

ÜBUNGEN ZUR WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK FÜR BIOLOGEN

Blatt 7

1. Aufgabe 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 in einer vorherigen Aufgabe 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

- 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.
- Ist der Stichprobenmittelwert der blauen Marsianer auf dem 95%-Niveau signifikant von 65 verschieden? Dies kann alleine mit den 95%-Konfidenzintervallen beantwortet werden.
- 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 berechnen.

2. Aufgabe 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 Strafbüchse 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?

- Formulieren Sie zunächst die Nullhypothese.
- 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.
- Führen Sie dann einen geeigneten Test mit R durch.

3. Aufgabe Die Ameise *Pseudomyrmex ferruginea* lebt in enger Gemeinschaft mit der Akazie *Acacia colinsii*. 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 $\hat{\theta}_{\text{mit}}$. 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} signifikant 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} ?

4. Aufgabe 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.

5. Aufgabe Herr Metzler verteilt seine 69 Biologie- und Statistikbücher auf vier Regalbretter links-oben, rechts-oben, unten-links und unten-rechts. Die Anzahlen der Bücher auf den Regalbrettern sind A, B, C, D :

	links	rechts	
oben	A	B	A+B=45
unten	C	D	C+D=24
	X=A+C	Y=B+D	X+Y=69

Auf den oberen Regalbrettern zusammen stehen Biologiebücher und auf den beiden unteren zusammen 24 Statistikbücher. Diese Ordnung – oben Biologie und unten Statistik – soll erhalten bleiben, auch in den folgenden Aufgabenteilen. Ob ein Buch links oder rechts steht, ist rein zufällig.

- Wieviele verschiedene Möglichkeiten hat Herr Metzler, zwei Biologiebücher und sieben Statistikbücher auszuwählen, um sie ins rechte Regal zu stellen? (Tipp: der R-Befehl für $\binom{n}{k}$ ist `choose(n, k)`)
- Wieviele Möglichkeiten hat Herr Metzler insgesamt, neun Bücher für die rechte Regalseite auszuwählen?
- Wenn Herr Metzler 9 Bücher rein zufällig auswählt und ins rechte Regal stellt – unter Beachtung der Regel, dass Biologiebücher oben und Statistikbücher unten stehen – wie groß ist dann die Wahrscheinlichkeit, dass $B = 2$ herauskommt?
- Welche möglichen Werte für B sind unter den Voraussetzungen von Aufgabenteil (c) die fünf unwahrscheinlichsten? Berechnen Sie deren Gesamtwahrscheinlichkeit.
- Was hat diese Aufgabe nun mit einem Test aus der Vorlesung zu tun?

6. Aufgabe 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?