



# Dipartimento di Scienze e Tecnologie

ANNO ACCADEMICO 2015/2016

primo ANNO

CORSO DI STUDIO IN Scienze e Tecnologie Genetiche

INSEGNAMENTO Analisi Statistica dei Dati Biologici

DOCENTE prof. Stefano M. Pagnotta

**Le distribuzioni di frequenza.** Fenomeno, collettivo e variabile; il campione come sottoinsieme speciale del collettivo; le tipologie di variabili: quantitative e qualitative; le distribuzioni di frequenze assolute e frequenze relative; il diagramma a barre (e sue varianti), il diagramma a torta; le distribuzioni in classi; l'istogramma e sua versione perequata; calcolo delle percentuali sull'istogramma; il sistema delle frequenze cumulate e retrocumulate; cenni alla curva di sopravvivenza. *Questi argomenti sono trattati nei capitoli 1 e 4 del libro 3 in bibliografia.*

**Gli indici di posizione, variabilità e forma.** La caratterizzazione delle distribuzioni per posizione, variabilità e forma. La media aritmetica e sua proprietà. Calcolo della media aritmetica su dati grezzi e distribuzioni. La mediana e sue proprietà; calcolo su dati grezzi. Proprietà di resistenza della mediana. La moda. Rapporto fra media, mediana e moda. I percentili e il diagramma boxplot. La media troncata, la media geometrica. La nozione di variabilità. Indici elementari di variabilità. La varianza e la deviazione standard; calcolo e proprietà. Gli scarti semplici e quadratici. La variabilità per le variabili qualitative: l'eterogeneità/omogeneità. Gli indici di Gini e Shannon. Le trasformazioni lineari e la standardizzazione. Proprietà degli indici di posizione e variabilità rispetto alle trasformazioni lineari. Forma di una distribuzione: simmetria e asimmetria; indice di Fisher e indice di Solomon. *Questi argomenti sono trattati nei capitoli 5 e 6 del libro 3 in bibliografia.*

**Le distribuzioni doppie di frequenza.** La distribuzione doppia di frequenze assolute e relative. Le distribuzioni condizionali e marginali. I profili riga e i profili colonna. Rappresentazione grafica delle distribuzioni condizionate con diagrammi a barre parallele; rappresentazione grafica dei profili. La dipendenza assoluta, Le frequenze assolute sotto ipotesi di indipendenza. L'indice del chiQuadro, l'indice di contingenza quadratica e V di Cramer. *Questi argomenti sono trattati nei capitoli 12 (no 12.4) del libro 3 in bibliografia.*

**La dipendenza in media e la dipendenza lineare.** Dipendenza in media: media, varianza e devianza totali e condizionati. Le devianze totale, between e within. La decomposizione della devianza totale e il rapporto di correlazione di etaQuadro di Pearson. La spezzata di regressione. Il diagramma di dispersione e individuazione della dipendenza nel grafico. La covarianza e sua interpretazione come misura assoluta della dipendenza lineare. Proprietà della covarianza per trasformazioni lineari. Covarianza come generalizzazione della varianza. Calcolo della covarianza: il momento misto del II ordine. La correlazione lineare di Pearson: proprietà della covarianza per trasformazioni lineari. Stima della linea di tendenza. Muto rapporto fra dipendenza assoluta, in media e lineare. *Questi argomenti sono trattati nei capitoli 12 (no 12.4) del libro 3 in bibliografia.*

**Elementi di calcolo delle probabilità.** Mutuo rapporto fra prova evento e probabilità. Evento come risultato di un esperimento soggetto al caso. Spazio degli eventi e operazioni sugli eventi. Eventi compatibili e incompatibili, eventi elementari e composti. Partizioni di eventi. Probabilità secondo Kolmogorov; quantificazione oggettiva, frequentista e soggettiva. Regola additiva della probabilità; eventi indipendenti e regola moltiplicativa della probabilità. Eventi dipendenti, probabilità condizionata ed estensione della regola moltiplicativa della probabilità. Il modello bernoulliano, binomiale, geometrico; risoluzione delle urne con sottoprove indipendenti (con reintroduzione) e dipendenti (senza reintroduzione), il modello ipergeometrico. Il teorema delle probabilità totali, il teorema di Bayes. *Questi argomenti sono trattati nel capitolo 8 del libro 1 in bibliografia.*

**Le variabili casuali.** Nozione di variabile casuale. Il valore atteso e proprietà. Modelli discreti di vc: bernoulliano, uniforme, binomiale, geometrico, poisson e ipergeometrico. Modelli continui di vc: la normale e le distribuzioni derivate. Calcolo delle probabilità sulla normale.

**Cenni alla teoria dell'inferenza statistica e teoria della stima.** I paradigmi dell'inferenza statistica: stima puntuale, teoria dei test e stima intervallare. La popolazione e il campione casuale; la distribuzione di probabilità del campione casuale, la nozione di statistica; la distribuzione campionaria della media campionaria per la popolazione gaussiana, il teorema del limite centrale; la distribuzione campionaria della media campionaria per la popolazione bernoulliana. La nozione di stimatore e discussione sulle proprietà di sufficienza, non distorsione, efficienza e consistenza.

**Teoria dei test di significatività e stima intervallare.** Significato del test; le ipotesi; gli errori di primo e secondo tipo; le probabilità associate agli errori, il livello di significatività del test e la potenza. La statistica test, le regioni di accettazione e di rifiuto. Il p-value. I test sulla media: z-test e t-test; confronto fra le medie di due popolazioni; confronto fra le medie di K popolazioni (ANOVA). Il test sulla frequenza relativa e confronto fra due frequenze relative. I test del chiQuadro: il test di indipendenza, il test di buon adattamento, test sull'equilibrio di Hardy-Weinberg.

**Il modello di regressione semplice.** La formalizzazione del modello: la componente deterministica e la componente stocastica. Interpretazione dei parametri. Ipotesi deboli sul modello e la stima OLS (dei minimi quadrati) dei parametri e della varianza degli errori. Rappresentazione grafica della retta stimata. Calcolo dei valori predetti e dei residui; rappresentazione grafica dei residui. La decomposizione della devianza totale e l'indice di determinazione lineare (calcolo e interpretazione). Il teorema di Gauss-Markov. Il sistema delle ipotesi forti e la distribuzione campionaria degli stimatori dei parametri. Il test sul coefficiente angolare del modello.

## Bibliografia

Testi per studenti di primo anno

- 1) D. Piccolo - Statistica per le decisioni. La conoscenza umana sostenuta dall'evidenza empirica - (2010; 676 pagine, copertina verde, ISBN-10: 8815137335) Il Mulino
- 2) M. Blend - Statistica medica - (2013, ISBN-10: 8838786119) Apogeo
- 3) Di Ciaccio, Borra - Introduzione alla statistica descrittiva - (2003, ISBN-10: 8838607230) The McGraw-Hill Companies Italia