

Mathematik II für Biologen

Übungsblatt 3 (Abgabe am 30.04.2007)

Das Übungsblatt besteht aus 4 Seiten (Aufgaben 7, 8, 9 & 10).

Aufgabe 7 MATLAB

(10 Punkte)

Die Datei `states.dat` auf der Vorlesungshomepage gibt für jeden Staat der USA den Prozentsatz der über 65-Jährigen an der Bevölkerung an.

- Zeichnen Sie mit MATLAB (oder einem anderen Programm) ein Histogramm der Daten und einen Boxplot (bitte ausdrucken).
- Berechnen Sie
 - den arithmetischen Mittelwert \bar{x} ,
 - die empirische Standardabweichung s_x ,
 - die Quartilsdifferenz,
 - den Median und
 - die Medianabweichung MAD. (Achtung: `mad` in MATLAB bedeutet etwas anderes.)
- Wieviel Prozent der Daten liegen in den Intervallen $\bar{x} \pm s_x$, $\bar{x} \pm 2s_x$ und $\bar{x} \pm 3s_x$? Vergleichen Sie Ihr Ergebnis mit der Vorhersage der Faustregel aus der Vorlesung.
- Die Prozentzahlen welcher Staaten weichen um mehr als 2 Standardabweichungen s_x vom Mittelwert \bar{x} ab? Warum könnten gerade diese Staaten extreme Werte aufweisen? Identifizieren Sie diese Daten im Boxplot.

MATLAB-Code¹, unvollständig:

```
>> load states.dat
>> hist(states)
>> boxplot(states)
>> xquer=mean(states)
>> s=std(states)
>> iqr(states)
>> med=median(states)
>> MAD=median(abs(states-???)
>> n2 = xquer-2*s < states & states < xquer + 2*s
    % Ergibt einen Vektor n2 mit 1 an der Position i, falls
    % xquer-2*s < states(i) < xquer + 2*s und 0 sonst.
>> sum(n2)/length(states)
```

Aufgabe 8

(10 Punkte)

Es wird ein Loch in einen Permafrostboden gebohrt. In den Tiefen von 0, 2, 4, 10, 15 und 20 Metern wird dann die Bodentemperatur gemessen. Diese könnte im Sommer folgende Werte haben: $3^\circ C$, $1^\circ C$, $0^\circ C$, $-0,2^\circ C$, $-0,5^\circ C$, und $-0,2^\circ C$.

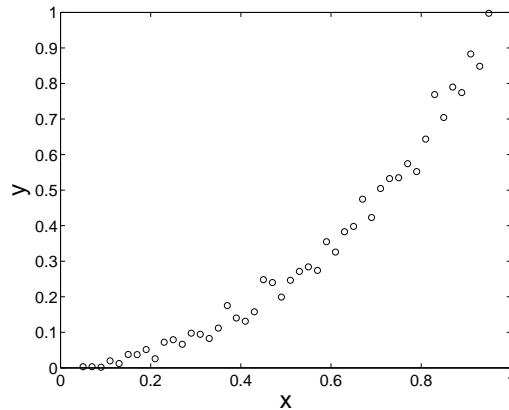
- Berechnen Sie die Pearsonsche Produktmomenten-Korrelation und die Spearmansche Rangkorrelation.
- Zeichnen Sie ein Streudiagramm der Temperatur in Abhängigkeit von der Tiefe. Interpretieren Sie die Vorzeichen der in Aufgabe (a) bestimmten Korrelationskoeffizienten anhand des Streudiagrammes.

¹Verwenden Sie die Hilfe-Funktion, um mehr über die verwendeten Befehle zu erfahren.

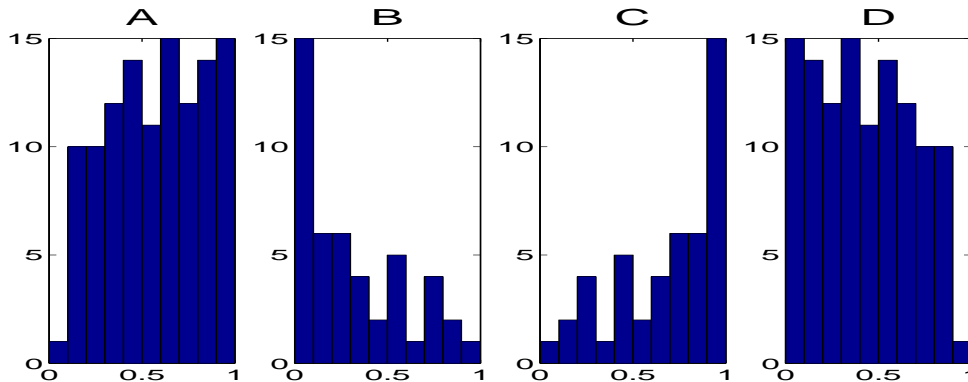
Aufgabe 9

(10 Punkte)

In einem Experiment wurde eine Stichprobe $(x_1, y_1), \dots, (x_{46}, y_{46})$ von Zahlenpaaren gewonnen, die in folgendem Streudiagramm dargestellt wird.

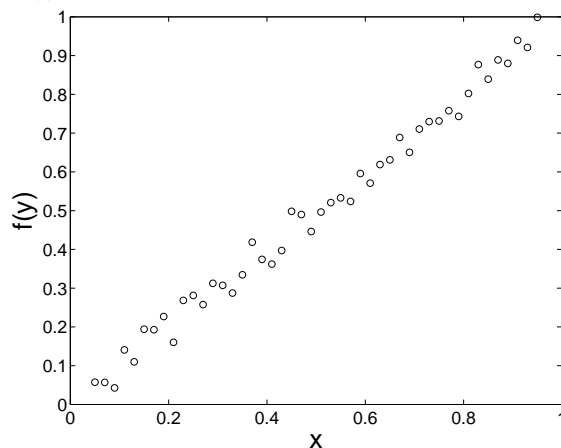


a) Welches der folgenden Histogramme ist das Histogramm der Stichprobe y_1, \dots, y_{46} ? Begründen Sie kurz Ihre Wahl.



b) Nun wurden die Daten so transformiert, dass das Streudiagramm an eine gerade Linie erinnert. Für welche der folgenden Funktionen ist das folgende Streudiagramm das Streudiagramm der Stichprobe $(x_1, f(y_1)), \dots, (x_{46}, f(y_{46}))$? Begründen Sie kurz Ihre Wahl.

- (i) $f(y) = e^y$
- (ii) $f(y) = y^2$
- (iii) $f(y) = \log y$
- (iv) $f(y) = \sqrt{y}$
- (v) $f(y) = 1/y$



c) Legen Sie nach Augenmaß eine Regressionsgerade durch das Streudiagramm in Aufgabe (b) und lesen Sie deren Gleichung ab.
 d) Finden Sie damit eine Funktion g , für die

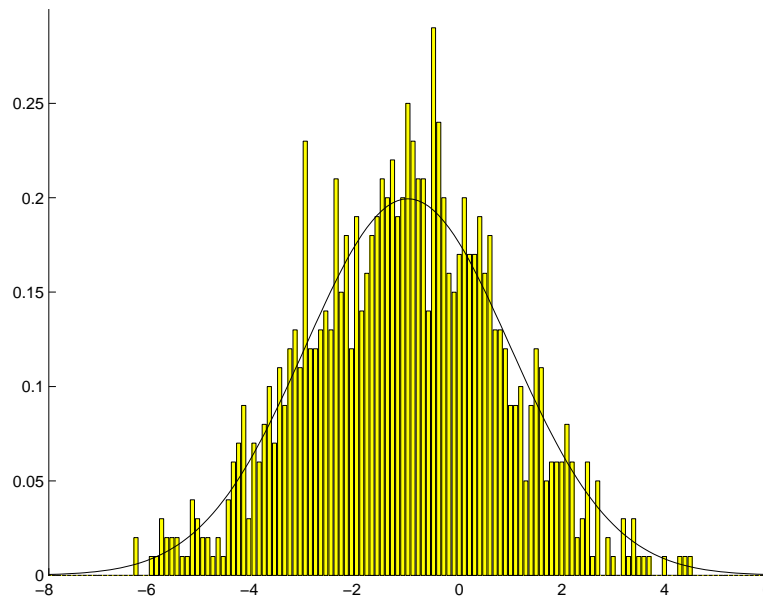
$$y_i = g(x_i) + \text{kleiner Fehler}$$

gilt.

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

(10 Punkte)

Die folgende Abbildung zeigt das Histogramm zu einer eindimensionalen Stichprobe x_1, \dots, x_n der Größe $n = 1000$.



(Die Höhen der Balken des Histogramms wurden dabei so normiert, dass die Gesamtfläche des Histogramms gleich 1 ist.) Auch wenn das Histogramm noch recht unregelmäßig ist, so ist es doch in etwa "glockenförmig". Man mag versucht sein, eine glatte Kurve hindurchzulegen, die das Histogramm beschreibt. Dies ist in der Abbildung bereits geschehen, und zwar ist dort der Graph der Ihnen aus der Mathematik I bekannten Gauß-Funktion

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

für $\mu = -1$ und $\sigma = 2$ eingezeichnet. ($f_{\mu,\sigma}$ ist die sogenannte Dichte der Normalverteilung mit Erwartungswert μ und Standardabweichung σ , die später noch näher behandelt wird.) Augenscheinlich beschreibt diese Kurve das Histogramm recht gut, d.h. es gibt eine geeignete μ und ein geeignetes σ , für das das Histogramm gut durch $f_{\mu,\sigma}$ angenähert wird. Für andere $\mu \neq -1$ oder $\sigma \neq 2$ ist dies i.a. nicht der Fall.

Die folgende Darstellungsart, genannt Quantil-Quantil-Diagramm (Q-Q-Plot, auch normal plot und manchmal leider wenig spezifisch einfach nur Wahrscheinlichkeits-Diagramm genannt) erlaubt es einem, zu überprüfen, ob eine Stichprobe bzw. deren Histogramm gut durch irgendeine Gauß-Kurve beschrieben werden kann, ohne dass man sich dabei über die geeignete Wahl von μ und σ Gedanken machen muss. Auch können mit diesem Diagramm Abweichungen von der Gauß-Kurve leichter beurteilt werden.

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

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

(Z.B. ist $q_{0,975}^{(\Phi)} = 1.96$. Bem.: $\Phi(\infty) = 1$.) Der Q-Q-Plot ist dann das Streudiagramm zur bivariaten Stichprobe

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

wobei $(x_{(i)})_{i=1, \dots, n}$ die der Größe nach geordnete ursprüngliche Stichprobe ist. (Bitte wenden.)

Fortsetzung Aufgabe 10

Unten sind zu 6 Stichproben der Größe $n = 200$ jeweils das Histogramm und der Q-Q-Plot gezeigt, wobei die Beschriftung der y -Achse weggelassen wurde. Finden Sie die richtigen Paare und begründen Sie kurz Ihre Entscheidung. Welches Histogramm lässt sich am besten durch eine Gauß-Kurve beschreiben? Wie sieht man dies dem zugehörigen Q-Q-Plot an?

