Universität Tübingen
Mathematisches Institut
Dr. Stefan Keppeler

Sommersemester 08

Mathematik II für Biologen

Übungsblatt 4 (Abgabe am 07.05.2008)

Aufgabe 11 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?

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 12 MATLAB (10 Punkte)

Der MATLAB-Befehl `x=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 0.6 und 0.8 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.

Aufgabe 13 MATLAB (10 Punkte)

In dieser Aufgabe werden wir mittels (Monte-Carlo-)Simulation zu unterscheiden versuchen zwischen Ereignissen, die "durchaus" eintreten können, und solchen, die "praktisch unmöglich" sind.

Angenommen, Sie vermuten, dass Ihr Würfel gezinkt, d.h. nicht ideal, nicht fair ist. Und zwar vermuten Sie, dass die 6 nicht so häufig erscheint, wie sie es sollte, wenn der Würfel fair wäre. Um Ihre Vermutung zu testen, werfen Sie den Würfel 100 Mal. Sei X die Anzahl der dabei geworfenen Sechsen. Sie beobachten $X = 12$. Nun stellt sich die Frage, ob diese Beobachtung mit der (Null-)Hypothese vereinbar ist, dass der Würfel fair ist, oder ob sie die (Alternativ-)Hypothese unterstützt, dass der Würfel nicht fair ist.

Um herauszufinden, welche Werte von X durchaus beobachtet werden können, falls der Würfel fair ist, simulieren wir das Experiment, einen fairen Würfel 100 Mal zu werfen, und wiederholen diese Simulation 10000 Mal. MATLAB Programm dazu:

```
>> n=10000;
>> mat=unidrnd(6,100,n); % Ergibt eine 100 x n Matrix mat mit Eintraegen,
     % die zufaellig und unabhaengig voneinander 1,2,3,4,5 oder 6 sind,
     % wobei keine Zahl bevorzugt wird. Auf diese Weise wird ein idealer
     % Wuerfel simuliert. (Versuchen Sie es mit kleineren Zahlen n und
     % lassen Sie das Semikolon weg, wenn Sie sehen wollen, wie mat
     % aussieht.)
>> ist6= mat==6; % neue Matrix mit Eintrag 1, falls entsprechender Eintrag
                 % in mat eine 6 ist und 0 sonst.
>> wieoft6=sum(ist6)  % summiert ueber Spalten
>> r=hist(wieoft6,0:35)/n;
     % r = Vektor der relativen Haeufigkeit von Experimenten, in denen
     % keinmal 6 gewuerfelt wurde, genau einmal, genau zweimal, ...,
     % genau 34 Mal, mindestens 35 Mal.
>> bar(0:35,r) % zugehoeriges Stabdiagramm (ein Histogramm)
```

a) In wieviel Prozent der Fälle ist bei Ihnen in 100 Würfen genau 12 Mal die 6 aufgetreten? Wie liest man dies aus der Computer-Ausgabe ab?

b) In wieviel Prozent der Fälle ist bei Ihnen in 100 Würfen höchstens 12 Mal die 6 aufgetreten? Wie liest man dies aus der Computer-Ausgabe ab?

c) Benutzen Sie den Vektor r, den Sie erhalten haben, um zu entscheiden, welche Werte "praktisch unmöglich" sind. Nehmen Sie hierbei $\alpha = 5\%$ als Signifikanz-Niveau, d.h. erklären Sie (etwa) $\alpha/2 = 2.5\%$ der kleinsten tatsächlich beobachteten Werte von X und $\alpha/2 = 2.5\%$ der größten beobachteten Werte von X für "praktisch nicht beobachtbar", wenn das Experiment nur ein einziges Mal durchgeführt wird. Gehört $X = 12$ demgemäß zu den (für $\alpha = 5\%$) "praktisch unmöglichen" Werten? HINWEIS:

```
>> cumsum(r)      % kumulative Summe von r: An der n-ten Stelle steht
                  % die Summe aller Eintraege von r links der n-ten
                  % Stelle einschliesslich der n-ten Stelle selbst.
```

d) Angenommen, jemand schlägt vor, genau dann die (Null-)Hypothese H_0 : *Der Würfel ist fair* zu verwerfen und stattdessen an die (Alternativ-)Hypothese H_A : *Der Würfel ist unfair* zu glauben, wenn bei der einmaligen Durchführung des Experimentes entweder $X \leq 10$ oder $X \geq 22$ beobachtet wurde.
In wieviel Prozent der oben simulierten 10000 Fälle würde man dann H_0 verwerfen müssen, obwohl wir wissen, dass H_0 wahr ist?