

Statistik und Bioinformatik 2026, 1. Termin, Gedächtnisprotokoll

Achtung: Fragen können in Klausur etwas anders gewesen sein (habe es aber versucht so gut es geht zu rekonstruieren)

generell: Klausuren sind oft recht ähnlich aufgebaut, größte Vielfalt an Aufgaben im Bereich deskriptive Statistik, mit Needleman-Wunsch/Smith-Waterman-Algorithmus kann man gut Punkte holen, Klausurvorbereitungs-Vorlesung in Statistik wichtig, auch Fragen-Tutorium wichtig, Fragestunde zu Bioinformatik auch empfehlenswert, so kann man möglicherweise schon Hints auf Klausurfragen bekommen

Statistik (25 Punkte)

Deskriptive Statistik (8 Punkte)

- a) Berechnen Sie Mittelwert, Varianz, Standardabweichung und Standardfehler von folgenden Werten: [-1, -1, 1, 1] (1 Punkt)
- b) Single-Choice-Frage zum 25-Quantil (1 Punkt)
- c) 3 Abbildungen von Korrelationen, welche Abbildung hat den geringsten Pearson-Korrelationskoeffizient (1 Punkt)
- d) Unterschied zwischen den Korrelationskoeffizienten die in Vorlesung behandelt wurden (Pearson, Spearman, Kendall) (1 Punkt)
- e) Single-Choice-Frage zu PCA Abbildung (1 Punkt)
- f) Skalenniveaus angeben von 2 Variablen (Variable 1 (Gewicht): numerisch/metrisch, Variable 2 (Behandlung): nominal) (1 Punkt)
- g) Abbildung mit drei Boxplots (Kontrolle, Treatment 1 und 2) beschriften, Strich in der Mitte, was gibt die Box an, Whisker beschriften (Achtung Whisker unterschiedlich lang, einmal Minimum, einmal $1,5 \cdot$ Interquartilsabstand (1 Punkte)
- h) Geben Sie ein Beispiel für eine Kreuztabelle an. (1 Punkt)

Schließende Statistik (11 Punkte)

(ziemlich identisch zu Altklausur 2024)

- a) Nullhypothese des t-Tests angeben, Verteilungsannahme angeben (1 Punkt)
- b) Nullhypothese der ANOVA, Unterschied zum t-Test (1 Punkt)
- c) ANOVA aus R auswerten, ist ANOVA signifikant?, wird H_0 abgelehnt?, wie würde man das in Bachelorarbeit angeben? (1 Punkt)
- d) p-Wert definieren (1 Punkt)
- e) mit welcher Methode kann man Wahrscheinlichkeit der Nullhypothese berechnen? (1 Punkt)
- f) Formel FDR angeben, kann man $FDR < 5\%$ erreichen, wenn ja: wie?, wenn nein: warum nicht? (1 Punkt)
- g) Abbildung aus R, Abhängigkeit zwischen Lufttemperatur und Wind, welche Form der Abhängigkeit wurde für Modell angenommen?, woran sieht man ob es eine Abhängigkeit gibt? (in Abbildung markieren), Abhängigkeit in welche Richtung?, wo wird CI angegeben? (2 Punkte)
- h) MLE definieren (1 Punkt)

i) Modell zu Überleben auf der Titanic von Geschlecht und Passagierklasse aus R, um welche GLM handelt es sich, dazu Verteilungsannahme und Linkfunktion (1 Punkt), gibt es einen signifikanten Unterschied zwischen 2. und 3. Klasse (1 Punkt)

Versuchsdesign (6 Punkte)

(ziemlich identisch zu Altklausur 2024)

- a) Was ist ein Störfaktor und warum ist er ein Problem? (1 Punkt)
- b) Ist Rauchen ein Störfaktor in einer Beobachtungsstudie zu Glyphosatexposition und Krebserkrankungen bei Landwirten? (1 Punkt)
- c) Für welche Annahme wäre Rauchen ein Störfaktor? (nachvollziehbare Annahme erfinden) (1 Punkt)
- d) Bei Ihrer Bachelor Arbeit hat Ihr Vorgänger ein ähnliches Experiment gemacht, warum können Sie seine Werte für die Kontrolle nicht übernehmen? (1 Punkt)
- e) Erklären warum randomisiertes Blockdesign dem einfachen Blockdesign vorzuziehen ist. (1 Punkt)
- f) Wie kann man Power/Validität durch Versuchsplanung erhöhen? (zwei Möglichkeiten aufzählen) (1 Punkt)

Bioinformatik (25 Punkte)

Sequenzvergleich (4 Punkte)

- a) Sie wollen eine Aminosäure durch eine andere ersetzen. Welche Aminosäure wählen Sie und warum? (1 Punkt)
Folgende Scores: $s(a, a_i) = +1$, $s(a, a_k) = 0$ $s(a, a_l) = -1$
- b) Sie wollen Aminosäure E durch eine möglichst unähnliche Aminosäure ersetzen. Welche wählen Sie und warum? Verwenden Sie dazu die Blosum80 Matrix. (1 Punkt)
- c) Vervollständigen Sie den nachfolgenden Needleman-Wunsch-Algorithmus. Füllen Sie die grauen Felder aus, die Werte können Sie der Blosum80 Matrix entnehmen. Gap Score = -3 (2 Punkte)

		X		
		
Y		-1	...	2
		
		1		

Strukturvorhersage (5 Punkte)

- a) Beschreiben Sie das Grundprinzip der Proteinfaltung, auf dem die theoretischen Methoden der Strukturvorhersage beruhen. (2 Punkt)
- b) Nennen Sie vier wichtige Schritte des Arbeitsablauf zur Homologiemodellierung und beschreiben Sie diese kurz. (2 Punkte)
- c) Welches zweite Konzept neben der Homologiemodellierung gibt es zur Strukturvorhersage? Benennen Sie dieses und beschreiben Sie es kurz. (1 Punkt)

Neuronale Netze (4 Punkte)

a) Sie haben folgende Wahrheitstabelle gegeben. Entwerfen Sie ein dazu passendes neuronales Netz und geben sie die zugehörige Schwellenwertfunktion und die jeweiligen Gewichte an. (3 Punkte)

x_1	x_2	y
0	0	1
0	1	1
1	0	1
1	1	0

b) Um welche Art von Funktion handelt es sich in der oberen Aufgabe? (1 Punkt)

Molekulardynamik (4 Punkte)

a) Was ist die bioinformatische Sichtweise eines Proteins? (1 Punkte)

b) Ist die Aussage richtig oder falsch? (0,5 Punkte)

Die Energieminimierung von Proteinsystemen führt normalerweise zu einem globalen Energieminimum.

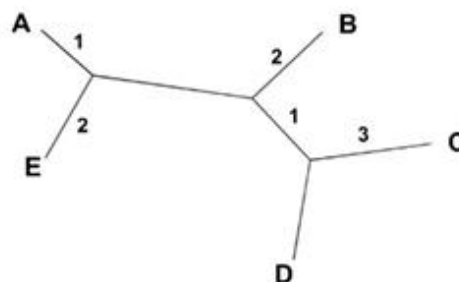
c) Nennen Sie vier der fünf Kraftfeldparameter/Wechselwirkungspotentiale. (2 Punkte)

d) Erläutern Sie kurz die konzeptionellen Näherungen (keine Formel) mit der Bindung zwischen zwei Atomen und die Atome selbst in der klassischen molekularmechanischen Simulation beschrieben werden. (0,5 Punkte)

Phylogenie (2 Punkte)

Gegeben sei für fünf Sequenzen A – E die links gezeigte Distanzmatrix. Können Sie nachweisen, dass dies eine additive Matrix ist? (1 Punkt) Tragen Sie die fehlenden Gewichte im rechts gezeigten Baum ein. (1 Punkt)

	A	B	C	D	E
A	0	7	9	7	3
B		0	6	4	8
C			0	4	10
D				0	8
E					0



(aus dem Übungszettel)

