POLITECNICO DI MILANO SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE CORSO DI STUDI IN INGEGNERIA FISICA

Insegnamento di STATISTICA – Anno Accademico 2020/2021

Crediti

5 CFU

Docente

• Alessandro Toigo

Studio: Dipartimento di Matematica, via Bonardi, 9, La Nave, II piano

Tel: 02 2399 4650

 $e ext{-}mail$: alessandro.toigo@polimi.it

Aula Webex: https://politecnicomilano.webex.com/meet/alessandro.toigo

Orario di ricevimento: su appuntamento

Esercitazioni e laboratorio

• Tommaso Andreoli (esercitazioni)

 $e ext{-}mail$: tommaso.andreoli@polimi.it

Aula Webex: https://politecnicomilano.webex.com/meet/tommaso.andreoli

Orario di ricevimento: su appuntamento

• Francescopaolo Manicone (laboratorio)

 $e ext{-}mail$: francescopaolo.manicone@polimi.it

 $Aula\ Webex: \ https://politecnicomilano.webex.com/meet/francescopaolo.manicone$

Orario di ricevimento: su appuntamento

Per tutto il semestre il ricevimento verrà svolto via Cisco Webex.

Orari e Aule

martedì	8:15 - 11:15	online	lezione / laboratorio
martedì	16:15 - 18:15	aula 5.1.1	esercitazione (squadra 1 – cod. persona dispari)
giovedì	10:15 - 12:15	aula OSVALDO DE DONATO	esercitazione (squadra 2 – cod. persona pari)

Pagina web

Su BeeP, il portale di supporto alla didattica per l'intero Ateneo, compariranno: avvisi agli studenti, testi e soluzioni delle prove d'esame, valutazioni, materiale didattico.

Obiettivi

Obiettivi dell'insegnamento sono:

- 1. fornire tecniche e procedure per la descrizione sintetica e grafica delle informazioni fornite da insiemi di dati;
- 2. introdurre al linguaggio e ai modelli per la rappresentazione e l'analisi di fenomeni aleatori;
- 3. introdurre ai metodi e agli strumenti dell'inferenza statistica;
- 4. applicare i metodi e le tecniche dell'analisi statistica a insiemi di dati reali per mezzo di un opportuno software statistico.

Programma

- 1. Statistica descrittiva. Tipi di dati. Tabelle di distribuzione di frequenza. Indici di posizione: media, moda, mediana. Quantili. Indici di dispersione: varianza e scarto interquartile. Rappresentazioni grafiche di distribuzioni: istogrammi e box-plot.
- 2. Elementi di calcolo delle probabilità. Esperimenti aleatori, variabili aleatorie e probabilità. Variabili aleatorie discrete e assolutamente continue: densità, funzione di ripartizione, media e varianza. Distribuzioni di Bernoulli, Binomiale, Poissoniana, Uniforme discreta e continua, Esponenziale e Normale. Variabili aleatorie indipendenti. Il Teorema Centrale del Limite. Approssimazione Normale per le distribuzioni Binomiale e Poissoniana.
- 3. Statistica inferenziale. Campioni casuali e statistiche. Stimatore e stima di un parametro: distorsione, errore quadratico medio, formula di propagazione degli errori. Intervalli di confidenza (IC) per un parametro. La verifica delle ipotesi. Criteri di decisione per la scelta tra due ipotesi: errore di primo tipo ed errore di secondo tipo. Test: regione di rifiuto, livello di significatività e funzione potenza. Potenza e dimensione campionaria. p-value di un test. Test e IC per la media di una popolazione a varianza nota o incognita. Test e IC per la varianza di una popolazione normale. Test e IC per una proporzione. Test di Shapiro-Wilks. Normal probability plot. Test per il confronto tra due medie. Test per il confronto tra due proporzioni. Test per il confronto tra due varianze.
- 5. Regressione lineare. Modelli empririci gaussiani di regressione lineare (semplice e multipla). Stima ai minimi quadrati dei parametri del modello. Diagnostica: analisi dei residui, R-quadro. Metodi per la selezione delle variabili. Test e IC per i coefficienti di una regressione. Previsione di una nuova osservazione. Modelli con variabili dummy.

Sono previste esercitazioni di laboratorio informatico su piattaforma R sugli argomenti trattati a lezione ed esercitazione. L'attività di laboratorio è fondamentale per l'apprendimento.

Testi consigliati

- 1. Montgomery D.C., Runger G.C., Hubele N.F., Statistica per ingegneria II edizione italiana, Egea, 2012
- 2. Ross S.M., Probabilità e statistica per l'ingegneria e le scienze, Apogeo, 2008
- 3. Dalgaard P., Introductory Statistics with R, Springer, 2008
- 4. IEVA F., MASCI C., PAGANONI A.M., Laboratorio di Statistica con R, Pearson, 2016

Laboratori informatici

Il laboratorio informatico si svolgerà esclusivamente in modalità online. I dati e gli script da utilizzare saranno previamente caricati su BeeP oppure resi disponibili dal docente durante le lezioni. Per partecipare ai laboratori informatici è necessario aver scaricato e installato il software R sul proprio computer. Tale software è reperibile gratuitamente ai link seguenti:

• Windows: https://cran.r-project.org/bin/windows/base/

• MacOS: https://cran.r-project.org/bin/macosx/

• Linux: https://cran.r-project.org/bin/linux/

Modalità di verifica

La verifica dell'apprendimento viene effettuata per mezzo di:

- una prova d'esame scritta obbligatoria, nell'ambito degli appelli d'esame previsti;
- un eventuale esame orale integrativo, a esclusiva discrezione del docente;
- un progetto di analisi dati facoltativo, da presentare e discutere durante apposite sessioni concordate col docente.

Prova d'esame scritta (obbligatoria). Ogni prova scritta consta di più problemi da risolvere individualmente e autonomamente. Durante l'esame lo studente potrà utilizzare il formulario preparato dal docente e le tavole delle distribuzioni notevoli – reperibili entrambi sulla pagina Beep del corso – e la calcolatrice. È vietato l'uso di altro materiale. Non è nemmeno consentito utilizzare software statistici.

Il voto massimo conseguibile nella prova scritta è pari a 29. La prova si intende superata se il suo voto, integrato dall'eventuale prova orale, è almeno 18. Non è possibile ottenere né 30 né la lode senza aver presentato il progetto di analisi dati facoltativo.

Esame orale integrativo (a discrezione del docente). L'orale può vertere su qualunque argomento del corso (teoria, esercizi o laboratorio). Non sono richieste le dimostrazioni dei teoremi.

L'eventuale convocazione all'esame orale avviene contestualmente alla pubblicazione dei risultati dello scritto.

Il voto ottenuto allo scritto non è in alcun modo vincolante per il voto finale di scritto + orale.

Progetto di analisi dati (facoltativo). Il progetto di analisi dati è un lavoro di gruppo, da svolgere obbligatoriamente in gruppi di 3 studenti. Si deve configurare come una consulenza di tipo statistico, e deve contenere:

- un problema concreto da affrontare;
- un insieme di dati inerenti al problema;
- un'analisi dati eseguita formalizzando adeguatamente il problema e utilizzando, fra gli strumenti statistici introdotti a lezione, quelli più adatti considerata la natura del problema e dei dati;
- la vostra conclusione sul problema.

Gli studenti sono invitati a scegliere l'argomento della tesina fra quelli di proprio interesse, universitario o extra universitario, e a reperire quindi i dati utili per l'analisi.

Alcune linee guida più dettagliate saranno illustrate a laboratorio e messe a disposizione su BeeP al termine del corso.

Ogni gruppo deve preparare:

- una *relazione scritta*, organizzata come una breve sintesi del lavoro, corredata da tutti gli allegati del caso, da inviare via mail al docente due giorni prima della presentazione;
- una presentazione a diapositive (o simili) da proiettare tramite PC, organizzata considerando di avere a disposizione 10 minuti massimi a testa per la presentazione (30 minuti massimi in totale).

Ogni membro del gruppo riceve una valutazione individuale, che potrà andare da un minimo di -1 punto a un massimo di +5 punti.

La presentazione del progetto avviene in prossimità di uno dei tre appelli di giugno, luglio e settembre, scegliendo una data fra quelle proposte nella form d'iscrizione dei gruppi. Chi vuole presentare il progetto deve iscriversi alla form entro e non oltre il 20 giugno 2021.

Il voto ottenuto col progetto è valido per almeno un anno, e viene sommato al voto del primo scritto (o scritto + orale) sufficiente. Se il voto risultante supera 30, si ottiene la lode.