

## COURSE SYLLABUS

### Fundamentals of Bioinformatics and Biostatistics

2223-1-E0201Q081

---

#### Obiettivi

L'insegnamento si propone di spiegare i fondamenti della bioinformatica e della biostatistica, con una particolare attenzione alle applicazioni di carattere biotecnologico.

In particolare, saranno affrontati i principali problemi legati all'analisi di sequenze biologiche e alla ricostruzione di genomi (con particolare riferimento alle tecniche di sequenziamento di ultima generazione), all'analisi del trascrittoma, e alla predizione di strutture proteiche. Verranno inoltre introdotti i concetti base sulle tecniche di apprendimento automatico e intelligenza artificiale per applicazioni in contesto biologico.

Saranno inoltre trattati i principali metodi di statistica descrittiva e statistica inferenziale, per guidare lo studente nel processo che, a partire dal disegno di uno studio sperimentale, termina nell'interpretazione statistica e nella comprensione della rilevanza biologica dei risultati ottenuti.

#### 1. 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 relativamente alla scelta dei metodi statistici più adeguati per l'analisi di dati in ambito biologico/clinico, e per l'interpretazione dei rispettivi risultati.

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

Al termine dell'insegnamento lo studente dovrà essere in grado di applicare le conoscenze acquisite al punto 1 per la soluzione di problemi in ambito computazionale e statistico, oltre a:

- organizzare e trattare automaticamente i dati biologici o risolvere semplici problemi tramite l'utilizzo di fogli di calcolo (esercitazioni in laboratorio);
- sviluppare semplici algoritmi e programmare in linguaggio Python (esercitazioni in laboratorio).

#### 3. 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.

#### 4. 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.

#### 5. 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 base di metodi euristici di ispirazione biologica, apprendimento automatico e intelligenza artificiale.  
Piattaforme di sequenziamento, allineamenti di sequenze, ricostruzione di genomi.  
Analisi del trascrittoma.  
Analisi di strutture proteiche.

Biostatistica:

Statistica descrittiva.  
Statistica inferenziale.

Esercitazioni in laboratorio:

Fogli di calcolo e ricerche in banche dati biologiche.  
Elementi di programmazione in Python.

## Programma esteso

Bioinformatica

- Pensare in modo algoritmico ed elementi di programmazione. 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.
- 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 del trascrittoma, identificazione di geni differenzialmente espressi.
- Analisi e predizione 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 dei dati. Ricerca in banche dati biologiche. Elementi di programmazione in Python.

## **Prerequisiti**

Prerequisiti: nessuno.

Propedeuticità: nessuna.

## **Modalità didattica**

Lezioni frontali in aula (40 ore) ed esercitazioni in aula di calcolo (30 ore), supportate da presentazioni PowerPoint. Tutta l'attività didattica verrà videoregistrata e resa disponibile tramite la piattaforma Moodle.

## **Materiale didattico**

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

Testi consigliati:

- P.J. Deitel, H.M. Deitel. Introduzione a Python. Per l'informatica e la data science. Pearson, 2020
- M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole. Fondamenti di Bioinformatica. Zanichelli, 2018
- F. Amaldi, P. Benedetti, G. Pesole, P. Plevani. Tecniche e metodi per la biologia molecolare. Casa Editrice Ambrosiana - Zanichelli, 2020
- 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**

Esame scritto della durata di 2 ore, con 9 domande a risposta multipla e 1 domanda a risposta aperta, riguardanti gli argomenti trattati durante le lezioni frontali e 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 QUALITÀ | IMPRESE, INNOVAZIONE E INFRASTRUTTURE

---