

SYLLABUS DEL CORSO

Probabilità e Statistica per l'Informatica (blended)

1920-2-E3101Q127

Obiettivi

Fornire abilità di base di; **statistica descrittiva**, **statistica inferenziale** e **calcolo delle probabilità**.
Lo studente al termine del corso sarà in grado di

- analizzare e sintetizzare dati quantitativi sia univariati che multivariati
- impostare e condurre un ragionamento in condizioni incerte caratterizzate da eventi multipli
- utilizzare le distribuzioni notevoli per modellare fenomeni e sistemi reali
- effettuare stime puntuali ed intervallari per parametri che caratterizzano fenomeni incerti
- impostare e condurre verifiche di ipotesi su affermazioni quantitative e qualitative
- studiare il legame esistente tra un insieme di variabili di input ed una variabile di output
- affrontare tutti i punti precedenti tramite il liguaggio di programmazione open source R

Contenuti sintetici

Il corso presenta elementi di **statistica descrittiva**, univariata, bivariata e multivariata. Verranno presentati concetti di **calcolo delle probabilità** come variabili aleatorie, continue e discrete, unidimensionali, bidimensionali e multidimensionali. Verranno poi presentate le principali **distribuzioni di probabilità**, continue e discrete, i **teoremi fondamentali di convergenza** come il teorema limite centrale e la legge dei grandi numeri. Verranno forniti elementi di **stima puntuale ed intervallare**. Il corso presenterà rudimenti sulla **verifica di ipotesi parametrica e non parametrica**. Infine, verranno presentate l'**analisi di regressione lineare** unidimensionale e

multidimensionale. Il corso porrà particolare attenzione al **linguaggio di programmazione R** in modo da completare le competenze metodologiche acquisite tramite lezioni teoriche ed esercitazioni numeriche con competenze operative. In questo modo si fornirà allo studente una maggiore autonomia progettuale ed implementativa.

Programma esteso

?1. Statistica Descrittiva

- 1.1. Rappresentazioni numeriche e grafiche di dati statistici
- 1.2. Indici di tendenza centrale e di variabilità
- 1.3. Rappresentazione per caratteri bidimensionali??

??2. Calcolo delle Probabilità

- 2.1. Definizioni di spazio campione, probabilità e funzione di ripartizione
- 2.2. Probabilità condizionata ed indipendenza stocastica
- 2.3. Formula del prodotto, delle probabilità totali e teorema di Bayes
- 2.4. Variabili aleatorie continue e discrete
- 2.5. Variabili aleatorie unidimensionali e multidimensionali
- 2.6. Indici di tendenza centrale e variabilità

3. Distribuzioni Notevoli

- 3.1. Distribuzioni discrete: Bernoulli, Binomiale, Poisson, Geometrica, ...
- 3.2. Distribuzioni continue: normale, beta, esponenziale, t di Student, F, Chi-quadro, ...

4. Teoremi di Convergenza

- 4.1. Convergenza in distribuzione
- 4.2. Legge dei grandi numeri
- 4.3. Teorema limite centrale

5. Stima di Parametri

- 5.1. Campionamento e campioni
- 5.2. Principali distribuzioni campionarie
- 5.3. Stimatori e stime puntuali
- 5.4. Stime intervallari: intervalli di confidenza per la media e la varianza

6. Verifica di Ipotesi: test parametrici

- 6.1. Introduzione alla verifica di ipotesi
- 6.2. Errori del I e del II tipo
- 6.3. Test sulla media e sulla varianza di una popolazione
- 6.4. Test sulla differenza delle medie e sulla differenza della varianza di due popolazioni

7. Verifica di Ipotesi: test non parametrici

- 7.1. Test per la bontà dell'adattamento: Kolmogorov-Smirnov e test Chi-quadro
- 7.2. Test per il confronto delle distribuzioni di popolazioni, test dei segni e test dei ranghi
- 7.3. Test di indipendenza

8. Regressione Lineare

- 8.1. Introduzione alla regressione e alla regressione lineare
- 8.2. Stima delle costanti del modello, intervalli di confidenza per i valori dei singoli individui
- 8.3. Attendibilità di un modello lineare, Analisi dei residui
- 8.4. Regressione lineare multipla: stima parametri, attendibilità, importanza variabili

Prerequisiti

Analisi Matematica, capacità di progettare e sviluppare programmi secondo il paradigma procedurale.

Modalità didattica

Il corso viene erogato in modalità *blended-learning* ed *in lingua Italiana*. Ogni capitolo è composto da un insieme di moduli formativi.

Un *modulo formativo* prevede le seguenti sessioni:

- presentazione e discussione di una o più componenti metodologiche e/o teoriche
- esercitazione che mostra istanze concrete delle componenti metodologiche e/o teoriche

- laboratorio di progettazione e realizzazione di codice in linguaggio R delle componenti metodologiche e/o teoriche
- autovalutazione, al termine della quale viene suggerito allo studente quali argomenti rivedere e/o approfondire

Il corso stimola e favorisce l'*interazione* tra *docente* e *studente*, tra studente e *tutor online* e tra studente e studente tramite la predisposizione di *forum tematici* per ogni capitolo presentato e forum dedicati alla programmazione in *linguaggio R*.

Il corso rende disponibile immediatamente tutto il materiale e stimola lo studente ad essere presente in aula dopo aver preso visione del materiale della lezione che verrà svolta in modo tale da *umentare interazione*, *discussione* e pertanto *apprendimento*.

Importante segnalare che circa il 70% degli argomenti del corso sono disponibili sotto forma di **video lezioni progettate e registrate dal docente del corso**. Pertanto, lo studente può liberamente decidere se assistere alle lezioni in aula o meno. Anche nel caso in cui decidesse di assistere alla lezioni in aula, le video lezioni offrono un valido strumento di ripasso.

Materiale didattico

Il libro di testo è il seguente; **Franco Pellerey (2007)**. *Elementi di statistica per le applicazioni. con esercizi*, CELID. Inoltre, viene reso disponibile il seguente materiale didattico;

- Supporti digitali, audiovideo, realizzati dal docente.
- Slide realizzate dal docente.
- Esercizi numerici con svolgimento a cura del docente, tramite slide o supporto digitale.
- Forum tematici per ogni capitolo del corso e per il linguaggio R.
- Quiz di autovalutazione e strumenti di simulazione per la preparazione alla prova d'esame.

Periodo di erogazione dell'insegnamento

Secondo semestre

Modalità di verifica del profitto e valutazione

- **Itinere**; lo studente è stimolato a seguire il corso durante il periodo di erogazione, a studiare settimanalmente sfruttando i moduli formativi per giungere alle due prove in itinere preparato. La prima prova interessa i capitoli 1, 2, 3 e 4, mentre la seconda i capitoli 5, 6, 7 ed 8. Ogni prova verifica la preparazione dello studente tramite **10 quiz a risposta chiusa** (ogni quiz riguarda le nozioni presentate nel corso ed assegna 1 punto), **2 esercizi numerici** (ogni esercizio numerico prevede lo sviluppo ed il calcolo di una soluzione ad un problema assegnato ed assegna un massimo di 9 punti), **1 domanda a risposta aperta** (domanda relativa a nozioni presentate nel corso che assegna un massimo di 5 punti) ed eventualmente, a richiesta dello studente, **1 domanda sul linguaggio R** (prevede di commentare un listato in linguaggio R ed assegna un massimo di 3 punti) ed una **prova orale** (prevede domande di ragionamento e deduzione ed assegna un massimo di 2 punti). Il voto verbalizzato è ottenuto come arrotondamento per eccesso della media aritmetica dei voti delle due prove in itinere..
- **Ordinaria**; verifica la preparazione dello studente su tutti i capitoli del corso tramite **10 quiz a risposta chiusa** (ogni quiz riguarda le nozioni presentate nel corso ed assegna 1 punto), **3 esercizi numerici** (ogni esercizio numerico prevede lo sviluppo ed il calcolo di una soluzione ad un problema assegnato ed assegna un massimo di 6 punti), **2 domande a risposta aperta** (ogni domanda è relativa a nozioni presentate nel corso ed assegna un massimo di 2 punti) ed eventualmente, a richiesta dello studente, **1 domanda sul linguaggio R** (prevede di commentare un listato in linguaggio R ed assegna un massimo di 1 punto) ed una **prova orale** (prevede domande di ragionamento e deduzione ed assegna un massimo di 2 punti).

Orario di ricevimento

Su appuntamento
