Universität Tübingen
Mathematisches Institut
Dr. Stefan Keppeler

Sommersemester 2010

Mathematik II für Biologen

Übungsblatt 9 (Abgabe am 18.6.2010)

Aufgabe 32 MATLAB (10 Punkte)

Die Datei `fishy.dat` enthält in der ersten Spalte die Länge (in cm), in der zweiten Spalte das Gewicht (Einheit leider unbekannt) und in der dritten Spalte den DDT-Gehalt (in ppm) von $n = 96$ Welsen, die im Tennessee River in Alabama, USA, gefangen wurden (Quelle: Mendenhall, Sincich: *Statistics for Engineering and the Sciences*, Prentice Hall. Appendix III. Daten leicht geändert, daher das y im Dateinamen.).

a) Bestimmen Sie den empirischen Median der DDT-Gehalte.
b) Ist dieser empirische Median signifikant (zum Signifikanz-Niveau $\alpha = 5\%$) von 10 verschieden? Beantworten Sie diese Frage mit einem zweiseitigen Vorzeichentest.
c) Bestimmen Sie das zum Vorzeichentest gehörige 95%-Vertrauensintervall $[a, b]$ für den "wahren, theoretischen" Median des DDT-Gehaltes eines solchen Fisches, indem Sie den Test aus Aufgabe (b) für verschiedene Werte wiederholen und dabei beobachten, ob der Test verwirft. Bestimmen Sie dabei a und b so genau, dass die erste Ziffer nach dem Komma sicher stimmt.

Unvollständiger MATLAB-Code:

```
load fishy.dat
ddt=fishy(:,3);
[p,h]=signtest(ddt,10)  % help signtest
```

Aufgabe 33 (10 Punkte)

Die Inkubationszeit X (in Monaten) einer bestimmten ansteckenden Krankheit wird als *log-normal* verteilt mit $\mu = 0$ und $\sigma = 0{,}4$ modelliert, d.h. man nimmt an, dass $\log X \sim \mathcal{N}(0; 0{,}4^2)$, wobei log der natürliche Logarithmus ist.

a) Geben Sie die Wahrscheinlichkeit an, dass die Inkubationszeit mehr als 3 Monate beträgt.
   HINWEIS: Der MATLAB-Befehl `normcdf(x)` berechnet $\Phi(x)$.
b) Berechnen Sie den Median med von X.
c) Für eine $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariable Y gilt bekanntlich

$$\begin{aligned} P[\mu - \sigma \leq Y \leq \mu + \sigma] &\approx 68\% \qquad \text{und} \\ P[\mu - 1{,}96\sigma \leq Y \leq \mu + 1{,}96\sigma] &\approx 95\%. \end{aligned}$$

   Finden Sie Konstanten c_1 und c_2, so dass

$$\begin{aligned} P\left[\frac{\text{med}}{c_1} \leq X \leq \text{med} \cdot c_1\right] &\approx 68\% \qquad \text{und} \\ P\left[\frac{\text{med}}{c_2} \leq X \leq \text{med} \cdot c_2\right] &\approx 95\%. \end{aligned}$$

Aufgabe 34 (Quantil-Quantil-Diagramm, Q-Q-Plot) (10 Punkte)

Uns liege das Histogramm einer Stichprobe vor, das auf den ersten Blick mehr oder weniger glockenförmig (wie eine Gauß-Kurve) aussieht. Daher stellen wir uns die Frage, ob es Parameter μ und σ gibt, so dass der Plot der Dichte der Normalverteilung,

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

das Histogramm gut beschreibt.

Die folgende Darstellungsart, genannt Quantil-Quantil-Diagramm oder kurz Q-Q-Plot (auch normal plot und manchmal leider wenig spezifisch einfach nur Wahrscheinlichkeits-Diagramm), erlaubt es einem, dies zu überprüfen, ohne dafür zunächst die passenden Parameterwerte μ und σ zu bestimmen (bzw. zu raten, zu schätzen oder auszuprobieren). Auch können mit diesem Diagramm Abweichungen von der Gauß-Kurve leichter beurteilt werden.

Dazu definieren wir zunächst für $0 < \alpha < 1$ das (theoretische) α-Quantil $q_\alpha^{(\Phi)}$ für die Dichte der Standardnormalverteilung, $f_{0,1}$, und zwar ist $q_\alpha^{(\Phi)} \in \mathbb{R}$ diejenige Zahl, für die

$$\Phi\left(q_\alpha^{(\Phi)}\right) = \alpha\,, \quad \text{wobei} \quad \Phi(x) := \int_{-\infty}^{x} f_{0,1}(x)\,\mathrm{d}x = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x\,,$$

z.B. ist $q_{0{,}975}^{(\Phi)} = 1{,}96$. Im Q-Q-Plot werden dann die Punkte

$$\left(q_{(i-1/2)/n}^{(\Phi)},\, x_{(i)}\right)_{i=1,\ldots,n}$$

in einem zweidimensionalen Diagramm eingetragen, wobei $(x_{(i)})_{i=1,\ldots,n}$ die der Größe nach geordnete ursprüngliche Stichprobe ist.

Kurz gesagt, trägt man also die theoretischen Quantile der Normalverteilung gegen die empirischen Quantile der Stichprobe auf.

a) Welchen Wert hat die empirische Verteilungsfunktion $F(x)$ für x etwas kleiner als $x_{(i)}$ und für x etwas größer als $x_{(i)}$?
b) Erzeugen Sie mit dem MATLAB-Befehl `qqplot` einen Q-Q-Plot der Daten aus `states.dat` aus Aufgabe 4. Markieren Sie darin die beiden Ausreißerstaaten aus Aufgabe 4d (mit Namen).
c) Lassen sich die Daten aus `states.dat` gut durch eine Normalverteilung beschreiben? (Bearbeiten Sie, bevor Sie diese Frage beantworten, zuerst die nächste Aufgabe).
d) Schätzen Sie mithilfe des Q-Q-Plots aus Teil b grob die Parameter μ und σ der in Teil c erwähnten Normalverteilung. Wie sind Sie bei der Schätzung vorgegangen und warum?

Aufgabe 35 (10 Punkte)

Unten wird für die 6 Stichproben aus Aufgabe 1 jeweils der Q-Q-Plot gezeigt. Ordnen Sie die Q-Q-Plots a-f den Histogrammen A-F aus Aufgabe 1 zu, und begründen Sie kurz Ihre Entscheidung. Welche Histogramme lassen sich besser, welche schlechter durch eine Gauß-Kurve beschreiben? Wie sieht man dies den zugehörigen Q-Q-Plots an?