

## Statistik und Bioinformatik

### Deskriptive Statistik:

1. Standardabweichung, Schiefe ausrechnen von -2, 0, 1, 1
2. Median, Mittelwert, 1. Und 3. Quartil in Histogramm einzeichnen
3. Wenn man 20% aus diesem Histogramm zufällig raus rauszieht -> wie ändert sich Varianz und Standardabweichung vom Mittelwert
4. Drei Boxplots (Kontrolle, Treatment 1 und 2 von Pflanzen-Gewicht) -> eine Box beschriften was der Strich in der Mitte ist, was die Form der Box ist, was die Whisker sind (-> Median, Quartile)
5. Skalenniveaus angeben von Gewicht und Behandlung bei den Pflanzen
6. Kann man aus den Boxplots schließen, dass es signifikante Unterschiede gibt, wenn sich die Boxen nicht überschneiden?

### Schließende Statistik:

1. Nullhypothese des t-Tests zum vorherigen Beispiel mit den Pflanzen, Verteilungsannahme
2. Was hat die ANOVA für eine Nullhypothese, was ist der Unterschied zum t-Test?
3. ANOVA aus R auswerten -> obs signifikant ist, ob man  $H_0$  ablehnen kann, wie würde man das in der Bachelorarbeit formulieren?
4. Beispielsatz aus Paper über p-Wert (Medikament ist zu 3% nicht wirksam, da p-Wert 3%), Fehler finden über Definition des p-Wertes
5. Beispiel -> welche Regression wurde gemacht? -> LM, obs signifikant ist, wo des steht
6. Beispiel Sterberate auf der Titanic Geschlecht / 2. Oder 3. Klasse -> GLM -> ob Unterschied zwischen den Mortalitäten der Klassen signifikant ist, wo kann man Konfidenzintervall ablesen
7. MLE -> was wird maximiert?
8. Formel FDR
9. Unterschied Fehler 1. Art und FDR

### Versuchsdesign:

1. Was ist ein Störfaktor und warum ist er ein Problem?
2. Zwei Kriterien, die Störfaktor beschreiben
3. Beispiel mit Zusammenhang Glyphosatexposition – Krebs -> Untersuchung bei Landwirten  
Warum ist Rauchen KEIN Störfaktor?
4. Was müsste passieren, damit Rauchen ein Störfaktor ist? Nenne ein Szenario, egal wie unwahrscheinlich
5. Was ist die Idee von Blockdesign und warum macht man des?
6. Was hat Einfluss auf Power und wie? (mit +, -, 0 markieren)?  
Stichprobengröße  
Datum des Versuchs  
Effektgröße  
Effektrichtung  
Varianz  
Balance

## Bioinformatik:

1. Smith-Waterman Algorithmus eintragen von A=YW und B=CY mit Gap -4 und BLOSUM 62 gegeben
2. BLAST-Ausgabe von signifikanten Protein-Sequenzen erklären, also Struktur von A erklären
3. Suffix-Baum zeichnen
4. Hidden Markov Modell Zustandsdiagramm zeichnen mit allen Wahrscheinlichkeiten (zwei Münzen)
5. HMM: Davon Viterbi-Variablen ausrechnen
6. HMM: MSA gegeben, Zahl der Matches angeben und Emissionswahrscheinlichkeit des Matches 7
7. Homologie-Modellierung: warum macht man das und auf welchem Prinzip beruht es? Wann ist Genauigkeit des Modells wichtiger? Bei Planung einer Punktmutation oder bei Substratspezifität?
8. Log10 odd Score, fünf Sequenzen gegeben, log10 odd score von A in der 5. Spalte berechnen, Sequenzlogo zeichnen
9. Phylogenetischer Baum: ist es eine additive Matrix? An zwei Kanten die Zahlen ergänzen
10. Quartett-Puzzle Zahlen an Kanten ergänzen, sagen wo man E einfügt
11. Nearman Pearson Lemma -> irgendetwas mit Berechnung, was die Scores sagen oder so ähnlich
12. Transkriptomik: erklären? O.ä.  
zwei Methoden zum Quantifizieren nennen
13. Python-Code: einmal was sagt der aus, was ist die Ausgabe und was ist die Laufzeit (begründen)
14. Anderer Python-Code: Syntax-Fehler finden