

Programma di teoria per l'esame di Metodi Statistici per la Bioingegneria 25/26:

- Statistica descrittiva (parte 1 nel canale B):
 - Tutto quello visto a lezione incluse le formule.
- Ripasso di probabilità e variabili aleatorie (parte 2 nel canale B):
 - Tutto quello visto a lezione incluse le formule.
- Modelli di variabili aleatorie (parte 3 nel canale B):
 - è richiesta la conoscenza delle formule solo per la variabile aleatoria normale.
- Indici di forma (parte 3 nel canale B):
 - richiesta la loro definizione ed interpretazione, range numerico di definizione (positivo, negativo, ecc..), non richieste le formule degli stimatori campionari.
- Test statistici parametrici (parte 4 e 5 nel canale B):
 - Nozioni generali sulla costruzione e applicazione di un test statistico.
 - Per tutti i test statistici visti a lezione (t-test omoschedastico, eteroschedastico, appaiato, ...) conoscere il sistema di ipotesi che verificano, le assunzioni su cui si basano, che distribuzione ha la statistica del test quando vale l'ipotesi nulla (sono esclusi i test di gaussianità).
 - E' richiesta la conoscenza della formula della statistica del test solo per il t test a un campione, il t-test per due campioni omoschedastici, e l'F test per verificare che due campioni normali abbiano la stessa varianza.
- Verifica dell'assunzione di normalità (parte 6 nel canale B):
 - Test di normalità/gaussianità: conoscere i nomi dei principali test, l'ipotesi che testano, ma non i dettagli delle statistiche utilizzate.
 - Cos'è e come interpretare un Q-Q plot.
 - Come interpretare gli indici di forma.
- Test statistici non parametrici (parte 6 nel canale B):
 - Tutto quello visto a lezione.
- Correzione per test multipli (parte 7 nel canale B)
 - Tutto quello visto a lezione.
- Prima parte sulla regressione lineare (parte 8 e 9 nel canale B):
 - Tutto quello visto a lezione, incluse le formule.

- Cose escluse:
 - la dimostrazione con cui si ricava la formula dello stimatore ai minimi quadrati.
 - Le formule delle statistiche dei test usati per generare i p-value associati alle stime dei parametri.
- Metodi per gestire la multicollinearità (parte 10 nel canale B):
 - Numero di condizionamento e indici di condizionamento
 - Analisi di correlazione
 - Metodo VIF
 - Esclusi altri metodi menzionati nella lezione del canale A.
- Confronto tra modelli in competizione (parte 10 nel canale B):
 - R-quadro adjusted, inclusa la formula
 - AIC e BIC, incluse le formule
 - F test, esclusa la formula della statistica del test.
- Minimi quadrati pesati ed equivalenza tra regressione e correlazione (parte 10 nel canale B):
 - Tutto quello visto a lezione, incluse le formule.
- Regolarizzazione (parte 11 nel canale B):
 - I 3 metodi di shrinkage visti a lezione incluse le formule della funzione obiettivo da minimizzare. Per Ridge conoscere anche la formula dello stimatore.
 - Metodi per la stima del parametro di regolarizzazione/iperparametri: conoscere a grandi linee come funzionano k-fold cross-validation e leave-one-out senza i dettagli sul loro funzionamento e le formule.
- ANOVA (parte 12 nel canale B):
 - ANOVA a una via: tutto quello visto a lezione incluse le formule.
 - ANOVA a due vie: tutto quello visto a lezione escluse le formule delle statistiche del test (fatte solo nel canale A)
- Regressione logistica (parte 13 nel canale B):
 - Tutto quello visto a lezione esclusa la formula dello stimatore di massima verosimiglianza.
- Analisi di sopravvivenza e modello di Cox (parte 14 e 15 nel canale B):
 - Obiettivi dell'analisi di sopravvivenza
 - Definizione del tempo di sopravvivenza

- Definizione di funzione di sopravvivenza, hazard function, cumulative hazard function. Legame tra funzione di sopravvivenza e hazard function
 - Il metodo Kaplan-Meier
 - Il log-rank test escluse le formule
 - Il modello di Cox, incluse le formule
 - Esclusi il modello esponenziale e il modello di Weibull (fatti solo nel canale A)
 - C-index: definizione, formula stimatore e interpretazione del valore.
- Clustering (parte 16 nel canale B):
 - Tutto quello visto a lezione eccetto:
 - Gap statistic ed elbow method (fatte solo canale A)
 - Formula dell'indice cophenetico (da sapere invece a cosa serve l'indice cophenetico e su che misure si basa)
 - PCA (parte 17 nel canale B):
 - Tutto quello visto a lezione escluso:
 - Passaggi matematici che dimostrano che i loadings sono gli autovettori della matrice di covarianza (dimostrazione fatta solo nel canale A)
 - Rotazione degli assi fattoriali (fatta solo nel canale A)

Argomenti esclusi dal programma d'esame perché trattati solo in uno dei canali:

- Argomenti fatti solo nel canale A:
 - Formule della statistica del test ANOVA a due vie
 - Modello esponenziale e modello di Weibull per l'analisi di sopravvivenza
 - Nella PCA, dimostrazione che i loadings sono gli autovettori della matrice di covarianza e la rotazione degli assi fattoriali
- Argomenti fatti solo nel canale B:
 - Codifica delle variabili qualitative mediante variabili dummy (parte 10 nel canale B)
 - Formule della scomposizione dell'errore del modello di regressione in bias e varianza (parte 11 nel canale B)
 - Equivalenza tra ANOVA e regressione lineare (parte 12 nel canale B)