

Programma del corso
Metodi Matematici e Informatici per la Biologia
A.A. 2018-2019
Prof. G. Cavallaro, G. Basile

Introduzione alla statistica. La statistica in biologia e lo studio quantitativo della variabilità. Pianificazione di una ricerca in biologia. Passi principali del disegno di una ricerca in biologia. Metodi di raccolta dei dati: Sondaggi - Indagini campionarie: disegno campionario. Popolazione statistica e biologica. Campioni probabilistici e non. Definizione e scelta di un campione casuale semplice, di un campione a strati e di un campione a due o più stadi. Tavole dei numeri casuali. Parametri incogniti di una popolazione e statistiche campionarie. Errore di campionamento. Errore di copertura. Campioni non rappresentativi. Distorsione e variabilità in campionamenti ripetuti: come ridurle. Esperimenti: disegno di un esperimento. Disegno completamente randomizzato. Importanza della randomizzazione, delle repliche e del controllo in un disegno sperimentale. Variabili nascoste (effetti di confondimento). Esperimenti e relazioni causa-effetto. Studi di osservazione. Studi sul campo.

Introduzione all'analisi esplorativa dei dati. Tipi di variabili. Rapporti, percentuali, proporzioni, tassi. Distribuzioni di frequenze, frequenze relative e densità di una variabile. Rappresentazioni grafiche: grafici ramo-foglia, istogrammi, diagrammi a segmenti. Grafici a barre e grafici a torta. Le caratteristiche principali di una distribuzione: forma, centro, dispersione, outliers. Indici riassuntivi numerici di una variabile. Indici di posizione: media, moda, mediana, quartili, percentili. Frequenze cumulate percentuali. Indici di variabilità: intervallo di variazione, differenza interquartile, varianza e deviazione standard, coefficiente di variazione. Box-plot. Media e varianza campionarie: stimatori non distorti della media e varianza incognite di una popolazione. Proprietà di media e varianza. Una regola per individuare eventuali outliers. Indici di forma: curtosi, indice di simmetria.

Elementi di calcolo delle probabilità. Fenomeni aleatori e probabilità. Modelli probabilistici. Definizione frequentista della probabilità. Distribuzioni di probabilità di variabili aleatorie discrete: distribuzione uniforme, distribuzione binomiale, distribuzione di Poisson. Scelta di un modello opportuno per la modellizzazione di un fenomeno aleatorio osservato. Le distribuzioni di probabilità discrete come modelli di diffusione. Distribuzioni di probabilità di variabili aleatorie continue: la distribuzione normale. Dall'istogramma alla curva di densità normale caratterizzata da due parametri. Il modello normale. La regola 68-95-99,7. Unità standard e uso delle tavole della distribuzione normale standard. Plot dei quantili normali: come si costruisce e come si interpreta. Test del chi quadro.

Relazioni tra due variabili quantitative. Diagramma di dispersione. Coefficiente di correlazione di Pearson. Correlazione spuria.

Il modello di regressione lineare. La retta di regressione dei minimi quadrati. L'uso della retta di regressione per fare previsioni. I residui. Relazioni non lineari. Il controllo della bontà del modello. Grafico dei residui e analisi dei residui. Variabilità della y spiegata e non spiegata dal modello di regressione. Il coefficiente di determinazione. Osservazioni influenti e outliers. Le associazioni fra due variabili, le variabili nascoste e i rapporti di causa-effetto. Estrapolazione.

Vengono fornite dispense del corso.

Ciascun argomento è accompagnato da esercitazioni pratiche nel laboratorio di informatica.

Il software usato è Open Office Calc. Il materiale delle esercitazioni si può trovare sulla piattaforma elearning2, all'interno delle pagine dei docenti.