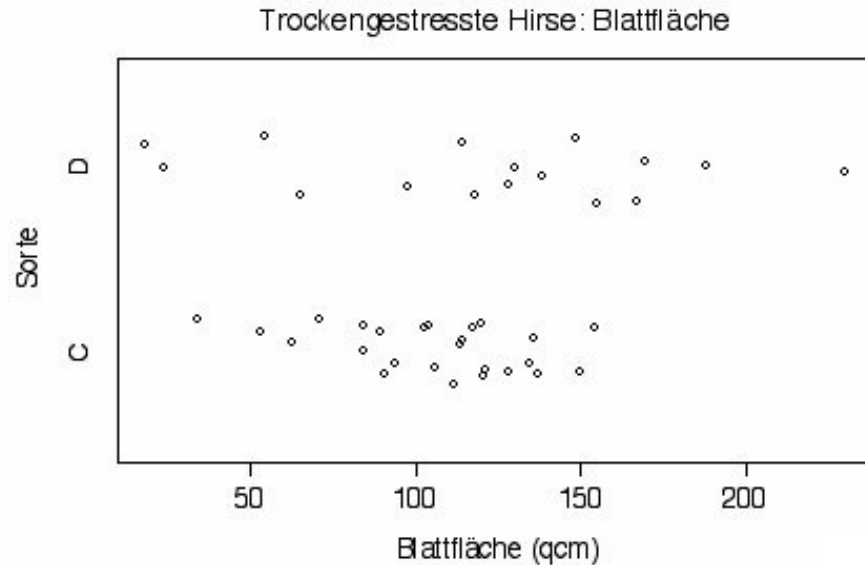


## 1. Aufgabe



- (b) Überprüfen Sie Ihre Ergebnisse, indem Sie die selben Werte nun mit R berechnen.
- (c) Legen diese Werte nahe, dass das mittlere Gewicht erwachsener Gnus in dieser Herde von dem bei Wikipedia angegebenen Wert abweicht?

**3. Aufgabe** Betrachten Sie nochmals die Schnabellängen der Darwin-Finken von den drei Inseln des Galápagos-Archipels aus dem Datensatz von H.S. Swarth (R Intro, Abschnitt 10, die Datei `swarth1.dat`): Berechnen Sie mit R getrennt nach den Inseln jeweils den Stichprobenmittelwert und den dazugehörigen Standardfehler. Würden Sie sagen, dass die mittlere Schnabellänge auf den Inseln unterschiedlich ist?

**4. Aufgabe** In einem Experiment wurden zwei Gruppen von zehn bzw. neun Küken mit zwei verschiedenen Futtersorten aufgezogen und nach 21 Tagen gewogen. Die Ergebnisse waren (in g):

Mit Sorte A

331 167 175 74 265 251 192 233 309 150

Mit Sorte B

204 281 200 196 238 205 322 237 264

Bestimmen Sie für jede der Stichproben den Mittelwert und den Standardfehler (geben Sie einen plausibel gerundeten Wert an). Angenommen, Sie sollten ein Experiment entwerfen, in dem das mittlere Gewicht nach 21 Tagen mit Futtersorte A bis auf einen Standardfehler von ungefähr 5g genau bestimmt werden kann. Wieviele Küken müsste man dafür etwa nehmen? Wieviele bräuchte man vermutlich, um dasselbe Problem für Sorte B zu lösen?

(Die Daten sind Teil des R-Beispieldatensatzes `ChickWeight` (die Gruppen `Diet==4` und `Diet==2`), siehe z.B. M. Crowder, D. Hand, *Analysis of Repeated Measures*, Chapman and Hall (1990).)

**5. Aufgabe** Ein Programmierer arbeitet an einem neuen Programm, `FAIREMÜNZE`, das faire Münzwürfe simulieren soll. Um das Programm zu testen, läßt er es eine Million mal laufen und läßt die Anzahl der Erfolge mitprotokollieren. Das Ergebnis ist 502015. Angesichts dieses Resultats denkt der Programmierer: „*Hm, um zweitausend und fünfzehn daneben. Ganz schön viel. Andererseits, vergleiche die Abweichung mit der Million: Zweitausend – ich vernachlässige 'mal die fünfzehn – zu einer Million ist zwei zu tausend oder eins zu fünfhundert, also 0.2 Prozent. Das ist ziemlich klein. Sehr gut, also besteht FAIREMÜNZE den Test.*“ Stimmen Sie dieser Argumentation zu? Begründen Sie Ihre Antwort.

**6. Aufgabe** Das R-Skript `samsim.R` enthält Befehle, um folgendes auszuführen:

- (a) Simuliere eine Population  $x$  von 1000000 Werten gemäß einer gewissen Verteilung.
- (b) Berechne Mittelwert  $\mu$  und Standardabweichung  $\sigma$  der Gesamtpopulation.
- (c) Ziehe 1000 Stichproben der Größe  $n = 10$  aus der Gesamtpopulation.
- (d) Berechne Stichprobenmittelwert  $\bar{x}$  und (Stichproben-)Standardabweichung  $s$  für jede der 1000 Stichproben.
- (e) Zähle, für welchen Anteil der Stichproben das Intervall  $[\bar{x} - s/\sqrt{n}, \bar{x} + s/\sqrt{n}]$  den Populationsmittelwert  $\mu$  überdeckt.
- (f) Vergleiche die Verteilung der Stichprobenmittel grafisch mit der Normalverteilung mit Mittelwert  $\mu$  und Standardabweichung  $\sigma/\sqrt{n}$ .

Öffnen Sie `samsim.R` als Skript in R und finden Sie heraus, welche Befehle zu den Schritten (a) bis (f) gehören. Am Ende des Skripts wird eine weitere Population  $y$  simuliert. Führen Sie obige Schritte für  $y$  mit verschiedenen Wahlen der Stichprobengröße  $n$  durch. Für welche  $n$  ist die Normalverteilung eine gute Approximation der Verteilung des Stichprobenmittels?