

1. Aufgabe Entgegen der landläufigen Meinung gibt es nicht nur grüne Marsmenschen, sondern auch rote und blaue. In der Tabelle `mars.txt` finden Sie die Größen (in cm) und Färbungen von allen 42 Marsbewohnern, die in den letzten 50 Jahren in verschiedenen havarierten Raumschiffen gefunden wurden. Wir gehen einmal davon aus, dass es sich dabei bezüglich der Größen um repräsentative Stichproben erwachsener Marsbewohner der verschiedenen Färbungen handelt.

1. Visualisieren Sie die Daten als Histogramm, als Boxplot und als Streudiagramm. Der R-Befehl für das Streudiagramm ist hier

```
> stripchart(size~color, method="jitter", pch=16,
             col=c("blue", "green", "red"), ylim=c(0.5, 3.5))
```

Dabei rüttelt „jitter“ die Punkte und `ylim=` setzt den Bereich der y-Achse. Vergleichen Sie die Vor- und Nachteile von Histogramm, Boxplot und Streudiagramm

2. Schätzen Sie aus den Daten die durchschnittlichen Größen roter, blauer und grüner Marsbewohner.
3. Wie genau sind diese Schätzungen? Berechnen Sie die Standardfehler!
4. Visualisieren Sie die Daten zusammen mit den Schätzungen und den Standardfehlern.

2. Aufgabe 10 Testpersonen haben in einem Versuch einen (linken) Schuh mit Sohlenmaterial A und einen (linken) Schuh mit Sohlenmaterial B getragen, nach einiger Zeit wurde die Abnutzung gemessen, mit folgendem Ergebnis:

Testperson	1	2	3	4	5	6	7	8	9	10
Abnutzung A	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8	8.8	13.3
Abnutzung B	14.0	8.8	11.2	14.2	11.8	6.4	9.8	11.3	9.3	13.6

Ist die Abnutzung bei den beiden Materialien signifikant verschieden? R kennt diesen (populären) Datensatz mittels `data(shoes, package='MASS')`. Nach diesem Befehl steht Ihnen der Datensatz als `shoes` zur Verfügung.

1. Führen Sie einen gepaarten t-Test „per Hand“ durch. Hierfür dürfen Sie alle R-Befehle ausser `t.test()` benutzen.
2. In R führt man einen t-Test mit dem Befehl `t.test()` durch. Einen gepaarten t-Test führt man in R entweder als `t.test(x-y)` oder als `t.test(x,y,paired=TRUE)` durch. Führen Sie nun den t-Test mit `t.test()` durch.
3. Formulieren Sie das Ergebnis in einem Satz.

3. Aufgabe Um unser Gefühl für Korrelationen zu verbessern, wollen wir uns Datenwolken zu gegebenen Korrelationen überlegen. Dabei gibt es keine eindeutige Lösungen; verschiedene Punkt- wolken passen zu den Korrelationen. Zeichnen Sie jeweils eine Punkt- wolke mit 8 Punkten.

1. Wie könnte eine Punkt- wolke mit Korrelation 0.9 aussehen?
2. Wie könnte eine Punkt- wolke mit Korrelation -0.4 aussehen?
3. Wie könnte eine Punkt- wolke mit Korrelation -1 aussehen?
4. Wie könnte eine Punkt- wolke mit Korrelation 0.01 aussehen?
5. Wie könnte eine Punkt- wolke mit Korrelation -0.85 aussehen?

Noch ein reales Datenbeispiel: Führen Sie den R-Befehl `> data(trees); attach(trees)` aus. Eine Beschreibung des Datensatzes erhalten Sie mit `> ?trees`. Berechnen Sie die Korrelation zwischen 'Volume' und 'Girth' mit dem R-Befehl `cor()`. Wie könnte die zugehörige Punkt- wolke aussehen? Bestätigen Sie Ihre Vermutung, indem Sie 'Volume' in Abhängigkeit von 'Girth' plotten.

4. Aufgabe Ein leidenschaftlicher Risiko-Spieler Max hat die letzten Spiele verloren, da er mit seinem Lieblingswürfel zu kleine Werte gewürfelt hat. Nun zweifelt er an seinem Würfel und möchte testen, ob wirklich im Mittel jedes sechste Mal die 6 kommt. Dazu würfelt er 120 mal und notiert die jeweilige Augenzahl. Es kam 11 mal die 6 und sonst andere Zahlen. Da nur jedes elfte Mal die 6 kam, fühlt sich Max bestätigt, möchte seine Vermutung nun aber auch statistisch belegen. Bitte helfen Sie ihm dabei. Hinweis: Max möchte zeigen, dass sein Würfel nicht fair ist. Ein Skeptiker würde entgegen, dass die Beobachtung reiner Zufall ist. Wie groß ist unter dieser Nullhypothese die Wahrscheinlichkeit, eine so große oder noch größere Abweichung vom Erwartungswert zu sehen? Sie müssen sich zunächst die Verteilung unter der Nullhypothese überlegen.

5. Aufgabe Es mag überraschen, dass bei der Stichprobenvarianz durch die Stichprobenlänge minus eins geteilt wird. Wir wollen dies nun untersuchen.

In einer früheren Aufgabe wurde die Varianz des Würfels berechnet mit dem Ergebnis $35/12$. Um Vertrauen zur Skalierung '1 / Stichprobenlänge minus eins' zu gewinnen, wollen wir diese Varianz aus Stichproben schätzen.

1. Generieren Sie eine Reihe von 20 Würfelwürfen. Verwenden Sie dazu den R-Befehl `sample()` mit der Option `replace=TRUE`. Schätzen Sie aus dieser Stichprobe die Varianz des Würfel durch die korrigierte Stichprobenvarianz. Die Stichprobenvarianz der Stichprobe sei mit s_1 bezeichnet.
2. Wiederholen Sie diese Schätzung 10000 mal, d.h. sampeln Sie 10000 mal eine Würfelreihe der Länge 20. Speichern Sie die zugehörigen 10000 (korrigierten) Stichprobenvarianzen s_1, \dots, s_{10000} in einen Vektor.
3. Berechnen Sie nun den Mittelwert dieser 10000 (korrigierten) Stichprobenvarianzen. Vergleichen Sie das Resultat mit $35/12$.
4. Wiederholen Sie obige Schritte mit der Skalierung '1 / Stichprobenlänge'. Vergleichen Sie das Resultat ebenfalls mit $35/12$.

6. Aufgabe Erzeugen Sie sich mit den folgenden Befehlen eine zufällige Verteilung V :

```
N <- 100000
V <- rbeta(N,0.5,0.5)*100+rgamma(N,20,0.5)
```

1. Visualisieren Sie die Verteilung in angemessener Weise und zeichnen Sie Mittelwert und Standardabweichung ein.
2. Ziehen Sie 1000 Stichproben, jeweils vom Umfang $n = 20$ aus V und berechnen Sie jeweils den Stichprobenmittelwert.
3. Erstellen Sie ein Dichtepolygon der 1000 Stichprobenmittelwerte. Tipp: versuchen Sie es mit Befehlen der Art:

```
h <- hist(???,probability=TRUE)
plot(h$dist,h$dens,type="l")
```

4. Berechnen Sie für die 1000 Stichprobenmittelwerte den Mittelwert m Standardabweichung s und fügen Sie zum Dichtepolygon die Glockenkurve der Normalverteilung mit Mittelwert m und Standardabweichung s hinzu. Tipp: versuchen Sie es mit Befehlen der Art:

```
x <- seq(from=???,to=???,by=???)
lines(x,dnorm(x,m,s))
```

(Probieren Sie aus, durch welche Zahlenwerte Sie die Fragezeichen ersetzen müssen.)

5. Wiederholen Sie die vorherigen Aufgabenteil auch für die Stichprobengrößen 5 und 50. Was fällt Ihnen auf?