla parte di programma svolto. Durante il corso con l'ausilio di R nell'ambiente RStudio e del marcatore di testo RMarkdown oppure del software SAS, gli studenti imparano ad elaborare documenti riproducibili per le analisi dei dati e la stima dei modelli proposti. Gli stessi vengono incoraggiati ad affrontare il problema applicativo con lo scopo ulteriore di sviluppare l'apprendimento cooperativo.

Modalità di verifica dell'apprendimento

L'esame è in forma scritta con orale facoltativo, non sono previste prove intermedie. Le seguenti modalità di verifica dell'apprendimento riguardano sia gli studenti frequentanti che non frequentanti. L'esame scritto ha durata complessiva massima di due ore e si svolge presso il laboratorio informatico. Durante la prova occorre risolvere gli esercizi applicati alla luce degli argomenti teorici sviluppati durante il corso e rispondere ad alcune domande di teoria. Le analisi sono condotte tramite l'ambiente R, Rstudio e RMarkdown e SAS. Gli esercizi permettono di verificare la capacità di comprensione del problema, la sua risoluzione tramite l'applicazione dei Bayesiani e di modelli per dati longitudinali avanzati a dati reali o simulati e l'elaborazione di report in cui si descrive il procedimento e si illustrano i risultati. Le domande di teoria permettono di verificare l'apprendimento delle nozioni teoriche impartite durante il corso.

Testi di riferimento

Il materiale didattico è costituito principalmente dalle dispense redatte dal docente. Queste riguardano sia gli argomenti di teoria che le applicazioni sviluppate con il software R oppure SAS. Il predisposto dal docente viene reso disponibile nella pagina della piattaforma e-learning dell'ateneo dedicata al corso. Il docente pubblica al termine di ogni lezione anche i file riferiti al materiale didattico che comprende: le slides, i programmi di calcolo, gli esercizi, i dataset, e le soluzioni degli esercizi. Nella stessa pagina vengono pubblicati anche alcuni precedenti testi d'esame.

I principali testi di riferimento sono elencati nella bibliografia delle dispense. Alcuni tra questi anche disponibili in e-book i seguenti:

Albert, J. (2009). *Bayesian computation with R*. Springer Science & Business Media.

Albet, J., Hu, J. (2019). *Probability and Bayesian modeling*. Chapman and Hall/CRC.

Bartolucci, F., Farcomeni, A., Pennoni, F. (2013). *Latent Markov Models for longitudinal data*, Chapman and Hall/CRC, Boca Raton.

Migon, H. S., Gamerman, D., Louzada, F. (2014). *Statistical inference: an integrated approach*. Chapman & Hall.

Penioni, F. (2021). *Dispensa di Inferenza Bayesiana -parte di teoria e applicazioni con R e SAS*. Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca.

Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods* (second edition). Springer-Verlag, New York.

Dipak, D. K., Ghosh, S. K., Mallick, B. K. (2000). *Generalized linear models: A Bayesian perspective*. CRC press.

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

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Lingua di insegnamento

Il corso viene erogato in lingua italiana. Gli studenti Erasmus possono utilizzare il materiale didattico in Inglese e richiedere al docente lo svolgimento della prova d'esame in lingua inglese.

9.15.2 Modelli Statistici II

<i>Codice insegnamento:</i>	F8203B011
<i>Docente/i:</i>	Fulvia Pennoni
<i>Settore scientifico disciplinare:</i>	SECS-S/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, I Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	42

Obiettivi formativi

Il corso introduce alle procedure analitiche ed inferenziali condotte tramite modelli statistici avanzati e alle simulazioni con l'intento di sviluppare una conoscenza critica delle assunzioni dei modelli alla base della teoria.

Conoscenza e comprensione

Questo insegnamento permette allo studente di:

- analizzare i dati con modelli statistici sviluppati sia per variabili risposta categoriali che continue
- sviluppare in modo autonomo le simulazioni;
- servirsi della semantica del software R anche tramite l'ambiente RMarkdown per creare dei documenti riproducibili contenenti il codice, risultati ed i commenti delle analisi;
- interpretare i risultati delle elaborazioni in modo rigoroso e fornire una descrizione esaustiva e chiara degli stessi per finalità divulgative.

Capacità di applicare conoscenza e comprensione

Questo insegnamento permette allo studente:

- sviluppare l'inferenza statistica tramite le moderne tecniche del bootstrap;
- stimare, selezionare ed interpretare i modelli di miscugli di distribuzioni per popolazioni eterogenee;
- adattare i modelli a variabili latenti;
- applicare le conoscenze teoriche per l'analisi dei dati raccolti in svariati ambiti tra cui l'epidemiologia, la medicina, la biologia, la genetica e la salute pubblica.
- sviluppare del codice in ambiente R.

Lo studente viene incoraggiato a presentare la teoria ed i risultati delle applicazioni in modo organico.

L'insegnamento fornisce i concetti essenziali per lo sviluppo dei metodi statistici parametrici e non parametrici sia in ambito teorico che applicato per i contesti lavorativi di sbocco degli studenti del corso di laurea in Biostatistica (biostatistico/statistico/demografico e affini). L'insegnamento risulta pertanto indispensabile per il successivo percorso universitario.

Contenuti sintetici

Nella prima parte del corso vengono impartiti i concetti di base per simulare delle realizzazioni da variabili casuali. Vengono introdotte le principali procedure di ricampionamento: bootstrap e Jackknife e la loro applicazione nell'ambito dell'inferenza statistica.

L'algoritmo Expectation-Maximization (EM) viene introdotto come metodo di imputazione dei dati mancanti attraverso la stima dei parametri di modelli lineare generalizzati. Viene illustrato il suo utilizzo per

la stima dei parametri dei modelli miscuglio (finite mixture models) e a variabili latenti. Lo studente approfondisce le competenze nell'utilizzo della semantica del software R.

Programma esteso

La prima parte del corso riguarda i metodi di simulazione e concerne anche i metodi lineari congruenziali per la generazione di numeri pseudo-casuali, i test grafici e statistici (tra cui il test Kolmogorov-Smirnov e il test Chi-Quadrato) per la verifica della pseudo-casualità. Vengono esaminati alcuni metodi per la generazione di realizzazioni da variabili casuali tra cui metodo della trasformata inversa. La teoria è affiancata da esempi applicativi utilizzando diversi modelli distributivi: la distribuzione esponenziale, binomiale e di Gauss. Si considera la convoluzione di variabili casuali e la generazione di realizzazioni dalla stessa.

Nella seconda parte si introducono i principali metodi di ricampionamento: jackknife e bootstrap. Si illustra l'utilizzo del bootstrap per l'inferenza statistica tramite gli intervalli di confidenza ottenuti con il metodo del percentile e con la correzione per la distorsione. L'algoritmo Expectation-Maximization viene illustrato dettagliatamente sia come algoritmo di stima dei parametri dei modelli a variabili latenti sia come metodo per l'imputazione dei valori mancanti in una tabella a doppia entrata riferita ai conteggi tramite un modello lineare generalizzato. Si introducono i modelli miscuglio per variabili risposta sia quantitative che categoriali, considerando in particolare i miscugli con componenti Gaussiane. Si illustra la stima dei modelli miscuglio con l'algoritmo Expectation-Maximization e l'interpretazione dei risultati comprendente la stima della densità e la classificazione delle unità statistiche con le probabilità a posteriori.

La teoria è affiancata da numerose applicazioni per l'analisi di dati reali e simulati riguardanti gli ambiti della biostatistica in modo da facilitare anche l'apprendimento dell'ambiente R con l'ausilio del marcatore di testo RMarkdown. Le principali librerie utilizzate sono boot, bootstrap e mclust. Lo studente è incoraggiato ad elaborare documenti riproducibili e sviluppare i commenti ai risultati delle analisi in modo critico anche tramite apprendimento cooperativo.

Prerequisiti

Per una più agevole comprensione dei contenuti del corso è consigliato conoscere le nozioni di Probabilità e di Inferenza Statistica. Lo studente deve inoltre conoscere la semantica di base del linguaggio di programmazione in ambiente R.

Metodi didattici

Sono previste lezioni frontali riguardanti la parte teorica sui concetti di base dei modelli statistici. Le lezioni di teoria sono affiancate da esercitazioni pratiche che permettono allo studente di sviluppare la capacità di analisi dei dati. Le lezioni si svolgono in laboratorio informatico. Vengono assegnati ogni settimana degli esercizi di riepilogo basati sull'applicazione dei modelli proposti a dati reali o simulati relativi alla parte di programma svolto. Durante il corso con l'ausilio di R nell'ambiente RStudio e l'interfaccia di RMarkdown, gli studenti imparano ad elaborare documenti riproducibili. Gli stessi vengono incoraggiati ad affrontare il problema applicativo con lo scopo ulteriore di sviluppare l'apprendimento cooperativo.

Modalità di verifica dell'apprendimento

Le seguenti modalità di verifica dell'apprendimento riguardano sia gli studenti che non frequentanti. L'esame è in forma scritta con orale facoltativo, non sono previste prove intermedie. L'esame scritto ha durata complessiva massima di due ore e si svolge presso il laboratorio informatico. Durante la prova occorre risolvere gli esercizi applicati alla luce degli argomenti teorici sviluppati durante il corso e rispondere ad

alcune domande di teoria. Le analisi sono condotte tramite l'ambiente R, Rstudio e RMarkdown. Gli esercizi permettono di verificare la capacità di comprensione del problema, la sua risoluzione tramite l'applicazione di modelli statistici avanzati a dati reali o simulati e l'elaborazione di report in cui si descrive il procedimento e si illustrano i risultati. Le domande di teoria permettono di verificare l'apprendimento delle nozioni teoriche impartite durante il corso.

Testi di riferimento

Il materiale didattico è costituito principalmente dalle dispense redatte dal docente. Queste riguardano sia gli argomenti di teoria che le applicazioni sviluppate con il software R oppure SAS. Il predisposto dal docente viene reso disponibile nella pagina della piattaforma e-learning dell'ateneo dedicata al corso. Il docente pubblica al termine di ogni lezione anche i file riferiti al materiale didattico che comprende: le slides, i programmi di calcolo, gli esercizi, i dataset, e le soluzioni degli esercizi. Nella stessa pagina vengono pubblicati anche alcuni precedenti testi d'esame.

I principali testi di riferimento sono elencati nella bibliografia delle dispense, tra gli altri si segnalano i seguenti. Alcuni tra questi anche disponibili in ebook i seguenti:

Bartolucci, F., Farcomeni, A., Pennoni, F. (2013). *Latent Markov Models for longitudinal data*, Chapman and Hall/CRC, Boca Raton.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis: theory and practice*. Springer Science & Business Media, New York.

Blitzstein J. K. and Hwang J. (2014). *Introduction to probability*, Chapman & Hall/CRC.

Gentle, J. E., Hardle W., Mori Y. (2004). *Handbook of computational statistics*. Springer-Berlin.

Lange, K. (2010). *Numerical analysis for statisticians*, 2nd Edition, Springer, New York.

Pennoni, F. (2021). *Dispensa di Modelli Statistici II, parte di teoria e applicazioni con R*. Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca.

Rizzo M. L. (2008). *Statistical Computing with R*, Chapman & Hall/CRC, New York.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Lingua di insegnamento

Il corso viene erogato in lingua italiana. Gli studenti Erasmus possono utilizzare il materiale didattico in Inglese e richiedere al docente la prova d'esame in lingua inglese.

9.16 Modelli Statistici I

9.16.1 Modello Lineare Generalizzato

<i>Codice insegnamento:</i>	F8203B010
<i>Docente/i:</i>	Giorgio Vittadini, Daniele Spinelli
<i>Settore scientifico disciplinare:</i>	SECS-S/01, SECS-S/02
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, III ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	47

Obiettivi formativi

Il corso ha quale obiettivo lo studio di modelli avanzati, partendo dal modello lineare classico fino ad arrivare ai modelli multivariati.

Contenuti sintetici

Il corso ha quale obiettivo lo studio di modelli più avanzati del modello lineare classico. Si presentano perciò
modelli lineari generalizzati,
modelli lineari multivariati
modelli multilevel

L'attività formativa è svolta attraverso lezioni teoriche e lezioni pratiche in laboratorio. Il materiale del corso e ulteriori informazioni verranno riportate sulla pagina web dedicata nella piattaforma e-learning unimib:

<http://elearning.unimib.it/>.

Programma esteso

Il corso ha quale obiettivo l'introduzione alla specificazione, stima e verifica di modelli interpretativi dei dati di tipo lineare più avanzati del modello lineare classico. Si presentano perciò

- Modelli lineari generalizzati che non rispettano le ipotesi del modello lineare classico: modelli con errori eteroschedastici e correlati, modelli non lineari, trattamento di outlier
- Modelli lineari multivariati di tipo classico e non
- Modelli multilevel

Ciascun ambito sarà l'oggetto specifico di un modulo del corso. L'attività formativa è svolta attraverso lezioni teoriche e lezioni pratiche in laboratorio statistico-informatico nelle quali si affronteranno analisi su casi empirici mediante l'uso del software SAS. Il materiale del corso (sia delle lezioni teoriche sia delle lezioni pratiche) e ulteriori informazioni verranno riportate sulla pagina web dedicata nella piattaforma e-learning unimib: <http://elearning.unimib.it/>.

Prerequisiti

Si richiede una buona conoscenza della

Statistica descrittiva univariata : indici di posizione; indici di variabilità; indici di simmetria e di curtosi.

Statistica descrittiva bivariata: connessione, dipendenza in media, correlazione lineare, regressione lineare bivariata, multipla, multivariata, polinomiale, non lineare.

Teoria della probabilità: popolazione e campione; significato di probabilità nella versione classica ; elementi di calcolo combinatorio; tipi di campionamento; distribuzioni di variabili casuali univariate; variabili casuali Normale , t di Student, F d Snedecor ; distribuzioni casuali campionarie

Inferenza: teoria della stima, proprietà dello stimatore puntuale; stima intervallare; verifica di ipotesi, test di ipotesi di Neyman Pearson; test di ipotesi sulle medie basati su Normale , t di Student; test d ipotesi sulla varianza.

Modello lineare classico: ipotesi; stima dei parametri del modello nel campione e nella popolazione; proprietà degli stimatori dei minimi quadrati; test di ipotesi sui parametri basati su Normale , t di Student, ; test di ipotesi sul modello su gruppi di parametri , su un parametro basata F di Snedecor

Package statistici R e SAS

Si suggerisce a chi non provenga da corsi triennali di statistica o economia di seguire preventivamente i corsi introduttivi del corso di laurea di biostatistica calcolo delle probabilità, introduzione all' inferenza statistica, introduzione ai modelli statistici, modelli statistici per dati categoriali e conoscano i pacchetti statistici R o SAS.

Metodi didattici

Le lezioni si distinguono in parte teorica e parte applicata. Durante la parte teorica vengono presentate i framework metodologici relativi al corso, che vengono poi applicati durante le lezioni pratiche in laboratorio. In laboratorio si utilizza il software SAS, e si apprende la stesura del codice e la lettura degli output dei modelli. Lezioni ed esercitazioni saranno registrate sulla piattaforma e-learning

Modalità di verifica dell'apprendimento

L'esame si svolge attraverso una prova da sostenere presso il laboratorio informatico e consiste in due domande di teoria e un esercizio pratico. L'esercizio riguarda uno dei temi proposti durante le esercitazioni svolte a lezione e riguarda la risoluzione di un problema tramite il software R o SAS e il commento ai risultati.

Testi di riferimento

Johnston, J. (1993), *Econometrica*, 3a edizione, Franco Angeli, Milano

Freund, R. J., Wilson, W. J., and Sa, P. (2006), *Regression Analysis: Statistical Modeling of a Response Variable*, 2nd edition, Academic Press

Baltagi B. H. (2008), *Econometrics*, fourth Edition, Springer Berlin

Littell, R. C., Freund, R. J., and Spector, P. C. (2002), *SAS for Linear Models*, 4th Edition, Cary, NC: SAS Institute Inc.

Manual SAS/STAT 9.2

Manual SAS/STAT 9.3

Manual SAS/ETS 9.3

Lingua di insegnamento

Italiano

9.16.2 Modelli Lineari per Dati Categoriali

<i>Codice insegnamento:</i>	F8203B010
<i>Docente/i:</i>	Nadia Solaro
<i>Settore scientifico disciplinare:</i>	SECS-S/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, II Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	42

Obiettivi formativi

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

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

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

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

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

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

Contenuti sintetici

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

Programma esteso

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

Prerequisiti

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

Metodi didattici

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

Modalità di verifica dell'apprendimento

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

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

Le domande pratiche riguardano sia l'individuazione, la costruzione e l'utilizzo della modellistica più opportuna con riferimento a problemi tratti da situazioni e da dati reali, sia la definizione del disegno dell'analisi più adeguato a soddisfare obiettivi di studio definiti a priori. L'analisi statistica di dati da preparare prima e presentare poi in sede d'esame costituisce la parte dell'esame in cui questi aspetti hanno

maggior enfasi, poiché richiede allo studente di lavorare in modo critico e in piena autonomia, soprattutto nella definizione e nella realizzazione degli obiettivi dello studio. Le domande pratiche consentono in definitiva di verificare la capacità di comprensione delle problematiche sottoposte e di proporre soluzioni in termini di analisi, la competenza nel leggere e interpretare i risultati delle analisi, e l'abilità nel realizzare le analisi richieste mediante le procedure di SAS.

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

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

Testi di riferimento

- Materiale didattico della docente pubblicato sul sito e-learning del corso (ad accesso riservato con password)
- Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, New York: John Wiley & Sons
- Dobson, A. (1990), *An Introduction to Generalized Linear Models*, London: Chapman & Hall
- Littell, R. C., Freund, R. J., and Spector, P. C. (2002), *SAS for Linear Models*, 4th Edition, Cary, NC: SAS Institute Inc.
- Searle, S. R. (1971), *Linear Models*, New York: John Wiley & Sons

Lingua di insegnamento

Italiano

9.17 Modelli Statistici per la Genetica

<i>Codice insegnamento:</i>	F8203B017
<i>Docente/i:</i>	Mariacristina Monti; Marco Gnesi
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, IV Ciclo
<i>Tipo di attività:</i>	Opzionale
<i>Ore:</i>	44

Obiettivi formativi

Il corso si propone di fornire le conoscenze di base per un approccio statistico rigoroso per il mappaggio e l'identificazione di loci implicati in patologie o caratteri nell'uomo. Alla fine del corso lo studente avrà appreso gli elementi per comprendere ed utilizzare strumenti base di statistica genetica, metodi di epidemiologia genetica, e saprà quindi leggere criticamente un articolo scientifico e interpretare i risultati derivanti da analisi statistiche di dati genetici.

Contenuti sintetici

Mendel vs Malattie genetiche complesse, Analisi di segregazione, Analisi di linkage, Analisi di Associazione genetica, Dimensione campionaria e potenza, Analisi genome-wide e analisi di dati di sequenziamento

Programma esteso

Mendel e le Malattie genetiche complesse;

Equilibrio di Hardy-Weinberg;

Fattori che "complicano" l'identificazione del tipo di ereditarietà;

Analisi di segregazione;

Analisi di linkage: fondamenti teorici e strategie per il mappaggio genetico di patologie mendeliane e di tratti complessi;

Analisi di associazione: Linkage disequilibrium, Studi casi-controllo, Studi familiari: TDT-Trasmission Disequilibrium Test

Gli studi GWA (Genome Wide Analyses) e di sequenziamento

Dimensione campionaria e potenza di uno studio

Prerequisiti

Nessuno

Metodi didattici

Il corso è organizzato in lezioni frontali ed esercitazioni con software ad hoc mirate tanto all'applicazione dei concetti teorici presentati su set di dati sperimentali, quanto all'interpretazione/comprendimento delle evidenze scientifiche derivanti da una corretta applicazione delle tecniche statistiche.

Modalità di verifica dell'apprendimento

Prova orale con quesiti aperti e domande chiuse; lo studente deve dimostrare non solo di saper applicare le tecniche di analisi corrette, ma di saper interpretare i risultati ottenuti e comunicare in modo scientificamente corretto le evidenze riscontrate (problem solving).

Testi di riferimento

Ziegler A, König IR. A. Statistical Approach to Genetic Epidemiology; II edition, Wiley-Blackwell; 2010

P. Sham. Statistics in Human Genetics. Oxford University Press; 1997

Articoli Scientifici ad hoc forniti durante il corso

Lingua di insegnamento

Italiano

9.18 Popolazione, Disuguaglianze e Benessere

<i>Codice insegnamento:</i>	F8203B038
<i>Docente/i:</i>	Stefania Maria Lorenza Rimoldi
<i>Settore scientifico disciplinare:</i>	SECS-S/04
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, III Ciclo
<i>Tipo di attività:</i>	Opzionale
<i>Ore:</i>	42

Obiettivi formativi

L'obiettivo del corso è fornire il quadro di riferimento e gli strumenti necessari a comprendere i fenomeni della disuguaglianza e del benessere degli individui, nel campo della salute e degli stili di vita.

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- le disuguaglianze nella salute;
- la valutazione del benessere degli individui.

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- misurare i differenziali socio-demografici nella salute e nella mortalità;
- interpretare le relazioni tra fattori di contesto e benessere.

L'insegnamento consente allo studente di acquisire solide basi nell'applicazione della statistica al contesto lavorativo biostatistico/statistico/demografico.

Contenuti sintetici

- Le transizioni, sanitaria e demografica (prima e seconda)
- Invecchiamento delle popolazioni
- Disuguaglianze nella salute
- Misure della mortalità
- Benessere, qualità della vita.
- Speranza di vita felice.

Programma esteso

Le transizioni sanitaria e demografica:

- la transizione sanitaria;
- evoluzione delle cause di morte;
- modelli di diffusione di epidemie e pandemie.

L'invecchiamento delle popolazioni:

- invecchiamento per slittamento del limite e/o rettangolarizzazione;
- disuguaglianze nell'invecchiamento;

Disuguaglianze nella salute

- disuguaglianze socio-economiche;
- fragilità, disabilità, cronicità;

- mortalità e disabilità;
- misurare la buona salute.

Misure di mortalità:

- differenziali per sesso ed età (metodo di Arriaga);
- differenziali sociali: la povertà.

Benessere e qualità della vita:

- definizioni e misure;
- speranza di vita felice.

Prerequisiti

Nessuno

Metodi didattici

Durante le lezioni vengono illustrati gli argomenti teorici. Il corso si avvale di alcune lezioni svolte in laboratorio, in cui gli studenti dovranno applicare i metodi a set di dati loro assegnati e produrre un elaborato originale strutturato.

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 degli elaborati prodotti in autonomia dagli studenti durante il corso (prove intermedie). 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

- Eyal, N., Hurst, S. A., Norheim, O. F., & Wikler, D. (Eds.). 2013. Inequalities in health: concepts, measures, and ethics. Oxford University Press.
- Lutz, W. (2017). Studying the Demography of Sustainable Human Wellbeing: Empowered Life Years (ELY) as Sustainability Criterion.
- Micheli G.A. 2011. Demografie. McGraw-Hill (Cap. 1,3,4 e 8-Lez. 8.3)
- Pinnelli A., Racioppi F., Rettaroli R. 2003. Genere e demografia. Il Mulino (Parte Quinta: Genere e sopravvivenza)
- Siegel, J. S. 2011. The demography and epidemiology of human health and aging. Springer Science & Business Media (capitoli selezionati).

Ulteriore materiale didattico sarà fornito dal docente attraverso la piattaforma e-learning.

Lingua di insegnamento

Italiano

9.19 Statistical Models in Epidemiology

<i>Codice insegnamento:</i>	2021-2-F8203B008
<i>Docente/i:</i>	Rino Bellocco
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	12
<i>Periodo:</i>	II Semestre, III/IV Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	84

Obiettivi formativi

1. **Conoscenza e capacità di comprensione (knowledge and understanding):**

Il corso si propone di introdurre metodi e tecniche statistiche per lo studio (1) dei tempi di sopravvivenza: stima della distribuzione dei tempi di sopravvivenza, confronto tra curve di sopravvivenza, modelli semi-parametrici, modelli parametrici, rischi competitivi, modelli multivariati di sopravvivenza, regressione quantilica, disegno dello studio; (2) dei dati longitudinali di tipo continuo e discreto: modelli marginali, modelli a effetti misti, modelli GEE, dati mancanti.

2. **Capacità di applicare conoscenza e comprensione (applying knowledge and understanding):**

Sulla base delle conoscenze acquisite, lo studente sarà in grado di applicare le conoscenze acquisite per analizzare con strumenti appropriati i tempi di sopravvivenza e i dati longitudinali, utilizzando il software statistico Stata (StataCorp), di ampio uso in epidemiologia e biostatistica. Applicazioni in R and SAS saranno presentate brevemente in appositi seminari.

3. **Autonomia di giudizio (making judgements):**

Utilizzando le metodologie presentate a lezione, lo studente dovrà utilizzare le tecniche statistiche presentate per analizzare come variabili di esposizione, fattori di rischio, trattamenti, provenienti da studi epidemiologici o clinici possono avere un impatto sui tempi di sopravvivenza e sulle misure ripetute di dati longitudinali, e in modo critico valutare la performance del modello statistico utilizzato.

4. **Abilità comunicative (communication skills)**

Lo studente dovrà essere in grado di comprendere il materiale presentato in classe e di comunicare in modo autonomo le tecniche statistiche imparate in classe e esporre in modo chiaro come interpretare i risultati ottenuti.

5. **Capacità di apprendimento (learning skills):**

Lo studente acquisirà le capacità e conoscenze necessarie per analizzare i dati di sopravvivenza e longitudinali provenienti da uno studio epidemiologico o da un studio clinico, attraverso un percorso composto da lezioni teoriche e da applicazioni in Stata.

Contenuti sintetici

1. Analisi dei Dati di Sopravvivenza: Introduzione all'analisi della sopravvivenza. Censura e troncamento. Funzioni di sopravvivenza, densità e rischio. Procedura "st" in Stata. Stima e confronto delle funzioni di sopravvivenza. Il modello di regressione di Cox. Analisi dei rischi competitivi. Modelli di sopravvivenza parametrici. Regressione quantilica. Disegno dello studio.
2. Analisi dei dati longitudinali: Introduzione all'analisi dei dati longitudinali: i modelli marginali e i modelli a effetti casuali per risposte continue e discrete.

Programma esteso

1. Analisi dei Dati di Sopravvivenza: Introduzione all'analisi della sopravvivenza. Definizione ed esempi. Censura e troncamento. Funzioni di sopravvivenza, densità e rischio. Procedura “st” in Stata. Stima e confronto delle funzioni di sopravvivenza di Kaplan-Meier e Nelson-Aalen. Tavole di sopravvivenza. Log-rank test, Wilcoxon-Breslow-Gehan, Tarone-Ware test. Il modello di regressione di Cox. Formulazione e ipotesi. Verosimiglianza Parziale. Interpretazione dei parametri del modello. Diagnostica: Martingale e Deviance dei residuals, valutazione delle ipotesi del modello. Analisi dei rischi competitivi: definizione di net e crude survival, la funzione di incidenza cumulativa, hazard causa specifica e la funzione subhazard. Il modello di Fine e Gray. Modelli parametrici: esponenziale e weibull. La regressione quantilica. La sopravvivenza multivariata. Introduzione al disegno dello studio. Applicazioni in R e SAS.
2. Analisi dei dati longitudinali: Introduzione all'analisi dei dati longitudinali: esempi e motivazioni. Modelli di risposta per lo studio dei dati longitudinali: estensione dei modelli standard. Struttura della matrici di varianza e covarianza. Analisi dei profili di risposta. I modelli a effetti casuali per risposte continue e discrete. Modelli a effetti fissi. Introduzione ai valori mancanti: definizione, trattamento e possibili metodi di stima. Introduzione ai dati mancanti.

Prerequisiti

Le lezioni vengono svolte in lingua inglese, anche il materiale distribuito e' in lingua inglese: dunque la conoscenza della lingua inglese e' considerata un requisito importante per la partecipazione attiva in classe, per la comprensione delle slides utilizzate, e per la lettura di tutto il materiale del corso.

Metodi didattici

Blended/Elearning: Lezioni frontali e online , group work, seminars, utilizzo di Stata. Le lezioni online saranno svolte con un duplice scopo: il primo e' quello di consentire a studenti non frequentanti di seguire alcune lezioni, che verranno preparate e registrate e dunque rese disponibili nella piattaforma elearning. Il secondo e' quello di discutere e rivedere i compiti che gli studenti hanno preparato, evidenziando le problematiche presenti, proponendo un modello di soluzione al compito.

Modalità di verifica dell'apprendimento

1. Lavoro di gruppo (30% votazione finale) (max 5 - 6 studenti) che si basa sulla consegna di 9 compiti, uno a settimana, conta per il 30% della votazione finale.
2. Esame Finale (70% votazione finale).
3. Prova orale-facoltativa : La discussione orale dell'esame scritto serve per valutare la conoscenza dello studente laddove le risposte all'esame non siano state chiare e poste in modo impreciso.

Sia i compiti che si basano sul lavoro di gruppo, sia l'esame finale, hanno come fine quello di valutare il grado di apprendimento dello studente e se gli obiettivi del corso sono stati raggiunti in termine di formulazione delle ipotesi dello studio analizzato utilizzando Stata, analisi descrittive, inferenza statistica, interpretazione dei parametri, valutazione delle ipotesi dei modelli fittati.

L'orale può comportare sia l'aumento che la diminuzione della valutazione della prova scritta.

Per gli studenti che non possono partecipare al lavoro di gruppo, la valutazione si baserà sulla prova scritta e orale soltanto, con un peso del 100%.

L'esame scritto finale e' lo stesso per studenti frequentanti e studenti non frequentanti

Testi di riferimento

1. Kleinbaum, D.G. and Klein, M. Survival Analysis, A self-learning text. (2013). Springer.
2. Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2013). Applied Longitudinal Analysis, Chapman & Hall CRC.
3. Jewell, N.P. Statistics for Epidemiology. (2004). Chapman & Hall CRC.

Materiale integrativo sara' fornito durante le lezioni e messo a disposizione tramite la pagina web del corso.

Lingua di insegnamento

Inglese

9.20 Valutazione Statistica dei Sistemi Sanitari

<i>Codice insegnamento:</i>	F8203B020
<i>Docente/i:</i>	Marco Albini
<i>Settore scientifico disciplinare:</i>	SECS-S/01
<i>CFU:</i>	6
<i>Periodo:</i>	II Semestre, IV ciclo
<i>Tipo di attività:</i>	Opzionale
<i>Ore:</i>	42

Obiettivi formativi

L'obiettivo del corso riguarda la descrizione dei differenti modelli di Sistema Sanitario e i conseguenti differenti metodi di valutazione delle performance. In particolare si approfondiranno le metodologie statistiche che all'interno di un Sistema Sanitario sono utilizzate per la valutazione del sistema stesso, ma anche le metodologie statistiche che ciascun attore del sistema (ospedali o altri provider) utilizza per il miglioramento della propria performance. Si studierà in particolare la capacità dei modelli di regressione semplice, multilevel o di altre metodologie, di rispondere meglio all'obiettivo di analisi di sistema o di miglioramento della performance. A completamento della analisi della metodologia statistica verrà affrontato il tema della rappresentazione grafica degli esiti delle valutazioni.

Contenuti sintetici

I differenti modelli dei Sistemi Sanitari e le esperienze di valutazione

La valutazione dei sistemi sanitari rispetto all'efficacia, l'efficienza e la customer satisfaction.

Modelli statistici utilizzati per l'analisi di Sistema

Modelli di analisi delle performance per singolo Provider

Lettura ragionata di esempi reali ed esercitazioni pratiche

Programma esteso

L'obiettivo del corso consiste nel descrivere i principali ambiti di valutazione dei sistemi sanitari, con riferimento alle principali esperienze in essere in ambito nazionale e internazionale. Il corso si propone di illustrare le metodologie statistiche adatte a ciascun obiettivo di valutazione, ad esempio modelli di regressione semplice e multilevel, nonché altre metodologie di analisi. Durante il corso verrà inoltre affrontato il tema della rappresentazione grafica degli esiti delle valutazioni. Lo studente dovrà acquisire i concetti riguardanti i sistemi sanitari e la loro valutazione, le strategie modellistiche adeguate per la valutazione e la rappresentazione degli esiti.

Prerequisiti

Nessuno

Metodi didattici

Lezioni frontali teoriche dove vengono presentati gli argomenti del corso con la descrizione degli elementi di politica ed economia sanitaria e le metodologie statistiche per la valutazione. Durante la lezione verranno presentati e discussi casi concreti di valutazione.

Modalità di verifica dell'apprendimento

Preparazione di un elaborato in cui lo studente dovrà impostare tutti gli elementi utili alla valutazione del sistema sanitario o del miglioramento di una struttura sanitaria, utilizzando i contenuti appresi durante le lezioni. L'elaborato sarà oggetto di discussione durante l'esame orale.

Testi di riferimento

Slides presentate durante le lezioni

G. Vittadini (2012), Manuale del Sistema di valutazione della performance degli ospedali lombardi, Aracne editrice, Roma

Snijders, Tom A.B., and Bosker, Roel J. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, second edition. London etc.: Sage Publishers, 2012.

Eugene C. Nelson, Paul B. Batalden, Joel S. Lazar Practice-based Learning and Improvement: A Clinical Improvement Action Guide, second edition, Joint Commission Resources, Incorporated, 2007

Lingua di insegnamento

Italiano

9.21 Programmazione SAS per la Biostatistica

<i>Codice insegnamento:</i>	F8203B035
<i>Docente/i:</i>	Vincenzo Bagnardi
<i>Settore scientifico disciplinare:</i>	MED/01
<i>CFU:</i>	6
<i>Periodo:</i>	I Semestre, IV Ciclo
<i>Tipo di attività:</i>	Obbligatorio
<i>Ore:</i>	44

Obiettivi formativi

Obiettivo dell'insegnamento è approfondire gli strumenti del linguaggio di programmazione SAS utili alla gestione dei dati, all'analisi statistica e alla reportistica di studi clinici e osservazionali..

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- Fondamenti del linguaggio di programmazione SAS;
- Principali procedure per la gestione e l'analisi statistica di dati clinici

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- Gestire database con il linguaggio di programmazione SAS;
- Implementare le principali procedure di analisi biostatistica con il linguaggio di programmazione SAS;
- Produrre la reportistica di studi clinici e osservazionali.

L'insegnamento consente allo studente di acquisire solide basi nella programmazione SAS nel contesto lavorativo delle scienze biomediche.

Contenuti sintetici

- introduzione al sistema SAS
- tecniche per la gestione dei dati longitudinali in SAS
- le principali procedure per l'analisi dei dati di uno studio clinico
- utilizzo del linguaggio SAS macro per l'automatizzazione dei processi
- implementazione di algoritmi statistici in SAS/IML (interactive matrix language)

Programma esteso

1. Introduzione al corso

Importanza del sistema SAS nell'ambito della ricerca clinica e ruolo del programmatore SAS nell'industria farmaceutica.

2. Introduzione al sistema SAS

L'ambiente di lavoro SAS

Data step e proc step
Importazione dei dati
Esempi di procedure
Gestione e manipolazione dei dati
Funzioni SAS
Tabelle a doppia entrata
Date e Formati
Unione di dataset
PROC SQL
Gestione degli output (ODS)

3. Tecniche per la gestione e l'analisi dei dati longitudinali in SAS

Selezione della prima e dell'ultima osservazione per un paziente in uno studio longitudinale
Formato wide e formato long
Procedure per l'analisi di misure ripetute

4. Le principali procedure per l'analisi dei dati di uno studio clinico

Procedure per l'analisi di risposte continue
Procedure per l'analisi di risposte binarie
Procedure per l'analisi dei tempi di sopravvivenza
Procedure SAS per la creazione di tabelle (PROC TABULATE) e grafici (PROC SGPLOT e PROC SGPANEL)

5. Utilizzo del linguaggio SAS macro per l'automatizzazione dei processi

Le macro variabili
I macro programmi

Implementazione di algoritmi statistici in SAS/IML (interactive matrix language)

Operazioni di base in SAS IML
Applicazioni statistiche in SAS IML
SAS/IML e R

Prerequisiti

Nessuno

Metodi didattici

Lezioni e esercitazioni al computer con SAS

Modalità di verifica dell'apprendimento

Prova scritta

La prova di esame consiste in una serie di brevi esercizi e problemi pratici da risolvere con gli strumenti SAS illustrati durante le lezioni.

La prova di esame si tiene in laboratorio.

Durante la prova d'esame è consentita la consultazione del materiale fornito a lezione.

Non sono previste prove di esame in itinere.

La prova scritta consente di verificare la capacità dello studente di impostare in modo indipendente, utilizzando il linguaggio di programmazione SAS, la risoluzione di esercizi e problemi pratici comuni nel contesto lavorativo delle scienze biomediche.

Prova orale facoltativa (su richiesta dello studente).

Le modalità di esame sono identiche per studenti frequentanti e non frequentanti.

Testi di riferimento

Tutto il materiale didattico (estratti di libri, articoli, codice SAS, dataset) distribuito agli studenti durante il corso

Lingua di insegnamento

Italiano