

100 Aufgaben zur kleinen S-PLUS-Kunde

H. P. Wolf

21. April 1999

1 Ziel und Form

Ziel dieser Aufgabensammlung ist es, parallel zu den Abschnitten der kleinen S-PLUS-Kunde Aufgaben anzubieten, an denen sich der Lernende üben kann. Die Überschriften sind entsprechend zu denen der S-PLUS-Kunde gewählt.

2 Hintergrund, Start, Ende — und ein winziges Beispiel

1: Suchen Sie im Internet nach "S-Plus", "GraS", "CRAN" sowie nach den Personen Richard A. Becker, John M. Chambers, Allen R. Wilks.

2: Starten Sie S-PLUS oder – falls Sie den public domain Interpreter R zur Verfügung haben – R.

3: Tippen Sie ein: `cat("Frisch gewagt ist halb gewonnen!\n")` – natürlich nach dem Prompt-Zeichen ">"! Jede Eingabe ist durch Druck auf der RETURN-Taste dem Interpreter zu übergeben. Was ändert sich, wenn die beiden Zeichen `\n` ausgelassen werden? Was passiert, wenn mitten im Text ein zusätzliches Tüddelchen (`"`) auftaucht?

4: Tippen Sie nacheinander ein: `mean(co2)`, `x<-mean(co2)`, `print(x)` – jeweils mit RETURN bestätigen – und begutachten Sie die Folgen. Falls unter R Fehlermeldungen auftauchen, tippen Sie bitte ein: `co2<-rnorm(10)`. Wer die Tastatur gut kennt, kann auch eingeben: `x<-1959+0:383/12` gefolgt von:
`co2<-cbind(1,x,x^2,sin(x*2*pi))%*%c(72709,-74.543,0.019187,2.6975)`

5: Öffnen Sie ein graphisches Fenster. (Hinweis: An unseren UNIX-Rechnern müssen Sie dazu zunächst unter S-PLUS-Befehl `motif()` eingeben. An einem PC mit WINDOWS ist im Interpreter `win.graph()` einzugeben.)

6: Versuchen Sie: `plot(co2)`, `tsplot(co2)`. Haben Sie eine Idee, was die Anweisung `plot(rnorm(100))` zeigt?

7: Welche Wirkung hat `dev.off()`?

8: Was verursacht `q()`?

3 Aktionen mit Hilfe von Zuweisungen, Operatoren und Funktionen

3.1 Daten ablegen

9: Probieren Sie nacheinander `x<-10`, `x`, `x<-20`, `x`, `30->x`, `x`, `x_40`, `x`, `x<-x<-50`, `x`.

3.2 Ausgaben veranlassen

10: Was passiert, wenn Sie nacheinander die beiden Tasten 7 und RETURN drücken? Natürlich haben Sie so etwas schon gemacht, jedoch sollen Sie jetzt über den Prozeß nachdenken.

11: Was ist der Unterschied von `print(x)` und `cat(x)`?

12: Probieren Sie `cat(x,x)` und `cat(x,x,"\n")`.

3.3 Eigene Daten eingeben

13: Tippen Sie ein: `scan(n=1)`, ihre Glückszahl und anschließend RETURN.

14: Tippen Sie ein: `scan(n=2)` und dann zwei Glückszahlen.

15: Tippen Sie ein: `scan()` und dann die sechs Zahlen, die bei der nächsten Lottoziehung gezogen werden, und drücken Sie anschließend noch einmal RETURN.

16: Weisen Sie auf die Variable `x` die Zahlen von 1 bis 10 zu.

17: Tippen Sie ein: `help(scan)` und studieren Sie den Hilfetext.

18: Was bewirkt: `scan(what="",n=1)`?

3.4 Datenobjekte kombinieren

19: Probieren Sie: `x<-1:10` und `print(x)`.

20: Was bewirkt `x_c(1,2,3,4,5,6,7,8,9,10)`?

21: Probiere: `c(x,x,x,x)`.

22: Probiere: `c(10:5,x)`.

3.5 Einfache Berechnungen anstellen

23: Überlegen Sie, was Sie zuletzt mit Ihrem Taschenrechner ausgerechnet haben, und wiederholen Sie diese Rechnung mit S bzw. R.

24: Rechnen Sie 100 DM in EURO um.

25: Rechnen Sie 0 Grad Fahrenheit in Celsius um.¹

26: Probiere: $10:1+2$, $10:(1+2)$, $10:5-2$, $10:7*2$, $10:3^2$, $10:(3^2)$, $(9:3)^2$, $-10:15$.

27: Weisen Sie auf `x` einen Vektor zu und stellen Sie mit diesem ähnliche Experimente an.

3.6 Kompliziertere Wünsche umsetzen

28: Berechnen Sie Maximum, Minimum, Median, Mittelwert, Range, Stichprobenvarianz und Standardabweichung von `co2`.

29: Ermitteln Sie die Summe aller Werte von `co2`.

30: Wie viele Werte sind auf `co2` abgelegt?

3.7 Deskriptive Statistiken berechnen

31: Setzen Sie die Funktion `summary` ein, um einige Statistiken des Datensatzes `co2` zu berechnen. Vergleichen Sie die Ergebnisse mit denen, die sich bei der Berechnung mittels der einfachen Funktionen ergeben.

3.8 Umfangbewahrende Handlungen anstoßen

32: Was liefert `rev(rev(x))` und warum?

33: Was ist der Unterschied zwischen `sort`, `order` und `rank`?

34: Für welche Fragestellung läßt sich in der Statistik `cumsum` verwenden?

3.9 Fehler und Hilfe

35: Studieren Sie den Text, den `help(help)` ausspuckt.

36: Listen Sie die Fehler auf, die Ihnen bisher unterlaufen sind. Versuchen Sie, diese zu klassifizieren. Überlegen Sie, wie Sie sich aus der Patsche geholfen haben.

4 Graphiken

4.1 Graphische High-Level-Routinen

37: Nehmen Sie den Ihnen nun schon bekannten Datensatz `co2` zur Hand und verfüttern Sie ihn an die Funktionen: `barplot`, `boxplot`, `hist`, `plot`, `qqnorm` und `tsplot`.

38: Ergründen Sie mit der Hilfefunktion, was `pairs` und `qqplot` leisten.

¹Celsius = $\frac{\text{Fahrenheit}-32}{9} \cdot 5$

39: Stellen Sie über `help` fest, welche zusätzlichen Parameter die graphischen High-Level-Routinen kennen und experimentieren Sie damit herum. Versuchen Sie im Besonderen, die Achsendarstellungen zu beeinflussen.

4.2 Graphische Low-Level-Routinen

40: Geben Sie ein: `plot(1:10,type="n")`. Ergänzen Sie eine Linie von den Koordinaten (1,1) über (5,2) nach (3,10). Markieren Sie die Punkte (1,1), (5,5) und (10,10) mittels Zentralsymbolen (siehe `points` oder `symbols`). Schreiben Sie als Titel das heutige Datum über das Bild.

41: Lesen die Hilfe von `legend` und `segments`.

42: Schreiben Sie in das bisherige Bild "hallo Welt" an die Stelle (8,8).

5 Umgang mit Datenobjekten

5.1 Teilzugriff auf Vektoren

43: Weisen Sie zu: `x<-10*(1:13)` und wiederholen Sie die im Abschnitt *Teilzugriff auf Vektoren* vorgestellten Zugriffsbeispiele.

44: Im Oktober 1997 antworteten 20 Studenten auf die Frage nach ihrer Größe: 180 174 171 181 180 163 178 160 175 180 184 0 169 170 168 182 180 176 177 175. Legen Sie diese Daten auf `x` ab.

45: Welches ist der fünft-kleinste Wert?

46: Wie viele Angaben lagen unter der Durchschnittslänge?

47: Berechnen Sie, welche Indexposition der Wert 0 hat.

48: Ersetzen Sie die falsche Angabe 0 durch den Median aller Werte.

49: Entfernen Sie aus dem Datensatz alle 180er Längen.

5.2 Operatoren und Funktionen mit Wahrheitswertresultaten

50: Was ergibt: `T&(F|F|T)|(T&F|F)?`

51: Was läßt sich aus dem Vergleich von `T|F&F` und `F&F|T` schließen?

52: Wiederholen Sie die Beispiele des Abschnitts und überlegen Sie, wofür man *logische Operationen* benötigen kann.

53: Gilt `co2<500` für alle Werte?

54: Gibt es einen Studenten, der einen Wert größer als 200 angegeben hatte?

5.3 Modes und Strukturen von Datenobjekten

55: Was ist der Unterschied zwischen dem Mode und der Struktur eines Objektes?

56: Versuchen Sie in der Literatur zu ermitteln, ob eine Funktion ein Vektor ist.

5.4 Erzeugung von Matrizen, Listen und data frames

57: Erstellen Sie eine (10×10) -Matrix mit den Zahlen von 1 bis 100.

58: Erstellen Sie eine (10×10) -Matrix mit den Zahlen von 1 bis 100, jedoch soll die Matrix zeilenweise mit `1:100` gefüllt sein.

59: Was bewirkt `t(x)`, wenn `x` eine Matrix ist?

60: Erstellen Sie eine (10×10) -Einheitsmatrix.

61: Erstellen Sie eine Liste mit den Elementen `1`, `1:2` und `1:3`. Legen Sie die Liste auf `x` ab. Vergeben Sie geeignete Elementnamen.

62: Was bewirkt `unlist(x)`?

63: Was sind die Besonderheiten eines data frame? Geben Sie das Beispiel aus dem zugehörigen Abschnitt ein.

5.5 Feststellung von Attributen

64: Nehmen Sie den oben erstellten data frame und prüfen Sie, ob es sich um einen Vektor, eine Matrix, eine Liste und einen data frame handelt.

65: Ermitteln Sie Spaltennamen des data frame.

66: Erstellen Sie eine (2×3) -Matrix und legen Sie diese auf `x` ab. Was liefert: `length(x)`, `ncol(x)`, `col(x)`, `dim(x)` und `nrow(x)`? Ist das Objekt eine Liste?

5.6 Teilzugriffe auf Matrizen, Listen und data frames

67: Ist die zweite Spalte ihrer (2×3) -Matrix eine Matrix?

68: Entfernen Sie aus Ihrer Matrix die zweite Spalte.

69: Extrahieren Sie aus der Matrix das Element $(2, 2)$.

70: Erstellen Sie eine Liste durch `x<-list("a"=1:3,"b"=4,"c"=5:10)`. Greifen Sie nun über den Namen `a` auf die Liste zu: `x$a`. Was ergibt sich für ein Objekt?

71: Ist `x` ein Vektor?

72: Was ist `x[1]`, `x[2:3]`?

73: Was ist der Unterschied zwischen `x[1]` und `x[[1]]`?

6 Weitere und eigene Funktionen

6.1 Anweisungen in Dateien

74: Schreiben Sie mit einem Editor ihrer Wahl `print("Hello world!")` in eine Datei und lesen Sie diese dann in ihre Umgebung mittels `source` ein. Gegebenenfalls können Sie in der Hilfe die Syntax überprüfen.

6.2 Die Definition einer eigenen Funktion an einem Beispiel

75: Schreiben Sie die Funktion `wurzel` und überprüfen Sie die in der kleinen S-PLUS-Kunde aufgeführten Versuche.

76: Gießen Sie die Beobachtungen in Regeln.

77: Definieren Sie eine Funktion, die den Mittelwert eines Zahlenvektors ermittelt.

78: Definieren Sie eine Funktion, die alle negativen Werte eines Vektors entfernt und von den restlichen den Median ermittelt.

79: Falls Sie die letzte Funktion direkt in der Umgebung eingegeben haben, schreiben Sie sie nun in eine ASCII-Datei und laden Sie sie anschließend mit `source` in die Umgebung.

80: Was leistet die Funktion `fix`?

81: Was verstehen Sie unter Formal- und Aktual-Parametern?

82: Lesen Sie in einem Buch den Abschnitt über die Erstellung eigener Funktionen.

6.3 Kontrollstrukturen

83: Was ist der Bedeutungs-Unterschied zwischen einer `for` und einer `while`-Konstruktion?

84: Wann ist eine `while`-Konstruktion einer `repeat`-Konstruktion vorzuziehen?

85: Geben Sie ein Beispiel für den Gebrauch von `if`, `if-else`, `switch` und von `ifelse`.

86: Was spricht dafür, ohne Sprungmarken zu programmieren?

6.4 Die Vermeidung von Schleifen

87: Legen die Gewichte von 5 Studenten in der zweispaltigen Matrix `x` ab. Was erhält man dann mittels `apply(x,2,mean)` und `apply(x,2,min)`?

88: Legen Sie auf `x` die Liste: `list(1,1:2,1:3,1:10)` ab. Was liefert für diese Liste `lapply(x,length)`?

7 Einige Beispielanwendungen

7.1 Konzentrationsmessung

89: Was liefert `cumsum(sort(x))/sum(x)`?

7.2 Der t-Test

90: Schreiben Sie eine Funktion, die zu einem Datensatz und μ_0 die Prüfgröße des t-Tests ermittelt. Schreiben Sie auch noch eine, die die Prüfgröße zu einem Chiquadrat-Unabhängigkeitstest berechnet.

7.3 Verteilungsmodelle

91: Ziehen Sie aus der Standardnormalverteilung eine Stichprobe vom Umfang 100 und erstellen Sie hierzu Jitterplot, Boxplot, empirische Verteilungsfunktion und Histogramm.

92: Wiederholen Sie die letzte Aufgabe mit einer normalverteilten Grundgesamtheit mit den Parametern $\mu = 10$, $\sigma = 5$.

93: Ziehen Sie 100 Stichproben aus der Standardnormalverteilung vom Umfang 10 und ermitteln Sie jeweils den Mittelwert. Zeichnen Sie zu diesen Mittelwerten einen QQ-Plot, um festzustellen, ob eine Normalverteilung ein passendes Modell ist.

94: Wiederholen Sie die letzte Aufgabe mit Stichproben aus einem Cauchy-Modell

7.4 Monte-Carlo-Simulationen

95: Wiederholen Sie die angegebenen Beispiele.

7.5 Regression

96: Beschaffen Sie sich zwei gleich lange Vektoren \mathbf{x} und \mathbf{y} . Schätzen Sie nach der Methode der kleinsten Quadrate die Parameter der Gerade $y = ax + b$.

97: Probiere: `y<-co2` und `x<-19590:383/12`. Zeichne Scatterplot (x, y) mit Regressionsgeraden. Erstelle auch den Residualplot.

98: Was leistet `l1fit`?

8 Literatur

99: Blättern Sie in den angegebenen Literaturstellen und stöbern Sie in den dortigen Literaturverzeichnissen herum.

9 Zugabe

100: Sie haben sich um einen Job als Bankspion beworben und den Job wegen Ihrer sagenumwobenen S-PLUS-Kenntnisse erhalten. Mit der Eingabe von:

```
set.seed<-13; bank<-sample(100000,sample(1000,1))-sample(50000,1)
```

haben Sie sich den Vektor aller Kontostände der *Allgemeinen Pump- und Leihbank* beschafft. Ihr Auftraggeber erwartet die folgenden Informationen und Aktionen sowie die Dokumentation Ihres Schaffens in S-PLUS-Ausdrücken.

- Wie viele Kunden hat die Bank?
- Wie hoch ist der Gesamtbetrag aller Kontostände?
- Gibt es fast leere Konten, also mit Kontoständen zwischen 0 und 10?
- Wie hoch ist die Gesamtsumme auf den Konten mit negativen Kontoständen?
- Wie viele Konten weisen positive Kontostände auf?
- Wieviel *Miese* stehen im Durchschnitt auf einem Konto mit negativem Kontostand?
- Wie viele Konten haben einen Kontostand über 70000?
- Welches sind der kleinste und der größte Kontostand?
- Gibt es Konten mit Kontoständen zwischen 43050 und 53500?
- Sie erhalten 0.1% aller positiven Kontostände über 50000 als Honorar *ausbezahlt*. Geben Sie Ihre Gehaltformel an.
- Kurz vor Ende Ihres Jobs stiften Sie Verwirrung, indem Sie allen Kunden mit weniger als 3000 Miesen ihre Schulden erlassen. Merken Sie sich den angerichteten Schaden.
- Sorgen Sie dafür, daß die Konten mit mehr als 60000 Ihren Großmut aus der vorangegangenen Aufgabe zu gleichen Teilen übernehmen.
- Bevor Sie die Bank verlassen, runden Sie bitte aus Gründen der Optik alle Kontostände auf volle Tausender.

Natürlich werden Sie nach einigen Vorversuchen einige Anweisungen in (einer) Funktion(en) zusammenfassen. Sie können dabei sicher auch die Funktion `menu` geschickt verwenden. Und Sie könnten Ihre Auftraggeber sicher auch durch einige geschickte Graphiken für Folgeaufträge sensibilisieren.