**Metodi statistici per l’analisi di dati di Spettrometria di Massa**

L’utilizzo della spettrometria di massa si sta via via sempre più diffondendo come strumento per lo studio di miscele complesse e sta assumendo notevole rilevanza in molti ambiti che spaziano dalla medicina, all’ambiente, al comparto alimentare.

Una adeguata analisi dei dati generati dall’applicazione della spettrometria di massa è la naturale premessa a garanzia di una corretta interpretazione dei risultati ottenuti. Per questo, c’è una crescente necessità di fare formazione non solo sugli aspetti pratici di laboratorio, ma anche sui metodi statistici più adeguati alla gestione e analisi dei dati e all’interpretazione e comunicazione dei risultati ottenuti.

La scuola è parte di una serie di corsi di formazione pratica organizzati dalla Divisione di Spettrometria di Massa (DSM-SCI) ed è una iniziativa che vede la sinergia con la Società Italiana di Biometria e il Dipartimento di Medicina e Chirurgia dell’Università di Milano - Bicocca.

Il corso è articolato in due giornate e mezzo e prevede nel pomeriggio del primo giorno, l’introduzione dell’open software R e del suo utilizzo. Verranno spiegate le funzioni base di R per l’importazione dei dati e la loro gestione e per l’analisi descrittiva, inclusiva della parte di rappresentazione grafica.

Il secondo e terzo giorno saranno invece dedicati all’elaborazione dei dati di spettrometria di massa e all’interpretazione dei risultati ottenuti. Le lezioni saranno divise in sessioni teoriche e pratiche, dove verranno utilizzati dati reali. Nello specifico, prima verrà spiegata l’importanza del pre-processamento dei dati ed i metodi relativi (i.e. *baseline, smoothing, normalization, alignment e peak detection*). Successivamente, ci si concentrerà sui metodi statistici *unsupervised* e *supervised*. Alcune delle tecniche che verranno trattate riguardano metodi per la riduzione della dimensionalità del dato e la rappresentazione grafica in spazi bi e tri-dimensionali, quali *Principal Component Analysis (PCA),* t-distributed stochastic neighbor embedding (*t-SNE) e Uniform Manifold Approximation and Projection (UMAP), Partial Least-square Discriminant Analysis (PLS-DA)*. Verranno poi affrontati i metodi *unsupervised* per il clustering dei dati come *k-means, hierarchical clustering analysis e community detection*. Infine verranno discussi i metodi *supervised* di classificazione come la regressione logistica e multinomiale, ponendo l’enfasi sull’interpretazione dei risultati dei modelli e sulle differenze rispetto ai metodi di analisi univariata.

Verrà fornito materiale didattico per le sessioni teoriche e per le esercitazioni pratiche.

Per le esercitazioni è richiesto un laptop personale, con R e RStudio installato

**Calendario**: 10-12 Maggio 2023

**Docenti**

Prof. Davide Risso Università di Padova

Dott. Giulia Capitoli Università di Milano -Bicocca

**Programma**

|  |  |  |
| --- | --- | --- |
| **10 Maggio 2023** | | |
| **Giorno 1** | 13.00-13.45 | **Registrazione** |
| 13.45-14.00- | **Introduzione al Corso**  F. Magni e S. Galimberti |
| **Introduzione al software R**  Docente: G.Capitoli | |
| 14.00-15.30 | **Sessione pratica** |
| 15.30-16.00 | Coffee break |
| 16.00-17.30 | **Sessione pratica** |
| **11 Maggio 2023** | | |
| **Giorno 2** | **Il pre-processing dei dati di proteomica**  Docente: G.Capitoli | |
| 9.00-10.30 | **Sessione teorica** |
| 10.30-11.00 | Coffee break |
| 11.00-12.30 | **Sessione pratica in R** |
| 12.30-14.00 | Pranzo |
| **Analisi Unsupervised: metodi per la riduzione di dimensionalità** | |
| 14.00-15.30 | **Sessione teorica**  Docente: D.Risso |
| 15.30-15.45 | Coffee break |
| 15.45-17.15 | **Sessione pratica in R**  Docente: G.Capitoli |
| **12 Maggio 2023** | | |
| **Giorno 3** | **Analisi Unsupervised: metodi di clustering** | |
| 9.00-10.30 | **Sessione teorica**  Docente: D.Risso |
| 10.30-11.00 | Coffee break |
| 11.00-12.30 | **Sessione pratica in R**  Docente: G.Capitoli |
| 12.30-14.00 | Pranzo |
| **Analisi Supervised: dal t-test al modello di regressione penalizzato** | |
| 14.00-15.30 | **Sessione teorica**  Docente: D.Risso |
| 15.30-15.45 | Coffee break |
| 15.45-17.15 | **Sessione pratica in R**  Docente: G.Capitoli |