

ALLEGATO "A1"

Modulo di esercitazioni nell'ambito del Corso di Alta Formazione in "Statistics: the art and science of learning from data"

A.A. 2025/26

Destinatari: Dottorandi

Durata: 5 ore

COMPENSO: 50,00 €/ora

Il corso di *Statistics: the art and science of learning from data* è strutturato in modo da fornire una solida base teorica, arricchita da una forte componente pratica, per permettere ai dottorandi di acquisire e perfezionare le competenze necessarie a procedere in autonomia alla gestione dell'analisi di dati in ambito biomedico.

L'approccio didattico prevede una varietà di metodologie formative, dall'aula tradizionale alle sessioni online, dai laboratori multimediali alle esercitazioni pratiche, garantendo un apprendimento esperienziale e interattivo.

Obiettivi Formativi:

Gli studenti impareranno a:

- o Disegnare un protocollo sperimentale
- o Impostare l'analisi dei dati in maniera corretta
- o Trarre inferenze statisticamente valide
- o Rappresentare graficamente i risultati

Prerequisiti:

Non sono richiesti prerequisiti specifici, se non un'ottima conoscenza della lingua inglese.

Contenuti:

1. Statistica descrittiva

- Media, mediana, moda
- Varianza e deviazione standard
- Percentili e intervalli
- Grafici (istogrammi, boxplot)

2. Statistica inferenziale

- Test di ipotesi (t-test, chi-quadrato, ANOVA)
- p-value e intervalli di confidenza
- Confronto tra gruppi biologici

3. Analisi univariata

Analizza **una sola variabile alla volta**.

- Distribuzione dei dati
- Normalità
- Analisi della variabilità

4. Analisi multivariata

- Pattern biologici
- Riduzione della dimensionalità
- Classificazione dei campioni

5. Correlazione

- Correlazione di Pearson (lineare)
- Correlazione di Spearman (non parametrica)

6. Regressione

- Regressione lineare semplice
- Regressione multipla
- Regressione logistica (molto usata in biologia/medicina)

7. Clustering (Grouping)

- Clustering gerarchico
- k-means
- Analisi di similarità genetica/ecologica

8. PCA (Principal Component Analysis)

- Riassume molte variabili in poche componenti principali
- Evidenzia pattern nascosti nei dati biologici
- Molto usata in genomica, ecologia e metabolomica

Metodi Didattici:

Lezioni Frontali: introduzione teorica ai concetti chiave

Esercitazioni Pratiche: simulazioni su popolazioni dati create ad hoc e da trial sperimentali

Laboratori Multimediali: utilizzo di software per analisi statistiche

Team Work: lavori di gruppo per la preparazione di progetti e presentazioni.

Prove pratiche tra vari team per favorire l'apprendimento collaborativo e il feedback costruttivo.

Verifica dell'Apprendimento:

Test di Valutazione: test scritto finale sulle competenze acquisite.

Presentazioni Orali: valutazione delle competenze di attraverso colloqui individuali e di gruppo.

Materiale Didattico:

Slide delle lezioni, video, articoli scientifici e manuali di riferimento saranno forniti attraverso una piattaforma online dedicata, risorse online per approfondimenti individuali.

Tempistiche:

Le lezioni si svolgeranno, presumibilmente, nel mese di giugno 2026