Übungen zur Wahrscheinlichkeitsrechnung und Statistik in der Biologie

Blatt 7

Aufgabe 1 Entgegen der landläufigen Meinung gibt es nicht nur grüne Marsmenschen, sondern auch rote und blaue. Grüne und blaue Marsmenschen scheinen sich zudem in ihrer Körpergröße zu unterscheiden. Belegen Sie dies mit einem statistischen Test; eine Vermutung kann man schließlich nicht publizieren! Der Datensatz `mars.txt` wurde bereits mit folgenden Ergebnissen untersucht:

	grün	blau	rot
Anzahl	20	12	10
Mittelwert	60.87	71.35	62.31
Standardfehler	1.61	2.60	1.95

Bearbeiten Sie die folgenden Aufgabenteile ohne Zugriff auf den vollen Datensatz, also nur mit den in obiger Tabelle gegebenen Werten. (Aber R dürfen Sie verwenden.)

(a) Bestimmen Sie (ohne `R`, mit Taschenrechner) die 95%-Konfidenzintervalle der Schätzungen der durchschnittlichen Körpergrößen grüner, blauer und roter Marsbewohner. Die Quantil-Tafeln zur t-Verteilung auf der Vorlesungs-Homepage liefert die nötigen Quantile.

(b) Ist der Stichprobenmittelwert der blauen Marsianer auf dem 5%-Niveau signifikant von 65 verschieden? (Tipp: Das kann alleine mit den 95%-Konfidenzintervallen beantwortet werden.)

(c) Sie glauben, dass sich die mittleren Körpergrößen von grünen und blauen Marsbewohnern unterscheiden. Mit welchem Test können Sie dies mit Hilfe der vorliegenden Daten untermauern? Stellen Sie zunächst die Nullhypothese auf. Benutzen Sie dann einen geeigneten `R`-Befehl, um den p-Wert zu berechen.

Aufgabe 2 Kuss et al. ("The fouled player should not take the penalty himself": An empirical investigation of an old German football myth, *J. Sports Sciences* **25**, no. 9, 963–967, 2007) berichten über die Strafelfmeter in der 1. Fußballbundesliga (der Herren) von August 1993 bis Februar 1995:

	verwandelt	nicht verwandelt
der Gefoulte schießt selbst	74	28
anderer Spieler schießt	547	186

Stützen diese Daten die These, dass der Gefoulte den Elfmeter nicht selbst schießen sollte?

a) Formulieren Sie zunächst die Nullhypothese.

b) Führen Sie dann einen geeigneten Test „per Hand" durch, d.h. ohne Verwendung eines `R`-Befehl, in dem „test" vorkommt. Verwenden Sie dazu die Quantil-Tafeln von der Vorlesungs-Homepage.

c) Führen Sie dann einen geeigneten Test mit `R` durch.

Aufgabe 3 Die Ameise *Pseudomyrmex ferruginea* lebt in enger Gemeinschaft mit der Akazie *Acacia collinsii*. In den großen hohlen Dornen der Akazie findet sie Unterschlupf, von speziellen Drüsen der Pflanze wird sie ernährt. Dafür verteidigt sie die Akazie gegen Insekten, Tiere und andere Pflanzen. Um die Wirksamkeit dieser Verteidigung zu prüfen, wählte man unter 50 vergleichbaren Akazien rein zufällig 20 aus und entfernte die Ameisen. Nach 10 Monaten hatten neun der Pflanzen ohne Ameisen überlebt und 21 der Pflanzen mit Ameisen.
Schätzen Sie die Überlebenswahrscheinlichkeit θ_{mit} einer Pflanze mit Ameisen und die Standardabweichung Ihres Schätzers $\widehat{\theta}_{\text{mit}}$. Berechnen Sie ohne R das 95%-Wald-Konfidenzintervall für θ_{mit} und anschließend mit R das 95%-Wilson-Konfidenzintervall. Würden Sie anhand der Beobachtungen sagen, dass θ_{mit} signifkant von $0{,}5$ verschieden ist? Schätzen Sie auch θ_{ohne}, die Überlebenswahrscheinlichkeit einer Pflanze ohne Ameisen. Kann man mit Hilfe der vorher berechneten Konfidenzintervalle entscheiden, ob θ_{ohne} signifikant verschieden ist von θ_{mit}?

Aufgabe 4 In einer Stichprobe von 213 Monarchfaltern, die Anfang August gefangen wurden, befanden sich 125 Männchen. Ende August wurde eine weitere Stichprobe von 87 Monarchfaltern erhoben, darunter waren 62 Männchen. Schätzen Sie den Anteil Männchen in der Population Anfang und Ende August und berechnen Sie ohne R jeweils das 95%-Wald-Konfidenzintervall für den Männchenanteil. Berechnen Sie anschließend mit R die 95%-Wilson-Konfidenzintervalle.

Aufgabe 5 Ein HIV-Test habe eine Sensitivität von 99,9% (d.h. er schlägt bei 99,9% der Infizierten an) und eine Spezifität von 95% (d.h. er schlägt bei 5% der nicht-Infizierten an). In einer Population seien 0,2 % infiziert. Wenn eine zufällig aus der Population ausgewählte Person positiv getestet wird, wie wahrscheinlich ist es dann, dass sie tatsächlich infiziert ist?

Aufgabe 6 Auf einer Alpen-Exkursion entdecken Sie eine mir bisher unbekannte Eidechsenart. Abends auf der Berghütte beschließen Sie, am nächsten Tag Eidechsen zu suchen und den Männchenanteil in der lokalen Population zu bestimmen. (Sie beraten also bereits vor der Probenahme über die statistische Auswertung der Daten, da Sie ja wissen, dass diese Vorgehensweise für jede seriöse Untersuchung essentiell ist!) Sie selbst wollen die Daten frequentistisch analysieren, aber zwei Mitglieder Ihres Teams, Anton und Berta, verfolgen einen Bayesschen Ansatz und erläutern ihre jeweilige A-priori-Verteilung (kurz "Prior"). Am nächsten Tag finden Sie $k = 40$ männliche und $n - k = 60$ weibliche Eidechsen. Sie gehen davon aus, dass es sich um eine repräsentative Stichprobe aus der lokalen Population handelt.

(a) Berechnen Sie den Maximum-Likelihood-Schätzer und das Wald-Konfidenzintervall für den Männchenanteil in der lokalen Population.

(b) Berechnen Sie mit R auch das Wilson-Konfidenzintervall.

(c) Berechnen Sie für die möglichen Männchenanteile 0.3, 0.4, 0.5 und 0.6 jeweils die Likelihood.

(d) Anton hatte am Vorabend vorgeschlagen, der Einfachheit halber nur die gerundeten Werte 0.3, 0.4, 0.5 und 0.6 als mögliche Männchenanteile zu berücksichtigen. Da seiner Meinung nach ein Wert von 0.5 am wahrscheinlichsten ist, bei Abweichungen davon aber eher ein höherer Weibchenanteil biologisch plausibel ist, waren seine A-Priori-Wahrscheinlichkeiten $p_{0.5} = 0.8$, $p_{0.4} = 0.1$ und $p_{0.3} = p_{0.6} = 0.05$. Auf welche A-Posterior-Verteilung (kurz "Posterior") für den Männchenanteil in der lokalen Population kommt man mit diesem Prior und wie groß ist der Erwartungswert des Posteriors?

(e) Bertas Prior wird gut durch eine Beta(3,7)-Verteilung beschrieben. Welcher Posterior ergibt sich für diesen Prior aus den Daten und welchen Erwartungswert hat diese Verteilung?

(f) Stellen Sie Bertas Prior, die Likelihood-Funktion des Männchenanteils sowie Bertas Posterior grafisch dar. (Tipp: hier könnten die R-Befehle `dbeta` und `dbinom` nützlich sein.)

(g) Berechnen Sie ein 95%-Kredibilitätsintervall für den Männchenanteil in der Population gemäß Bertas Posterior.

(h) Wie wahrscheinlich ist es jeweils nach Bertas Prior und nach Bertas Posterior, dass der auf eine Nachkommastelle gerundete Männchenanteil in der lokalen Population 0.3, 0.4, 0.5 oder 0.6 beträgt? (Tipp: verwenden Sie den R-Befehl `pbeta` und berücksichtigen Sie Bereich der Werte, die jeweils auf den gerundeten Wert führen.)