

Mathematik II für Biologen

Übungsblatt 4 (Abgabe am 15.5.2009)

Aufgabe 10 MATLAB

(10 Punkte)

21 Labore bestimmten den Kupfergehalt in je einer Probe aus 9 verschiedenen Klärschlammlieferungen. Die Daten sind in der Datei `schlamm.dat` zu finden. Dabei befindet sich in der i -ten Zeile und j -ten Spalte der darin enthaltenen Matrix das Ergebnis der Untersuchung der j -ten Probe durch das i -te Labor.

- a) Zeichnen Sie in einem gemeinsamen Diagramm für jede Probe einen Boxplot für die zugehörigen 21 Labordaten:
» `load schlamm.dat; boxplot(schlamm)`
Welche Probe scheint am stärksten mit Kupfer belastet zu sein, welche am geringsten?
- b) Zeichnen Sie in einem gemeinsamen Diagramm für jedes Labor einen Boxplot für die zugehörigen 9 Proben:
» `boxplot(schlamm')` % Erinnerung: ' transponiert eine Matrix.
Zwei Labore tanzen aus der Reihe. Welche und auf welche Weise?
- c) Um die Frage aus (b) genauer zu untersuchen, berechnen wir zunächst von jeder Probe den Median der Messungen:
» `med=median(schlamm)`
Wir betrachten diese Mediane in erster Näherung als den wahren Kupfergehalt der jeweiligen Probe. Abweichungen von diesem Median können dann als Fehler des jeweiligen Labors interpretiert werden. Zeichnen Sie in einem gemeinsamen Diagramm für jedes Labor einen Boxplot für die 9 Fehler, die dieses Labor gemacht hat. Interpretieren Sie das Ergebnis.
HINWEIS: `ones(21,1)*med` ist das Produkt eines Spalten- mit einem Zeilenvektor, nämlich eine 21×9 -Matrix, deren Zeilen alle identisch gleich `med` sind.
- d) Betrachten Sie die Matrix
» `corrcoef(schlamm')`
In welcher Größenordnung liegen die meisten der Zahlen in dieser Matrix? Was bedeutet dies? Wie unterscheidet sich die letzte Spalte (oder Zeile) dieser Matrix von den anderen? Was bedeutet dies?
HINWEIS: Wo in der Matrix finden Sie die Werte, die `corr(schlamm(1,:)',schlamm(2,:))` und `corr(schlamm(1,:)',schlamm(21,:))` liefern?
BEMERKUNG: Sie müssen diese Matrix nicht ausdrucken, um diese Aufgabe zu bearbeiten.
- e) Tragen Sie in einem Streudiagramm die 9 Werte des zweiten Labors (auf der vertikalen Achse) gegen die Werte des ersten Labors (auf der horizontalen Achse) auf.
» `plot(schlamm(???,:),schlamm(2,:),'o')`
Tun Sie das Gleiche mit dem letzten und dem ersten Labor. Wie spiegeln die beiden Diagramme die Ergebnisse aus Aufgabe d) wider?

Aufgabe 11 MATLAB

(10 Punkte)

Um zu bestimmen, wie die Geschwindigkeit einer chemischen Reaktion $A + B \rightarrow C$ von der Temperatur abhängt, wird eine Lösung von A und B bei verschiedenen Temperaturen T (in $^{\circ}C$) jeweils 5 Minuten lang geschüttelt. Danach wird die Konzentration c (in mmol/ℓ) von C gemessen. Dies sind die Ergebnisse:

T	4	8	12	16	20	24	28	32	36	40
c	1.60	2.34	3.41	3.48	5.90	8.38	9.24	13.79	28.09	31.16

- Welches ist die erklärende (“unabhängige”) Variable, welches die Zielgröße (“abhängige”) Variable)?
- Stellen Sie die Daten mit MATLAB in einem Streudiagramm dar, wobei T auf der waagerechten und c auf der senkrechten Achse aufgetragen wird.
- Berechnen Sie Pearsons Produktmomenten-Korrelationskoeffizient (kurz: die Korrelation) von T und c (d.h. r_{Tc}).
- Um ein Streudiagramm zu erhalten, das eher an eine Gerade erinnert, transformieren wir die Daten und tragen $\log(c)$ auf der senkrechten Achse auf. Zeichnen Sie das entsprechende Streudiagramm und fügen Sie die Regressionsgerade ein.
- Lesen Sie grob die Gleichung der Regressionsgeraden aus Teil d) ab, und geben Sie den entsprechend erwarteten funktionalen Zusammenhang zwischen c und T an. Zeichnen Sie diesen auch zum Vergleich in das Diagramm aus Teil b) ein.
- Berechnen Sie Pearsons Produktmomenten-Korrelationskoeffizient (kurz: die Korrelation) von T und $L := \log(c)$ (d.h. r_{TL}).
- Vergleichen Sie die Ergebnisse aus c) und f) und erklären Sie den Unterschied.
- Bestimmen Sie Spearmans Rangkorrelationskoeffizienten (kurz: die Rangkorrelation) von T und c (d.h. $r_{Tc}^{(\text{SP})}$) und von T und L (d.h. $r_{TL}^{(\text{SP})}$). (Tipp: Nicht rechnen, sondern Daten anschauen.) Begründen Sie Ihr Ergebnis.

MATLAB-Code, unvollständig:

```
>> T=[4:4:40]
>> c=[1.60,2.34,3.41,3.48,5.90,8.38,9.24,13.79,28.09,31.16]
>> plot(T,c,'o')
>> corrcoef(T,c)
>> L=log(c);
>> plot(???,???, 'o')
>> lsline
>> plot(T,c,'o'); hold on; plot(T,??); hold off
```

Aufgabe 12 MATLAB

(10 Punkte)

Der MATLAB-Befehl $\mathbf{x}=\text{rand}(1,500)$ erzeugt einen Zeilenvektor $x = (x_1, \dots, x_n)$ von $n = 500$ Zufallszahlen zwischen 0.0 und 1.0. Finden Sie einen Vektor $y = (y_1, \dots, y_n)$ von Zahlen, so dass die bivariate Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ (mit hoher Wahrscheinlichkeit) Pearsonsche Korrelation zwischen 0,7 und 0,9 hat.

HINWEIS:

```
> x=rand(1,500);y=x;plot(x,y,'o');corrcoef(x,y)
liefert zu große Korrelationen, wohingegen sie für
> x=rand(500,1);y=rand(500,1);plot(x,y,'o');corr(x,y)
(betragsmäßig) zu klein sind.
```