

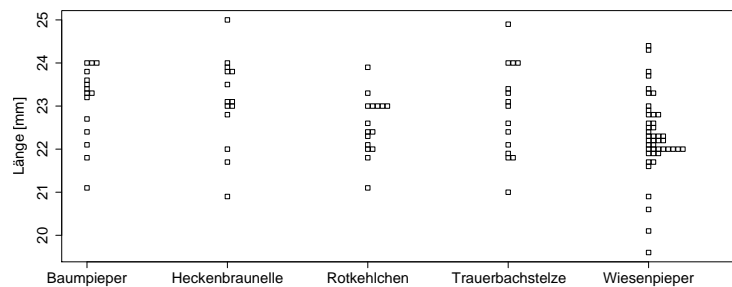
## ÜBUNGEN ZUR WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK FÜR BIOLOGEN

## Blatt 9

**Aufgabe 1** 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 2** 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 3** 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.

**Aufgabe 4** Aus drei verschiedenen Seen wurden jeweils fünf Welse gefangen. Es soll untersucht werden, ob es zwischen den Seen Unterschiede in Hinblick auf die Größe der Welse gibt. Die Datei `welse.txt` enthält die Gewichte der Welse in kg.

- (a) Visualisieren Sie die Daten in angemessener Weise.
- (b) Führen Sie Schritt für Schritt, also ohne Verwendung des R-Befehls `kruskal.test()`, einen Kruskal-Wallis-Test durch, um zu untersuchen, ob sich die Größenverteilungen der Welse in den Seen unterscheiden.
- (c) Überprüfen Sie Ihr Ergebnis mit dem R-Befehl `kruskal.test()`.
- (d) Führen Sie nun auch eine Varianzanalyse durch und vergleichen Sie die Ergebnisse der beiden Tests. Wie lassen sich die Unterschiede erklären?
- (e) Welches Fazit ziehen Sie in Hinblick auf die Welse? Formulieren Sie Ihr Ergebnis in einem Satz.

**Aufgabe 5** In einem Teich befindet sich eine unbekannte Anzahl  $N$  von Karpfen. Um  $N$  zu schätzen, wurden 50 Karpfen gefangen, markiert und wieder zurückgesetzt. Eine Woche später wurden 60 Karpfen gefangen, darunter waren 13 markierte.

**a)** Wie wahrscheinlich ist es, beim Fang von 60 Karpfen aus  $N - 50$  nicht-markierten und 50 markierten Fischen genau 13 markierte zu finden (wir nehmen dabei an, dass die markierten und die unmarkierten Fische sich wieder gut durchmischt haben)?

**b)** Schätzen Sie die Anzahl Karpfen  $N$  im Teich anhand der Informationen aus der Aufgabenstellung (bestimmen Sie beispielsweise den Maximum-Likelihood-Schätzer).

**Aufgabe 6** (simulierte Daten, freundlicherweise von Prof. S. Diehl zur Verfügung gestellt)

Man vermutet, dass es genetische Variabilität der Blattgröße in drei Spitzwegerich-Populationen gibt. Um dies zu prüfen, wurden aus jeder der drei Populationen zufällig Samen gesammelt. 30 Pflanzen, jeweils 10 Pflanzen pro Populationen, wurden gemeinsam im Gewächshaus gezogen. Nach 90 Tagen wurde die Blattfläche der Pflanzen in  $\text{mm}^2$  (als Mittelwert der jeweils drei größten Blätter einer Pflanze) bestimmt, mit folgendem Ergebnis:

Population A: 255 345 312 517 377 458 984 442 487 742

Population B: 344 389 517 598 734 1099 835 465 387 685

Population C: 649 756 484 527 534 388 734 645 865 1077

Stützen diese Daten obige Vermutung?