

## 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 von ihnen die Ameisen. Nach 10 Monaten hatten 9 der Pflanzen ohne Ameisen überlebt und 21 der Pflanzen mit Ameisen.

Schätzen Sie  $\theta_{\text{mit}}$ , die Überlebenswahrscheinlichkeit einer Pflanze mit Ameisen, und die Standardabweichung Ihres Schätzers  $\hat{\theta}_{\text{mit}}$ . Geben Sie das 95%-Wald-Konfidenzintervall für  $\theta_{\text{mit}}$  an. Würden Sie anhand der Beobachtungen sagen, dass  $\theta_{\text{mit}}$  signifikant von 0,5 verschieden ist? Schätzen Sie auch  $\theta_{\text{ohne}}$ , die Überlebenswahrscheinlichkeit einer Pflanze ohne Ameisen. Ist  $\theta_{\text{ohne}}$  signifikant verschieden von  $\theta_{\text{mit}}$ ? Beantworten Sie diese Frage mit Hilfe des vorher berechneten Konfidenzintervalls.

**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 geben Sie jeweils das 95%-Wald-Konfidenzintervall für den Männchenanteil an. Würden Sie sagen, dass sich der Männchenanteil im Laufe des Monats signifikant geändert hat?

**5. Aufgabe** Herr Metzler verteilt seine 69 Statistik-Bücher 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 45 Bücher über den  $t$ -Test und auf den beiden unteren zusammen 24 Bücher über den  $\chi^2$ -Test. Diese Themengebiete dürfen keinesfalls vermischt werden, auch nicht in den folgenden Aufgabenteilen. Ob ein Buch links oder rechts steht, ist ziemlich zufällig.

- Wieviele verschiedene Möglichkeiten hat Herr Metzler, 2 Bücher über den  $t$ -Test und 7 Bücher über den  $\chi^2$ -Test 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 obiger Nichtvermischungsregel), 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.

### 6. Aufgabe

In einem Bachsystem wurde an 70 zufällig ausgewählten Probennamestellen die Korngröße des Bodensediments gemessen:  $23 \times$  fein,  $19 \times$  mittel,  $28 \times$  grob. Die nebenstehende Tabelle zeigt, an wievielen der Stellen mit den verschiedenen Korngrößen vier verschiedene Arten (A, B, C, D) von Zuckmückenlarven vorkamen.

	A	B	C	D
fein	3	12	10	19
mittel	4	9	16	8
grob	7	15	18	9

- (a) Lässt sich mit den Daten die Vermutung belegen, dass die vier Arten in unterschiedlicher Weise an die Korngrößen angepasst sind, oder zeigen die Daten, dass alle vier Arten gleich auf die Korngrößen reagieren?
- (b) Kann man Aufgabenteil (a) durch einen Chi-Quadrat-Test beantworten und, wenn ja, mit wievielen Freiheitsgraden?
- (c) Berechnen Sie unter der Nullhypothese  $H_0$ , dass es unter den vier Arten keine Unterschiede bezüglich der Korngrößen gibt, die Wahrscheinlichkeiten dass an einer zufällig ausgewählten neuen Probenahmestelle *mittlere Korngröße und Art A* bzw. *grobes Korn und Art C* beobachtet werden. Schätzen Sie dabei die Häufigkeiten von “mittel (m)”, “grob (g)”, “A” und “C” aus den Zeilen- bzw. Spaltensummen der obigen Tabelle. Berechnen Sie dann die entsprechenden Erwartungswerte  $E_{m,A}$  und  $E_{g,C}$  der zugehörigen Tabelleneinträge.
- (d) Wir gehen nun von der etwas vereinfachten Nullhypothese aus, dass die  $O_{x,Y}$  zustande gekommen sind, indem  $n = 130$  Beobachtungen der Art “Art Y kommt an dieser Stelle mit Körnung  $x$  vor” stochastisch unabhängig jeweils mit den in (c) ermittelten Wahrscheinlichkeiten gemacht wurden. (Im Vergleich zur Nullhypothese des Chi-Quadrat-Test vernachlässigen also der Einfachheit halber die gegebenen Zeilen- und Spaltensummen.) Berechnen Sie für diese Annahme die Erwartungswerte von  $(O_{m,A} - E_{m,A})^2 / E_{m,A}$  und  $(O_{g,C} - E_{g,C})^2 / E_{g,C}$ .

Hinweis: Diese Aufgabe soll zeigen, welche Probleme bei Daten auftreten können.