

## COURSE SYLLABUS

### Fundamentals of Bioinformatics and Biostatistics

2425-1-E0201Q081

---

#### Obiettivi

L'insegnamento si propone di spiegare i fondamenti della bioinformatica e della biostatistica, con una particolare attenzione all'applicazione di queste discipline nel contesto delle biotecnologie.

Nell'ambito della bioinformatica, saranno affrontati i principali problemi computazionali legati all'analisi di sequenze biologiche (anche in riferimento ai limiti e ai vantaggi delle tecniche di sequenziamento di vecchia e nuova generazione), alla ricostruzione di genomi e alla predizione di strutture proteiche. Verranno inoltre introdotti i concetti base sulle tecniche di apprendimento automatico e intelligenza artificiale per applicazioni in contesti biologici, biotecnologici e biomedici.

Nell'ambito della biostatistica, saranno trattati i principali metodi di statistica descrittiva e statistica inferenziale al fine di guidare lo studente nel processo che parte dal disegno di uno studio sperimentale, procede con l'analisi dei dati raccolti e termina nell'interpretazione statistica e nella comprensione della rilevanza biologica dei risultati ottenuti.

#### *Conoscenza e capacità di comprensione.*

Al termine dell'insegnamento lo studente dovrà sviluppare la capacità di:

- "computational thinking", al fine di utilizzare in modo appropriato gli strumenti computazionali per la soluzione di un dato problema;
- analisi critica, al fine di scegliere i metodi statistici più adeguati per l'analisi di dati e per l'interpretazione dei rispettivi risultati.

#### *Capacità di applicare conoscenza e comprensione.*

Al termine dell'insegnamento lo studente dovrà essere in grado di applicare le conoscenze acquisite per la soluzione di problemi in ambito computazionale e statistico, oltre a:

- sviluppare semplici algoritmi;
- organizzare e trattare automaticamente i dati biologici o risolvere semplici problemi tramite l'utilizzo di fogli di calcolo.

#### *Autonomia di giudizio.*

Al termine dell'insegnamento lo studente dovrà essere in grado di elaborare quanto appreso, e saper riconoscere

le situazioni e i problemi in cui le metodologie bioinformatiche e biostatistiche apprese possano essere utilizzate.

*Abilità comunicative.*

Al termine dell'insegnamento lo studente dovrà essere in grado di esprimersi in modo appropriato nella descrizione delle tematiche affrontate, con proprietà di linguaggio e sicurezza di esposizione.

*Capacità di apprendimento.*

Al termine dell'insegnamento lo studente dovrà essere in grado di consultare la letteratura sugli argomenti trattati, nonché analizzare, applicare, integrare e collegare le conoscenze acquisite con quanto verrà appreso in insegnamenti correlati all'applicazione delle scienze computazionali e statistiche per l'analisi dei dati.

## **Contenuti sintetici**

### **Bioinformatica**

Algoritmi, computational thinking, elementi di programmazione strutturata, complessità computazionale.

Nozioni di metodi euristici di ispirazione biologica, apprendimento automatico e intelligenza artificiale.

Allineamenti di sequenze e ricostruzione di genomi (con cenni sulle piattaforme di sequenziamento).

Metodi computazionali per la previsione di strutture proteiche.

### **Biostatistica**

Statistica descrittiva.

Statistica inferenziale.

### **Esercitazioni in laboratorio**

Fogli di calcolo.

Ricerche in banche dati biologiche.

Applicazione di metodi di biostatistica.

## **Programma esteso**

### **Bioinformatica**

- Pensare in modo algoritmico. Definizione di algoritmo. Programmazione strutturata e pseudo-codice. Complessità computazionale.
- Metodi computazionali di ispirazione biologica, nozioni di base di metodi di apprendimento automatico supervisionato e non supervisionato, intelligenza artificiale.
- Cenni sulle piattaforme di sequenziamento di acidi nucleici. Algoritmi di allineamento fra sequenze, tecniche euristiche per ricerche in banche dati. Metodi di ricostruzione e annotazione di genomi. Banche dati biologiche.
- Analisi e previsione di strutture e interazioni proteiche (protein folding, molecular docking).

### **Biostatistica**

- Statistica descrittiva. Campioni e popolazioni, tipologie di dati e variabili. Disegno di esperimenti (cieco, strategie di campionamento). Il concetto di frequenza (assoluta, relativa, cumulativa). Rappresentazione grafica dei dati (istogrammi, diagrammi a torta, diagrammi a dispersione). Misure di centralità (media, mediana, moda). Misure di dispersione (range, deviazione standard, varianza). Misure di posizione (quartili, percentili). Analisi esplorativa dei dati (outlier, boxplot).
- Nozioni di teoria della probabilità. Distribuzioni di probabilità (uniforme, binomiale, normale, distribuzione t di

- Student). Distribuzioni campionarie.
- Statistica inferenziale. Metodi di stima con un campione (intervalli di confidenza, stima con un campione). Verifica di ipotesi con un campione. Correlazione e regressione.

### **Esercitazioni in laboratorio**

- I fogli di calcolo per l'elaborazione automatica dei dati.
- Ricerca in banche dati biologiche (geni, genomi, proteine).
- Applicazione di metodi di statistica descrittiva e inferenziale per l'analisi di dati biologici.

### **Prerequisiti**

Prerequisiti: nessuno.

Propedeuticità: nessuna.

### **Modalità didattica**

*Lezioni frontali.*

20 lezioni da 2 ore costituite da:

- una parte in modalità erogativa (didattica erogativa, DE) focalizzata sulla presentazione-illustrazione di contenuti, concetti, principi scientifici;
- una parte in modalità interattiva (didattica interattiva, DI), che prevede lavoro di gruppo in aula e dimostrazioni aggiuntive di applicazioni pratiche dei contenuti della parte erogativa (casi di studio, esercizi applicativi della teoria). Tutte le attività sono svolte in presenza.

*Esercitazioni in aula di calcolo.*

10 attività di esercitazione da 3 ore svolte in modalità interattiva in presenza.

### **Materiale didattico**

Slide e videoregistrazioni delle lezioni e delle esercitazioni reperibili sulla pagina Moodle dell'insegnamento.

Testi consigliati:

- M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole. Fondamenti di Bioinformatica. Zanichelli, 2018
- M.M. Triola, M.F. Triola, J. Roy. Fondamenti di statistica per le discipline biomediche (seconda edizione). Pearson, 2022
- M.C. Whitlock, D. Schluter. Analisi statistica dei dati biologici. Zanichelli, 2022

### **Periodo di erogazione dell'insegnamento**

Primo semestre.

## **Modalità di verifica del profitto e valutazione**

Prova scritta della durata di 90 minuti, costituita da:

- 12 domande chiuse a risposta multipla (per la verifica dell'apprendimento del programma del corso), di cui 6 relative ad argomenti di bioinformatica e 6 relative ad argomenti di biostatistica;
  - 2 domande a risposta aperta (per il controllo estensivo della comprensione e della capacità di analisi critica), di cui 1 relativa ad argomenti di bioinformatica e 1 relativa ad argomenti di biostatistica.
- Le 14 domande copriranno tutti gli argomenti trattati sia durante le lezioni frontali che le esercitazioni in laboratorio.

Non verranno svolte prove in itinere.

## **Orario di ricevimento**

Su appuntamento previa richiesta via e-mail.

## **Sustainable Development Goals**

SALUTE E BENESSERE | ISTRUZIONE DI QUALITÁ | IMPRESE, INNOVAZIONE E INFRASTRUTTURE

---