

SCHEDA INSEGNAMENTO: Statistica e Data Science

DOCENTE: Marcello De Giosa

A.A. 2020-2021

Insegnamento	Statistica e Data Science		
SSD	Mat/06		
Anno di Corso	2020-2021		
Codice Insegnamento			
Semestre	II		
Docente	Marcello De Giosa		
Crediti	8 (4 lezioni frontali + 2 laboratorio + 2 esercitazioni)		
Semestre	Dal 1 marzo al 15 giugno		
Propedeuticità	nessuna		
Prerequisiti	Per il raggiungimento degli obiettivi formativi è utile che lo studente posseda, seppur non approfonditamente, i concetti di base contenuti nel corso di Matematica e competenze generiche nelle materie scientifiche. Il possesso di tali prerequisiti sarebbe utile anche agli studenti lavoratori e non frequentanti.		
Obiettivi formativi	<p><i>Conoscenza e capacità di comprensione</i> Acquisizione delle metodologie di base della statistica e del data science più comuni in ambito ambientale. Vengono forniti gli strumenti propri del metodo scientifico applicato alla comprensione delle suddette metodologie. Il corso si articola in lezioni teoriche, esercitazioni e laboratorio con l'uso del software open source R, allo scopo di accrescere la capacità dello studente di individuare l'approccio più idoneo per la risoluzione di problematiche ambientali.</p> <p><i>Capacità di applicare conoscenza e comprensione</i> Acquisizione delle abilità connesse all'applicazione dei concetti teorici e pratici per la soluzione di problematiche ambientali concrete. L'acquisizione di tali abilità deve essere il risultato di esperienze pratiche ed esercitazioni di laboratorio.</p> <p><i>Autonomia di giudizio</i> Acquisizione della capacità di dedurre dai risultati ottenuti conclusioni utili alla soluzione di problematiche ambientali.</p> <p><i>Abilità comunicative</i> Acquisizione della capacità di discutere i concetti fondamentali delle tematiche di studio in modo chiaro ed esauriente, utilizzando un linguaggio scientifico adeguato. Al raggiungimento di questo obiettivo concorrono le discussioni durante le lezioni teoriche, le esercitazioni ed il laboratorio.</p> <p><i>Capacità di apprendimento</i> Acquisizione di capacità di integrare le conoscenze di base attraverso percorsi personali di approfondimento con materiale scientifico rigoroso.</p>		
Metodi didattici	Lezioni frontali	Laboratorio + Esercitazioni	Totale
<i>Ore didattica assistita</i>	36	30+30	96
<i>Ore studio individuale</i>	76	40	116
<i>Crediti</i>	4	2+2	8

Metodi di valutazione	<p>La valutazione dello studente prevede:</p> <ul style="list-style-type: none"> – una prova di laboratorio relativa all’uso del software R per rispondere, con l’uso di metodologie statistiche, a tipiche questioni in ambito ambientale. La prova dura 2 ore e consiste nella redazione di una breve relazione che descriva e commenti i codici utilizzati e gli output ottenuti. – una prova orale che generalmente consiste in tre domande relative a differenti argomenti del corso e/o chiarimenti circa la prova di laboratorio. <p>Il punteggio della prova d’esame è attribuito mediante un voto espresso in trentesimi. Una votazione eccellente è il risultato del soddisfacimento di gran parte dei seguenti criteri di valutazione.</p> <ul style="list-style-type: none"> • <i>Conoscenza e capacità di comprensione</i> Lo studente deve dimostrare di avere padronanza dei concetti di statistica trattati durante il corso. • <i>Capacità di applicare conoscenza e comprensione</i> Lo studente deve dimostrare di saper applicare i concetti statistici teorici acquisiti nel corso a problematiche concrete di interesse ed origine ambientale.. • <i>Autonomia di giudizio</i> Lo studente deve essere in grado di individuare autonomamente le metodologie statistiche idonee a rispondere ai quesiti posti. • <i>Abilità comunicative</i> Lo studente deve aver acquisito la capacità di comunicare compiutamente i concetti appresi e i risultati delle analisi statistiche svolte con l’uso di un linguaggio scientifico corretto. • <i>Capacità di apprendimento</i> Lo studente deve dimostrare di aver acquisito gli strumenti per arricchire le sue conoscenze anche attraverso percorsi di approfondimento individuali o di gruppo.
Programma	<p>Una breve introduzione ad R. Installazione. La console. Classi di oggetti. Vector, data.frame e factor. Function.</p> <p>Metodi esplorativi. Rango. Media. Mediana. Varianza e Deviazione Standard. Dotchart. Boxplot. Frequenza e densità. Probabilità. Stima di una Probabilità. Istogrammi.</p> <p>Densità di probabilità e stima.</p> <p>Distribuzioni: Normale, lognormale, t di student. Verificare la distribuzione di provenienza di un campione.</p> <p>Stima di una media: puntuale, intervalli di fiducia, test di ipotesi.</p> <p>Stima della differenza di due medie: puntuale, intervalli di fiducia, test di ipotesi.</p> <p>Il modello lineare. Definizione, stima dei parametri, partizione della dispersione, stima di una nuova osservazione.</p>
Testi di Riferimento	<p>Appunti dalle lezioni e materiale fornito dal docente.</p> <p>S.M. Ross (2014). Probabilità e statistica per l’ingegneria e le scienze. Apogeo.</p> <p>Manuals di R nel sito cran.r-project.org</p>
Testi di Approfondimento e strumenti a supporto della didattica	<p>Appunti distribuiti durante il corso.</p> <p>Daniel, W.W.; Cross, C.L. (2019). Biostatistica. Edises.</p>