

Programma di teoria per l'esame di Metodi Statistici per la Bioingegneria 24/25:

- Statistica descrittiva (parte 2 nel canale B):
 - Tutto quello visto a lezione incluse le formule.
 - Escluse media ponderata e media geometrica (fatte solo nel canale A).
- Ripasso di probabilità e variabili aleatorie (parte 3):
 - Tutto quello visto a lezione incluse le formule.
- Modelli di variabili aleatorie (parte 4 nel canale B):
 - è richiesta la conoscenza delle formule solo per la variabile aleatoria normale.
- Indici di forma (parte 4 nel canale B):
 - richiesta la loro definizione ed interpretazione, range numerico di definizione (positivo, negativo, ecc..), non richieste le formule degli stimatori.
- Test statistici parametrici (parte 5 e 6 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 due campioni, con varianze uguali (omoschedastico), e l'F test per verificare che due campioni normali abbiano la stessa varianza.
- Verifica dell'assunzione di normalità (parte 7 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 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.
 - La formula di RMSE.
 - 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
 - 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.
- 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.
- Minimi quadrati pesati ed equivalenza tra regressione e correlazione (parte 12 nel canale B):
 - Tutto quello visto a lezione, incluse le formule.
- Regressione logistica (parte 13 nel canale B):
 - Tutto quello visto a lezione esclusi i metodi per la stima dei parametri (massima verosimiglianza o minimi quadrati), deviance, likelihood ratio test e confronto tra modelli in competizione.
- 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:
 - Distanze diverse da quella euclidea (fatte solo canale A)
 - Fuzzy C-means (fatto solo nel canale A)
 - Formula dell'indice cophenetic (da sapere invece a cosa serve l'indice cophenetic e su che misure si basa)
 - Indice di inconsistenza (fatto solo nel canale A)
- ANOVA (parte 17 nel canale B):
 - Conoscere le assunzioni su cui si basa il test, l'ipotesi nulla che verifica e la distribuzione della statistica utilizzata (no formule).
- PCA (parte 18 e 19 nel canale B):
 - Concetti e passaggi principali del metodo, escluse le formule.
 - Metodi per la scelta del numero di componenti principali.
 - Esclusa la 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:
 - Media ponderata e media geometrica
 - Correlazione di Spearman
 - Formule dell'ANOVA a due vie
 - Altri metodi per gestire la multicollinearità nei modelli di regressione lineare non menzionati sopra
 - Metodi per la stima dei coefficienti del modello di regressione logistica
 - Modello esponenziale e modello di Weibull per l'analisi di sopravvivenza
 - Nel clustering: distanze diverse dalla euclidea, fuzzy C-means, indice di inconsistenza
 - Rotazione degli assi fattoriali nella PCA
 - Independent component analysis

- Argomenti fatti solo nel canale B:
 - Equivalenza tra ANOVA e regressione lineare
 - Codifica delle variabili qualitative mediante variabili dummy
 - Scomposizione dell'errore del modello di regressione in bias e varianza
 - Min-max scaling delle variabili
 - Metodi per la stima dei coefficienti del modello di regressione logistica
 - Nella regressione logistica: deviance, likelihood ratio test e confronto tra modelli in competizione.