



INFORMATICA E STATISTICA MEDICA

9 CFU - 1° semestre

Docente titolare dell'insegnamento

ALFREDO PULVIRENTI

Email: apulvirenti@dmi.unict.it

Edificio / Indirizzo: Stanza 35, Terzo Blocco Dipartimento di Matematica e Informatica.

Telefono: 095-7383087

Orario ricevimento: Martedì 10-11.

OBIETTIVI FORMATIVI

▪ INFORMATICA

Obiettivo del corso è l'acquisizione di metodi per l'analisi di sequenze e strutture biologiche e per la ricerca in database biologici (es. geni, sequenze, domini funzionali). Partendo da sequenze primarie di acidi nucleici o proteine è possibile ipotizzarne la funzione, la storia evolutiva e la struttura. Gli strumenti utilizzati per raggiungere questi obiettivi sono i database pubblici e i programmi di analisi e visualizzazione.

1. **Conoscenza e capacità di comprensione (knowledge and understanding):** Gli studenti acquisiranno una conoscenza sui metodi per l'analisi di sequenze biologiche e per la ricerca in database biologici. In particolare approfondiranno la ricerca su database di sequenze, di domini, ed una buona familiarità con i database pubblici e i programmi di analisi e visualizzazione. Infine gli studenti potranno acquisire gli strumenti di base per l'analisi del trascrittoma.
2. **Capacità di applicare conoscenza e comprensione (applying knowledge and understanding):** identificare gli strumenti idonei per manipolare i dati ed estrarre la conoscenza sottostante; risolvere problemi attraverso l'uso di software opportuni in ambito bioinformatico.
3. **Autonomia di giudizio (making judgements):** Attraverso le esercitazioni guidate, gli studenti acquisiranno le competenze di base necessarie per affrontare l'analisi di nuove sequenze biologiche, ipotizzandone la funzione, studiare il trascrittoma.
4. **Abilità comunicative (communication skills):** lo studente acquisirà le necessarie abilità comunicative e di appropriatezza espressiva nell'impiego del linguaggio tecnico nell'ambito generale dell'analisi dei dati biologici.
5. **Capacità di apprendimento (learning skills):** il corso si propone, come obiettivo, di fornire allo studente le necessarie metodologie di base teoriche e pratiche per poter affrontare e risolvere autonomamente problemi nell'ambito dell'analisi dei dati biologici.

▪ STATISTICA MEDICA

Descrizione generale sintetica

Il corso mira a far acquisire i principali concetti di base del calcolo delle probabilità e della statistica.

Obiettivi formativi generali dell'insegnamento in termini di risultati di apprendimento attesi:

1. **Conoscenza e capacità di comprensione (knowledge and understanding):** Il corso mira a far acquisire agli studenti abilità in merito alla descrizione di dati statistici; Comprendere i termini di base (popolazione, campione, variabile ecc); Calcolo e presentazione di distribuzioni di frequenza; Descrizione di dati con metodi grafici; Calcolo degli indici di tendenza centrale e variabilità; Comprendere i fondamenti della valutazione di probabilità di un evento e di una variabile aleatoria; Acquisire i concetti legati alla statistica inferenziale quali stime per intervalli di confidenza e test di ipotesi.
2. **Capacità di applicare conoscenza e comprensione (applying knowledge and understanding):** identificare le distribuzioni idonee a rappresentare la conoscenza sottostante; risolvere problemi di statistica inferenziale e calcolo delle probabilità.
3. **Autonomia di giudizio (making judgements):** Attraverso esempi concreti e casi di studio, lo studente sarà in grado di elaborare autonomamente soluzioni a determinati problemi e valutare l'idoneità della soluzione di un problema di statistica inferenziale e probabilità.
4. **Abilità comunicative (communication skills):** lo studente acquisirà le necessarie abilità comunicative e di appropriatezza espressiva nell'impiego del linguaggio tecnico nell'ambito generale dell'analisi dei dati attraverso metodi statistici
5. **Capacità di apprendimento (learning skills):** il corso si propone, come obiettivo, di fornire allo studente le necessarie metodologie teoriche e pratiche per poter affrontare e risolvere autonomamente problemi nell'ambito dell'analisi statistica dei dati.

MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

▪ INFORMATICA

Le lezioni sono tenute in aula con l'ausilio di slide, messe a disposizione degli studenti sul portale Studium. Le slide non sostituiscono i testi di riferimento, ma, oltre che agevolare la comprensione della lezione, forniscono un dettaglio puntuale sul programma svolto.

Le lezioni frontali teoriche sono intervallate da esercitazioni pratiche, svolte nella stessa aula di lezione. Gli studenti sono invitati a formare piccoli gruppi di lavoro (massimo 2-3 persone) per lo svolgimento delle esercitazioni proposte.

▪ STATISTICA MEDICA

Le lezioni sono tenute in aula con l'ausilio di slide, messe a disposizione degli studenti sul portale Studium. Le slide non sostituiscono i testi di riferimento, ma, oltre che agevolare la comprensione della lezione, forniscono un dettaglio puntuale sul programma svolto.

Le lezioni frontali teoriche sono intervallate da esercitazioni pratiche, svolte nella stessa aula di lezione. Gli studenti sono invitati a formare piccoli gruppi di lavoro (massimo 2-3 persone) per lo svolgimento delle esercitazioni proposte.

PREREQUISITI RICHIESTI

▪ **INFORMATICA**

Per la piena comprensione dei contenuti del corso non sono necessari prerequisiti.

▪ **STATISTICA MEDICA**

Conoscenze di matematica presenti in tutti i programmi delle scuole superiori.

FREQUENZA LEZIONI

▪ **INFORMATICA**

La frequenza delle lezioni è obbligatoria.

▪ **STATISTICA MEDICA**

La frequenza delle lezioni è obbligatoria.

CONTENUTI DEL CORSO

▪ **INFORMATICA**

Il corso è organizzato in lezioni che prevedono una base teorica affiancata a esercitazioni i per l'apprendimento dell'uso di programmi di analisi e visualizzazione dei risultati.

PROGRAMMA:

Introduzione alla bioinformatica:

-Cenni introduttivi

Pietre miliari della bioinformatica

Infrastrutture bioinformatiche

Indirizzi web delle più rilevanti infrastrutture bioinformatiche

Banche dati biologiche:

Introduzione

Sistemi di interrogazione delle banche dati biologiche

SRS

Entrez

Banche dati primarie e banche dati specializzate

Banche dati di sequenze nucleotidiche

Banche dati di sequenze proteiche

Banche dati di motivi e domini proteici

Banche dati di geni

Banche dati del trascrittoma

Banche dati di profili di espressione

Banche dati di polimorfismi e mutazioni

Banche dati di pathways metabolici

Indirizzi web per banche dati e risorse biologiche

Allineamento di sequenze di acidi nucleici e proteine

Introduzione al problema degli allineamenti di sequenze

Similarità di sequenze e algoritmi di allineamento

Allineamenti di sequenze biologiche con gap

Visualizzazione di dot matrix

Le matrici di sostituzione

Le matrici PAM

Le matrici BLOSUM

Metodi di allineamento esatto

Allineamenti globali e locali

Algoritmi dinamici di allineamento

L'algoritmo di Smith e Waterman per la ricerca di similarità locali

Metodi euristici di allineamento

BLAST

Principali risorse disponibili in rete

Allineamento multiplo di sequenze

Introduzione

Algoritmi per l'allineamento multiplo

Misura della qualità di un allineamento multiplo

Strumenti per la visualizzazione e manipolazione dei multiallineamenti

Principali risorse disponibili in rete

▪ **STATISTICA MEDICA**

Le tecniche di campionamento: il campione casuale semplice, sistematico, stratificato, a grappoli, a stadi.

Statistica descrittiva. Distribuzioni di frequenza. Tabelle. Grafici, istogrammi, diagramma a barre. Indici di tendenza centrale. Indici di dispersione.

Cenni di calcolo delle probabilità. Definizione di probabilità; Eventi; Probabilità condizionata; Teorema di Bayes; Principio delle probabilità totali; Variabili aleatorie discrete; Media, varianza e deviazione standard; Distribuzioni discrete notevoli: Bernoulli, binomiale, uniforme, geometrica, Poisson; Distribuzioni continue: uniforme, normale esponenziale.

Distribuzioni di campionamento: distribuzione t-student, distribuzione di Fisher;

Inferenza statistica: stima puntuale e stima di intervallo. Intervallo di confidenza: per una media, per la differenza tra due medie, per una proporzione, per la differenza tra due proporzioni.

Principi della verifica di ipotesi (l'ipotesi nulla. Gli errori di I tipo e II tipo. La potenza di un test).

Test parametrici per le verifica di ipotesi: una media; confronto tra due medie nel caso di campioni indipendenti e appaiati.

TESTI DI RIFERIMENTO

▪ **INFORMATICA**

Libri di testo

- Anna Tramontano "Bioinformatica", Zanichelli
- Krane, Raymer. "Fondamenti di Bioinformatica" Pearson
- Jambeck, Gibas "Developing Bioinformatics Computer Skills, O'Reilly
- Pascarella-Paiardini "Bioinformatica" Zanichelli

▪ **STATISTICA MEDICA**

Testi di riferimento:

Lantieri PB, Riso D, Ravera G: Statistica medica per le professioni sanitarie, II ed. McGraw-Hill

ALTRO MATERIALE DIDATTICO

▪ **INFORMATICA**

Sul portale studium.unict.it saranno forniti i lucidi delle lezioni.

▪ STATISTICA MEDICA

Il materiale didattico viene messo a disposizione degli studenti su studium.unict.it

PROGRAMMAZIONE DEL CORSO

INFORMATICA

Argomenti	Riferimenti testi
1 Introduzione alla bioinformatica: tipi di dati, problemi, strumenti.	materiale didattico fornito dal docente
2 Sequenze, ricerca tramite BLAST, allineamento pairwise e multiplo. Algoritmi.	materiale didattico fornito dal docente
3 Attività pratica su allineamento di sequenze	materiale didattico fornito dal docente
4 Banche dati biologiche presenti sul sistema dell'NCBI: nucleotide, protein, OMIM, PUBMED, GENE, SNP	materiale didattico fornito dal docente
5 Attività pratica su banche dati	
6 Banca dati UNIPROT	materiale didattico fornito dal docente
7 Attività pratica su uniprot	

STATISTICA MEDICA

Argomenti	Riferimenti testi
1 Introduzione, statistica descrittiva	Capitoli 1 e 3 + materiale didattico integrativo
2 Raccolta e organizzazione dei dati, Indici di tendenza centrale e dispersione	Capitoli 4 5 e 6 + materiale didattico integrativo
3 calcolo delle probabilità e distribuzioni di probabilità	Capitolo 8 + materiale didattico integrativo
4 Rappresentazione grafica dei dati	Capitolo 7 + materiale didattico integrativo
5 Campionamento e inferenza statistica	Capitolo 10 + materiale didattico integrativo
6 Stime di parametri per intervallo	
7 Test di ipotesi	

VERIFICA DELL'APPRENDIMENTO

MODALITÀ DI VERIFICA DELL'APPRENDIMENTO

▪ **INFORMATICA**

L'esame finale consiste in **una prova scritta** ed un **colloquio orale**.

La prova scritta è costituita da esercizi e domande di teoria.

Chi non supera la prova scritta, non può sostenere l'orale. La prova scritta può essere visionata prima delle prove orali.

Salvo diversa comunicazione:

- l'esame scritto si svolge alle **ore 9:00**

Note:

- Per sostenere gli esami è **obbligatorio prenotarsi** utilizzando l'apposito modulo del portale CEA.
- Non sono ammesse prenotazioni tardive tramite email. In mancanza di prenotazione, l'esame non può essere verbalizzato.

▪ **STATISTICA MEDICA**

L'esame finale consiste in **una prova scritta** ed un **colloquio orale**.

La prova scritta è costituita da esercizi e domande di teoria.

Chi non supera la prova scritta, non può sostenere l'orale. La prova scritta può essere visionata prima delle prove orali.

Salvo diversa comunicazione:

- l'esame scritto si svolge alle **ore 9:00**

Note:

- Per sostenere gli esami è **obbligatorio prenotarsi** utilizzando l'apposito modulo del portale CEA.
- Non sono ammesse prenotazioni tardive tramite email. In mancanza di prenotazione, l'esame non può essere verbalizzato.

ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI

▪ **INFORMATICA**

Durante il corso saranno forniti diversi esercizi risolti che verranno pubblicati sul portale studium.unict.it

▪ **STATISTICA MEDICA**

Sul portale studium.unict.it saranno messi a disposizione esercizi svolti.
