

ÜBUNGEN ZUR WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK FÜR BIOLOGEN

Blatt 8

Aufgabe 1 (simulierte Daten, freundlicherweise von Prof. S. Diehl zur Verfügung gestellt)

Es könnte sein, dass Dickhorn-Mutterschafe mehr (oder weniger) Nahrung benötigen, wenn sie Lämmer säugen. Um dies zu untersuchen, wurde bei 16 weiblichen Schafen jeweils die mittlere Zeit (in Minuten pro Tag) bestimmt, die sie mit Gras zu bringen, einerseits während der Zeit des Säugens, andererseits in einem Referenzzeitraum ohne Lämmer. Es gab folgendes Ergebnis:

Schaf Nr.	12	17	23	24	29	31	33	48	55	57	59	60	63	71	73	78
Zeit mit Lamm	272	262	256	260	252	261	253	252	256	260	249	243	254	270	247	250
Zeit ohne Lamm	256	244	267	268	258	262	245	257	273	277	264	254	256	291	264	271

Gras Mutterschafe mit Lämmern signifikant anders als ohne? Geben Sie auch ein Konfidenzintervall (beispielsweise zum Niveau 5%) für die Differenz der mittleren Futterzeiten an und formulieren Sie einen Antwortsatz.

Aufgabe 2 Sie besuchen eine fremde Stadt, von der Sie wissen, dass die Taxis dort fortlaufend mit $1, 2, \dots, N$ nummeriert sind, Sie kennen aber nicht die Gesamtzahl N . Während Sie vor dem Bahnhof auf den Bus warten, sehen Sie 12 Taxis vorbeifahren. Sie haben die Nummern 455, 190, 39, 542, 370, 289, 274, 237, 374, 899, 741 und 706.

- Wie groß ist die Wahrscheinlichkeit, genau diese Taxinummern in dieser Reihenfolge zu sehen, wenn $N = 912$ ist? (Modellieren Sie das Erblicken eines Taxis durch "Ziehen mit Zurücklegen", d.h. ein Taxi kann prinzipiell auch zweimal gesehen werden.)
- Wie groß ist die Wahrscheinlichkeit, genau diese Taxinummern in dieser Reihenfolge zu sehen für beliebiges N ? (Insbesondere: Was passiert mit der Wahrscheinlichkeit der Beobachtung für $N < 899$?)
- Berechnen Sie auf der Basis Ihrer Beobachtung den Maximum-Likelihood-Schätzer für N .
- Wie sähe der ML-Schätzer im allgemeinen Fall aus: Sie sehen n Taxis mit Nummern x_1, x_2, \dots, x_n ?

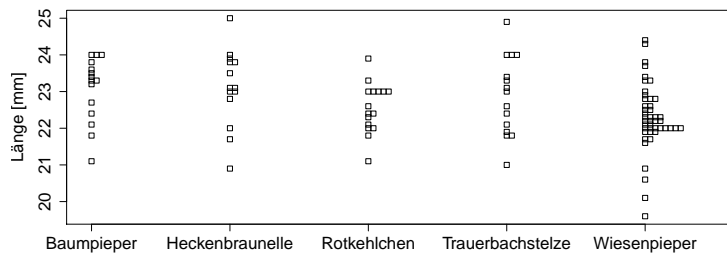
Aufgabe 3 Auf einem Abschnitt nicht-kodierender DNS der Länge 1000 Basenpaare werden zwischen Mensch und Schimpanse $N = 23$ Mutationen gezählt. Ein Wissenschaftler argumentiert: „Die Anzahl Mutationen auf einem Stück der Länge 1000 ist Binomial-verteilt mit Erwartungswert $np \approx n\hat{p} = 23$ und Standardabweichung $\sqrt{np(1-p)} \approx \sqrt{n\hat{p}(1-\hat{p})} \approx \sqrt{n\hat{p}} = \sqrt{23}$. Mit der asymptotischen Normalität folgt, dass $[13.6, 32.4]$ ein 95%-Konfidenzintervall für die erwartete Anzahl Unterschiede auf einem Stück DNS der Länge 1000 aus derselben genomischen Region ist.“

- Erklären Sie, wie dieses Intervall berechnet wurde.
- Erzeugen Sie unter der Annahme, dass das wahre $p = 0.023$ ist, zufällige Anzahlen von Mutationen und berechnen Sie aus den simulierten Daten Konfidenzintervalle nach dieser Methode (Hinweis: der R-Befehl `rbinom` simuliert binomialverteilte Zufallsgrößen). Mit welcher Wahrscheinlichkeit enthält das so konstruierte Konfidenzintervall den wahren Wert?
- Führen Sie Teil b) auch für $p = 0.007$ durch.

Aufgabe 4 Die Nester der winzigen Ameisen der Art *Temnothorax longispinosus* findet man in hohlen Eicheln. In einem Waldstück werden einige Nester gefunden und untersucht. An Stellen, wo der Boden etwas sandiger war, wurden drei Nester mit 16, 22 und 24 Arbeiterinnen gefunden. Auf anderem Boden, der als trocken klassifiziert wurde, wurden ebenfalls drei Nester gefunden. Sie enthielten 20, 28 und 32 Arbeiterinnen. Vier Nester mit 28, 34, 36 und 40 Arbeiterinnen wurden auf etwas feuchterem Boden gefunden.

- Visualisieren Sie die Daten in angemessener Weise.
- Führen Sie ohne Benutzung von R eine Varianzanalyse durch, um zu überprüfen, ob die Daten dafür sprechen, dass der Bodentyp einen signifikanten Einfluss auf die Nestgröße hat.
- Überprüfen Sie Ihre Rechnung, indem Sie die Varianzanalyse mit R durchführen.

Aufgabe 5 Die Datei `Kuckuckseier.txt` enthält die Längen von Kuckuckseiern, die in Nestern verschiedener Wirtsspezies gefunden wurden (ein Teil des Datensatzes aus O.H. Latter, *The Egg of Cuculus Canorus ...*, *Biometrika* **1**, 164–176 (1902)). Führen Sie mit diesen Daten eine Varianzanalyse durch und prüfen Sie die Hypothese, dass die Länge nicht von der Wirtsspezies abhängt. Sie können von Hand rechnen oder R benutzen, geben Sie die Werte der Quadratsummen und der F -Statistik an. Wie ist F unter der Hypothese verteilt, dass die mittlere Länge bei allen Wirtsspezies gleich ist?



Aufgabe 6 Eine Nutzpflanze wurde auf einer Versuchsfläche angebaut, die in 50 gleich große Abschnitte unterteilt war. Die Abschnitte wurden mit fünf verschiedenen Düngemitteln behandelt, je 10 Abschnitte mit dem selben Dünger. Die Erträge der einzelnen Abschnitte und die eingesetzten Düngemittel finden Sie in der Datei `duenger.txt`.

- Visualisieren Sie die Daten in angemessener Weise.
- Untersuchen Sie mittels einer Varianzanalyse, ob die Wahl des Düngemittels einen signifikanten Einfluss auf den Ertrag hat.
- Zwischen welchen Düngemitteln gibt es signifikante Unterschiede? Stellen Sie paarweise Vergleiche an.
- Verwenden Sie mindestens drei verschiedene Möglichkeiten, bei (b) die p -Werte in Hinblick auf das multiple Testen zu korrigieren, und Vergleichen Sie die Ergebnisse.