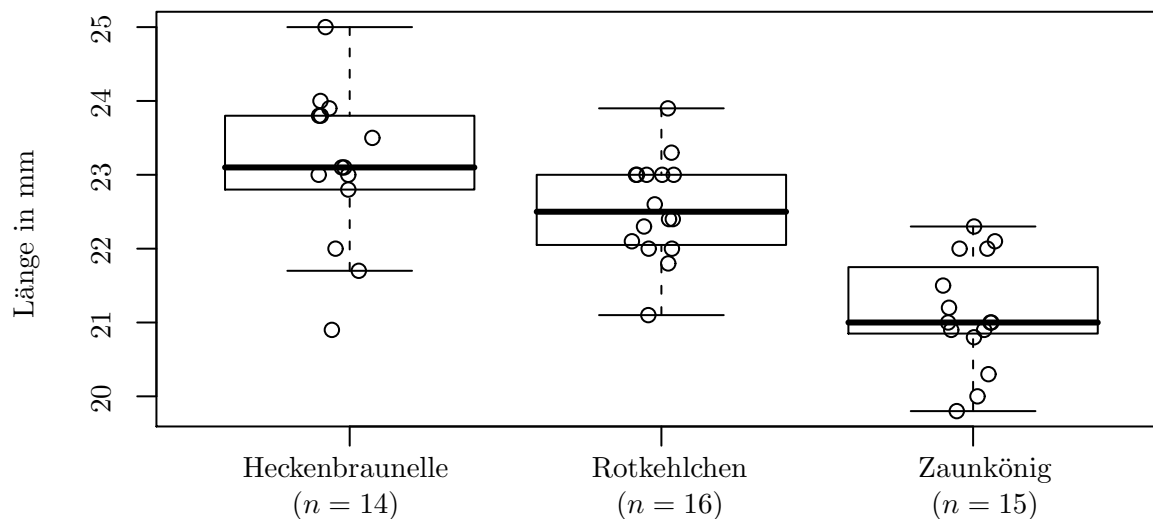


**1. Aufgabe** Die folgenden Boxplots zeigen die Länge von Kuckuckseiern, die in Nestern von drei verschiedenen Wirtsspezies gefunden wurden:



Geben Sie für jede der drei Wirtsspezies jeweils den Median, das 1. und 3. Quartil der beobachteten Längen sowie den kleinsten und den größten Wert an (mit einer plausiblen Genauigkeit).

(Daten aus O.H. Latter, The egg of *cuculus canorus*. An inquiry into the dimensions of the cuckoo's egg and the relation of the variations to the size of the eggs of the foster-parent, with notes on coloration, &c, *Biometrika* **1**, 164–176 (1902), zitiert nach dem R-Paket DAAG)

**2. Aufgabe** In einem Experiment (in dem es um verschiedenen Futtersorten ging) wurden neun Küken mit einem speziellen Futter aufgezogen und nach 21 Tagen gewogen. Dabei ergaben sich folgende Messwerte (in g):

204 281 200 196 238 205 322 237 264

Bestimmen Sie *von Hand* Median, Mittelwert und Standardabweichung der Beobachtungen (Sie sollten die Rechenschritte erklären können, für Grundrechenarten und Wurzeln dürfen Sie einen Taschenrechner benutzen).

(Die Daten sind Teil des R-Beispieldatensatzes `ChickWeight` (die Gruppe `Diet==4`), siehe z.B. M. Crowder, D. Hand, *Analysis of Repeated Measures*, Chapman and Hall (1990).)

**3. Aufgabe** Untersuchen Sie folgende Fragen, indem Sie sich kleine Beispiel-Datensätze ausdenken:

a) Unter welchen Umständen können Mittelwert und Median sehr unterschiedliche Werte annehmen?

b) Für eine Adler-Population sei  $x$  die Flügelänge und  $y$  das Gewicht. Kann es sein, dass bei einem bestimmten Adler aus einer Stichprobe sowohl das Verhältnis  $y/x$  von Gewicht zu Flügelänge als auch das Verhältnis  $x/y$  von Flügelänge zu Gewicht einen unterdurchschnittlichen Wert annimmt (wobei hier „unterdurchschnittlich“ verstanden werden soll als „ein Wert, der kleiner ist als der Mittelwert“)? Tipp: Untersuchen Sie einige extreme Beispiele mit drei Adlern.

c) Kann es sein, dass bei einem Adler sowohl  $x/y$  als auch  $y/x$  kleiner ist als der jeweilige Median?

**4. Aufgabe** Die Datei `schnupfen.dat` enthält die Inkubationszeit (in Stunden) für 1000 „simulierte Schnupfen“ (die Verteilung ist dem Artikel von P.E. Sartwell, *The distribution of incubation periods of infectious disease, Am. J. Hyg.* **51**, 310–318 (1950) entnommen). Speichern Sie die Datei in ein Verzeichnis ihrer Wahl, z.B. 'Pfad/Daten'. Starten Sie dann R. In der R Konsole muss zunächst das Verzeichnis durch

```
setwd("Pfad/Daten")
```

geändert werden. Der Befehl `'getwd()'` sollte nun das richtige Verzeichnis ausgeben. Laden Sie dann den Datensatz mit dem Befehl

```
schnupfen <- read.table("schnupfen.dat",header=TRUE).
```

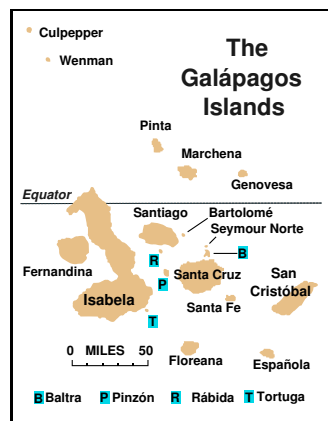
Ein Beispiel für das Einlesen von Daten finden Sie auf der Homepage in der Datei `'R_intro.R'` unter Abschnitt 10. Zum Umgang mit einer Datenstruktur wie `schnupfen`, siehe Abschnitt 9. Bitte beachten Sie, dass sie einige der im Folgenden genannten R-Befehle nicht direkt auf `schnupfen` anwenden müssen, sondern auf `schnupfen$zeit` oder `schnupfen[[1]]`.

- Berechnen Sie mit R Mittelwert und Standardabweichung für die 1000 Werte aus `schnupfen.dat`.
- Nehmen Sie aus diesen Beobachtungen eine zufällige Stichprobe der Größe 9 und berechnen Sie Mittelwert und Standardabweichung dieser Stichprobe. Stichproben können Sie mit dem R-Befehl `sample()` ziehen. Genaueres dazu finden Sie in Abschnitt 7 in `'R_intro.R'`.
- In der Übung: Vergleichen Sie Ihr Ergebnis mit dem der anderen Teilnehmer der Übungsgruppe. Erstellen Sie in der Übungsgruppe gemeinsam eine Tabelle der beobachteten Mittelwerte aller Teilnehmer, und bestimmen Sie Mittelwert und Standardabweichung dieser Werte.

**5. Aufgabe** a) Stellen Sie die Werte aus `schnupfen.dat` aus Aufgabe 4 als Histogramm dar, tragen Sie Mittelwert und Mittelwert plus/minus Standardabweichung ein. Hinweis: Abschnitt 11 in `'R_intro.R'`.

b) Wiederholen Sie Teil a) mit den Logarithmen der Werte aus `schnupfen.dat`. Wenden Sie dazu den Befehl `'log()'` an auf `'schnupfen$zeit'`. Was für Eigenschaften hat das Histogramm der logarithmierten Daten (im Gegensatz zu den nicht-logarithmierten Daten)?

### 6. Aufgabe



Die Datei `swarth1.dat` enthält die Schnabellängen von Exemplaren von Darwin-Finken (der Art *Geospiza fortis*) von drei verschiedenen Inseln des Galápagos-Archipels (Floreana, San Cristóbal, Santa Cruz). Erzeugen Sie mit R Histogramme und Boxplots der Schnabellänge

- für die gesamte Stichprobe,
- für jede der Inseln getrennt.
- Erstellen Sie eine Grafik, die Dichtepolygone für alle Inseln gemeinsam zeigt.
- Fassen Sie Ihre Beobachtungen in einem Satz (oder zwei) zusammen.

Hinweise: Abschnitte 11 und 12 in `'R_intro.R'`.

(Daten aus H. S. Swarth, *The avifauna of the Galapagos Islands* (1931), zitiert nach <http://datadryad.org/repo/handle/10255/dryad.155>.)