

Übung zur Vorlesung Statistik II für Biowissenschaftler

Übungsblatt 1 - Testen und lineare Modelle

Dr. Jochen Kruppa
AG-Leiter Statistische Bioinformatik
Institut für Biometrie und Klinische Epidemiologie
Charité - Universitätsmedizin Berlin

Version vom 13. April 2021

Aufgabe 1

1. Eine Forscherin hat in einem Experiment mit neun Ratten ein Test-Score-Modell entwickelt, aus dem sie die Tumorgroße vorhersagen will. Nach dem Tod der Ratten ergaben sich folgende Test-Score-Werte und die dazugehörigen Tumorgroßen pro Tier:

Tumorgroße [cm ³]	12	16	22	24	10	8	14	21	12
Test-Score	2	3.5	5.5	4	2.5	1	3	5	1.5
Test-Score (dichotomisiert)									

2. Stellen Sie das passende biologische Modell des Rattenexperiments in der Form $y \sim x$ auf!
3. Zeichnen Sie zu dem Modell den passenden Scatterplot aus den obigen Daten mit korrekter Beschriftung:
 - eine Regressionsgerade durch die Punkte aus b)!
 - Skizzieren/veranschaulichen Sie die Methode, wie Sie die beste Gerade durch die Punkte bestimmt haben an drei Beispielpunkten.
4. Bestimmen Sie die notwendigen Parameter für die Regressionsgerade aus der Zeichnung in b).
5. Stellen Sie die Regressionsgleichung auf! Welche Tumorgroße ergibt sich aus dem Modell für einen Test-Score von 4?

6. Die Forscherin möchte den Test für die spätere Anwendung in zwei Gruppen (Tumor klein vs. Tumor groß) einteilen. Welchen Cut-Off schlagen Sie für den Test-Score vor? Begründen Sie ihre Entscheidung! Füllen Sie entsprechend diesem Cut-Off die Zeile "Test-Score (dichotomisiert)" in der obigen Tabelle aus.
7. Zeichnen Sie für Ihre Test-Score Dichotomisierung die passenden Boxplots für den Gruppenvergleich!

Aufgabe 2

Ein Forscher untersucht 10 Mäuse auf Gewichtsverlust und erhält folgende Tabelle.

Maus	1	2	3	4	5	6	7	8	9	10
Gewichtsverlust [g]	10	9	7	12	6	2	4	1	0	3

Welche Annahmen an die Daten und Modifikationen haben Sie vorgenommen um folgende statistischen Tests zu rechnen?

1. Ein Gruppenvergleich mit dem Student t-Test und einen gepaarter t-Test
2. Wie müsste die Datenstruktur für einen X^2 -Test aussehen? Wären 10 Mäuse genug oder wie viele Mäuse sind das Minimum (Stichwort Zellenbelegung).

Aufgabe 3

Eine Forscherin vergleicht 4 Dosisstufen (none, low, mid, high) für ein neues Medikament miteinander. Erstellen Sie hypothetische Konfidenzintervalle, die beispielhaft mögliche Interpretationen der Studie beinhalten.

1. Zeichnen Sie eine Abbildung der entstehenden Konfidenzintervalle, die alle paarweisen Vergleiche beinhaltet, wenn der Endpunkt "Gewichtsreduktion" lautet.
2. Zeichnen Sie eine Abbildung der entstehenden Konfidenzintervalle, die alle paarweisen Vergleiche beinhaltet, wenn der Endpunkt "Auftreten eines Tumors (ja/nein)" lautet.

Aufgabe 4

In der Teilchenphysik ist es so: Abweichungen zwischen Theorie und Experiment können auch einfach nur Zufall sein. Deswegen verwendet man dort den „Sigma-Wert“. In der Teilchenphysik ist es üblich, erst bei einem Wert von 5 Sigma von einer „Entdeckung“ zu sprechen. Beim aktuellen g-2-Experiment des Fermilab liegt der Wert bei 4.2.

- Mit welcher Wahrscheinlichkeit ist 5 Sigma vermutlich verbunden?
- Mit welcher Wahrscheinlichkeit ist dann 4.2 Sigma verbunden?
- Wie sehe die Entscheidung in der „normalen“ Statistik aus?
- Visualisieren Sie die Zusammenhänge.