

Aufgabe 1 Es könnte sein, dass mit Chlor versetztes Wasser (wie in Schwimmbädern üblich) den Zahnschmelz beeinträchtigt. Dazu wurden 200 Schwimmer je nach dem, ob sie mehr oder weniger als 6 Stunden wöchentlich trainieren, in zwei Gruppen eingeteilt und der Zahnschmelz untersucht, mit folgendem Ergebnis:

Schwimmzeit pro Woche	Zahnschmelz angegriffen		
	Ja	Nein	Gesamt
mehr als 6 h	29	71	100
weniger als 6 h	19	81	100
Gesamt	48	152	200

- a) Stützen die Beobachtungen die Hypothese des ersten Satzes der Aufgabenstellung? Formulieren Sie eine Nullhypothese, testen Sie mittels eines χ^2 -Tests und formulieren Sie einen Antwortsatz.
- b) Schätzen Sie den Anteil θ der Personen mit angegriffenem Zahnschmelz unter den Viel-Schwimmern. Geben Sie auch ein 95%-Konfidenzintervall an.
- c) Nehmen wir an, analoge Ergebnisse wären in einer Studie des doppelten Umgangs beobachtet worden: Je 200 Viel- und 200 Wenig-Schwimmer wurden untersucht, 58 der Viel- und 38 der Wenig-Schwimmer hatten angegriffenen Zahnschmelz. Welchen Wert würde die χ^2 -Statistik dann annehmen, und wie lautete dann Ihr Befund?
- d) Welches Konfidenzintervall für θ würde sich mit den Daten aus c) ergeben?

Aufgabe 2 Sie wollen zwei Hirsesorten hinsichtlich ihrer Trockenempfindlichkeit vergleichen. Dazu werden die Pflanzen 7 Tage lang nicht gegossen und dann ihre mittlere Blattfläche bestimmt. Wie viele Pflanzen von jeder Sorte brauchen Sie für den Versuch, wenn ein Blattflächenunterschied von 20 auf dem Signifikanzniveau 5% bei einer Testmacht von 80% erkannt werden soll? Verwenden Sie dazu, dass die Standardabweichungen in einem Vorversuch 29.55 und 60.21 waren.

- a) Berechnen Sie die benötigte Stichprobenlänge in jeder Gruppe ohne Verwendung von R. Verwenden Sie dazu die Verteilungstafeln von der Vorlesungs-Homepage.
- b) Berechnen Sie die benötigte Stichprobenlänge in jeder Gruppe mit R.
- c) Angenommen Sie hätten die benötigte Stichprobenlänge nicht bestimmt und sich aus dem Bauch heraus für 20 Pflanzen pro Sorte entschieden. Wie groß ist dann die Testmacht? Drücken Sie diese Testmacht in Worten als relative Häufigkeit aus.

Aufgabe 3 In den Flüssen Trinidads sind die Guppys *Poecilia reticulata* verschiedenen Räubern ausgesetzt. Flussabwärts gibt es den großen und gefährlichen *Crenicichla alta*, flussaufwärts den weniger gefährlichen *Rivulus hartii*. Evolutionär-ökologische Überlegungen führen zu der Hypothese, dass die Nachkommen flussabwärts kleiner sind. Um diese Hypothese zu prüfen, werden Sie nach Trinidad fliegen, um das Gewicht von neugeborenen Guppys zu messen. Wie viele neugeborene Guppys müssen Sie vermessen, wenn ein tatsächlicher mittlerer Gewichtsunterschied von 0.4 mg auf dem Signifikanzniveau 5% erkannt werden

soll? Die Testmacht soll 99% Prozent sein, da die Messreihe wegen den Reisekosten recht teuer ist. Aus Voruntersuchungen wissen Sie, dass Sie in den Daten mit einer Standardabweichung von 0.5 mg rechnen können.

a) Ermitteln Sie die benötigte Stichprobenlänge an den beiden Flussabschnitten ohne Verwendung von R. Verwenden Sie dazu die Verteilungstafeln von der Vorlesungs-Homepage.

b) Berechnen Sie die benötigte Stichprobenlänge an den beiden Flussabschnitten mit R.

c) Angenommen, die Messreihe würde – rein hypothetisch – 100 mal durchgeführt werden. In wie vielen Fällen erwarten Sie bei einer Testmacht von 99%, dass die Daten keine Signifikanz liefern, wenn der wahre mittlere Gewichtsunterschied tatsächlich 0.4 mg beträgt.

d) Angenommen Sie hätten die benötigte Stichprobenlänge nicht bestimmt und sich aus dem Bauch heraus für 20 Neugeborene pro Flussabschnitt entschieden. Wie groß ist dann die Testmacht? Drücken Sie diese Testmacht in Worten als relative Häufigkeit aus.

Aufgabe 4 Experimente müssen sorgfältig durchdacht werden, um Fehler zu vermeiden.

a) Im Raum München wurden im Laufe eines Tages 30 Eichhörnchen (*Sciurus vulgaris*) gefangen und durchnummeriert zum Zwecke eines Verhaltensperiments. Wie sollen die 30 Tiere auf zwei Beobachtungsgruppen aufgeteilt werden?

b) Sicherlich hat die jahreszeitliche Temperatur einen Einfluss auf das Nuss-Sammelverhalten der Eichhörnchen. Nun möchte man herausfinden, ob auch die Zusammensetzung des Lichtes einen Einfluss auf das Sammelverhalten hat. Die Idee ist, die eine Gruppe blauem Licht ausgesetzt und die andere Gruppe rotem Licht. Wie könnte der Versuch konkret ablaufen? Welche Fehler könnten gemacht werden, welche Einflussgrößen könnten das Ergebnis verfälschen?

Aufgabe 5 Um den Anteil $p \in [0, 1]$ von Personen in einer (großen) Gruppe, die eine „heikle“ Eigenschaft haben, zu schätzen, betrachten wir folgendes Verfahren: Als Teilnehmer erhalten Sie die Frage, beispielsweise „Haben Sie schon einmal Marihuana geraucht?“ und zwei Würfel. Sie würfeln verdeckt: Wenn die Augensumme 5 oder 6 ist, antworten Sie „ja“, wenn sie 8 oder 9 ist, antworten Sie „nein“, ansonsten beantworten Sie die Frage.

a) Wenn der tatsächliche Anteil der Personen, die die heikle Eigenschaft haben, p ist, wie wahrscheinlich ist es dann, dass eine rein zufällig ausgewählte Person mit „ja“ antwortet?

b) Wie kann man anhand der Ergebnisse der Befragung von n Personen den Parameter p erwartungstreu schätzen? Beweisen Sie, dass der Schätzer tatsächlich erwartungstreu ist.

c) Berechnen Sie die Varianz des erwartungstreuen Schätzers.

Aufgabe 6 In R können Sie mit `MASS::oats` einen Datensatz einlesen, der Erträge (Y , engl. yield) aus einem Versuch mit verschiedenen Hafersorten (V , engl. variety) mit verschiedenen verschiedenen Anbauflächen (B , Block) und verschiedenen Mengen an Stickstoffdünger (N) zeigt.

(a) Erstellen Sie eine Abbildung, die (ohne Modellanpassung) zeigt, wie der Ertrag von den anderen Faktoren abhängt.

(b) Testen Sie, ob der Effekt unterschiedlicher Mengen des Düngemittels von der Sorte abhängt. Berücksichtigen Sie in Ihrem Modell, dass die Anbauflächen auch einen Effekt auf das Ergebnis haben können.

- (c) Welcher Ertrag wäre für die verschiedenen Sorten auf den verschiedenen Blöcken bei einer Düngemittelmenge von 0.3 [cwt/acre] zu erwarten, nach welcher Formel wird diese Vorhersage berechnet und welche Annahme/Vorhersage macht das Modell zur Verteilung der Werte? Passen Sie zur Beantwortung dieser Frage ein Modell an die Daten an, in dem Sie die Düngemittelmenge als kontinuierliche Variable mit linearem Einfluss auf den Ertrag verwenden. Tipp: erzeugen Sie eine kontinuierliche Variable mit `rep(0:3/5, 18)`.
- (d) Beantworten Sie die Fragen aus dem vorherigen Aufgabenteil auch für ein Modell, in dem Sie einen parabelförmigen Einfluss der Düngemittelmenge auf den Ertrag erlauben, indem Sie die quadrierte Düngemittelmenge als weitere Variable dazunehmen.
- (e) Welches der Modelle aus den drei obigen Aufgabenteilen wird vom Akaike-Informationskriterium bevorzugt?
- (f) Überprüfen Sie für die obigen Modelle jeweils die Modellannahmen mit den üblichen Visualisierungen.
- (g) Visualisieren Sie die Modellvorhersagen für die verschiedenen Anbauflächen, Hafersorten und Düngemittelmenge.