

Descrizione del Corso Base

Si tratta essenzialmente di un percorso di introduzione alla biostatistica che si articola in quattro giornate. Durante la prima giornata si fornisce una panoramica generale delle domande pertinenti alla biostatistica, introducendo con degli esempi concreti anche di ricerche svolte presso il nostro Ateneo i concetti chiave della metodologia di analisi, illustrando come la statistica venga utilizzata in medicina, in biologia e nelle scienze della Vita in generale. Inoltre, vengono presentati alcuni 'dataset' di dati di esempio che verranno utilizzati durante il corso, discutendo vengono discussi alcuni software gratuiti molto diffusi, come il linguaggio R, e l'interfaccia grafica JASP. Daremo uno sguardo approfondito anche sulle tematiche importanti della qualità dei dati e della privacy dei dati. Condurremo anche delle semplici analisi descrittive, e discuteremo su come scegliere gli opportuni indicatori di centralità e di dispersione dei dati.

Nella seconda giornata, affronteremo "l'aleatorietà" dei dati biomedici inquadrandoli nelle teorie del calcolo della probabilità, presentando le caratteristiche delle variabili aleatorie più comuni utilizzate in biostatistica, come la distribuzione normale, quella binomiale e quella di Poisson. Discuteremo delle applicazioni del teorema di Bayes, in una visione "temporale" a priori / a posteriori ed associandolo alla possibilità statistica di valutare le associazioni degli eventi ed alla valutazione dei rischi. Infine, discuteremo di come si possa quantificare l'affidabilità di una "congettura statistica", quale la stima della media incognita di una misura di popolazione per mezzo delle misure di variabilità dei dati di un campione, ossia del concetto fondamentale dell'inferenza statistica, alla luce della visione classica / frequentista e della più attuale visione bayesiana.

Durante la terza giornata ci concentreremo sulle varie tecniche statistiche che consentono di trarre conclusioni su una popolazione più ampia basandosi su dati di un campione. Passeremo in rassegna i dettagli del test t di Student, un test statistico fondamentale utilizzato per confrontare le medie di due gruppi. Spiegheremo come ad esso si riferiscano anche i diversi tipi di test t (ad esempio, test t per un campione, test t per campioni abbinati) e discuteremo i limiti del concetto di significatività statistica e di significatività clinica. Introduciamo anche i test t bayesiani come alternativa ai test t classici, discutendone la diversa impostazione conoscitiva. Ci eserciteremo ad analizzare differenze tra due o più gruppi, descrivendo diversi test statistici quali il test di Welch, il test di Mann-Whitney e le tecniche di analisi della varianza (ANOVA) ad "una via" e "a due vie". Affronteremo anche il problema dei confronti multipli e di come si possa agire per controllare l'eteroschedasticità dei dati.

Infine la conclusione del corso sarà dedicata all'introduzione ai modelli statistici, quali strumenti più avanzati e più adatti all'analisi multivariata. Presenteremo il concetto di modello lineare e descriveremo come esso venga utilizzato per analizzare le relazioni tra variabili. Spiegheremo come e perché si debba misurare la dispersione dei dati rispetto ad un modello statistico regressivo, al fine di adattare il modello di regressione ai dati. Infine, discuteremo la necessità di operare una diagnosi del modello statistico per verificarne l'adeguatezza, presentando l'analisi della covarianza (ANCOVA). Daremo un cenno di come si operi la selezione del modello minimale adeguato per mezzo del criterio di informazione di Akaike, trattando della regressione logistica.