Fachbereich Mathematik
M. Kohler
A. Fromkorth
J. Mehnert

TECHNISCHE UNIVERSITÄT DARMSTADT

SS 2009
12. Mai 2009

Lösungsvorschläge zum 3. Übungsblatt zur „Einführung in die Stochastik“

Lösung zur Aufgabe 10 (3 Punkte)

(a) Zunächst sollte man die Daten aufsteigend sortieren. Dies ergibt:

$$3.6, 7.7, 8.8, 12.7, 20.1, 22.4, 24.0, 24.8, 25.4, 30.4, 32.2, 33.0, 36.1, 36.4, 37.3,$$
$$38.6, 38.6, 49.4, 50.6, 51.5, 53.2, 54.8, 59.6.$$

Als Mittel erhält man 32.6607. Da die Anzahl ungerade ist, erhalten wir als Median den 12. Datenpunkt, also 33. Die Spannweite beträgt 56, die Varianz 254.8137 und die Standardabweichung 15.96288. Der Interquartilabstand der Messreihe ist 27 (Wert des 18. Datenpunkts - Wert des 6. Datenpunkts).

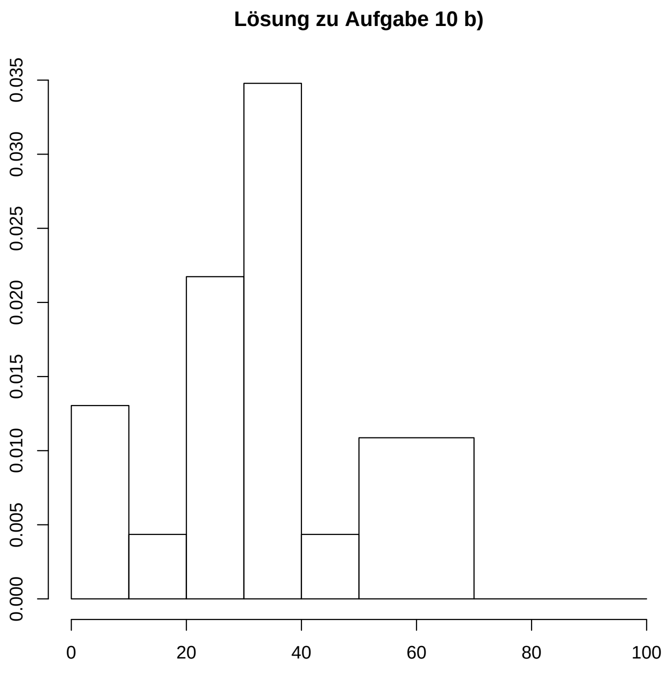


Abbildung 1: Abbildung zu Aufgabe 10 b).

Lösung zur Aufgabe 11 (3 Punkte)

Wie in der Vorlesung müssen wir die partiellen Ableitungen nullsetzen. Dies ergibt

$$\frac{\partial}{\partial a}F(a,b,c) = \frac{\partial}{\partial b}F(a,b,c) = \frac{\partial}{\partial c}F(a,b,c) = 0.$$

Wir berechnen also die partiellen Ableitungen

$$\begin{aligned}
0 = \frac{\partial}{\partial a}F(a,b,c) &= \frac{\partial}{\partial a}\left(\sum_{i=1}^{n}\left((y_i - (a + bx_i + cx_i^2)\right)^2\right) = \sum_{i=1}^{n}\frac{\partial}{\partial a}\left(\left(y_i - (a + bx_i + cx_i^2)\right)^2\right) \\
&= \sum_{i=1}^{n} 2\left(y_i - (a + bx_i + cx_i^2)\right)\frac{\partial}{\partial a}\left(y_i - (a + bx_i + cx_i^2)\right) \\
&= -2\sum_{i=1}^{n}\left(y_i - (a + bx_i + cx_i^2)\right) \\
&= -2\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i - \sum_{i=1}^{n} cx_i^2\right) \\
\Rightarrow \qquad & \frac{1}{n}\sum_{i=1}^{n} y_i = a + b\frac{1}{n}\sum_{i=1}^{n} x_i + c\frac{1}{n}\sum_{i=1}^{n} x_i^2.
\end{aligned}$$

$$\begin{aligned}
0 = \frac{\partial}{\partial b}F(a,b,c) &= \frac{\partial}{\partial b}\left(\sum_{i=1}^{n}\left((y_i - (a + bx_i + cx_i^2)\right)^2\right) = \sum_{i=1}^{n}\frac{\partial}{\partial b}\left(\left(y_i - (a + bx_i + cx_i^2)\right)^2\right) \\
&= \sum_{i=1}^{n} 2\left(y_i - (a + bx_i + cx_i^2)\right)\frac{\partial}{\partial b}\left(y_i - (a + bx_i + cx_i^2)\right) \\
&= -2\sum_{i=1}^{n}\left(y_i - (a + bx_i + cx_i^2)\right)\cdot x_i \\
&= -2\left(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} ax_i - \sum_{i=1}^{n} bx_i^2 - \sum_{i=1}^{n} cx_i^3\right) \\
\Rightarrow \qquad & \frac{1}{n}\sum_{i=1}^{n} x_i y_i = a\frac{1}{n}\sum_{i=1}^{n} x_i + b\frac{1}{n}\sum_{i=1}^{n} x_i^2 + c\frac{1}{n}\sum_{i=1}^{n} x_i^3.
\end{aligned}$$

und

$$\begin{aligned}
0 = \frac{\partial}{\partial c}F(a,b,c) &= \frac{\partial}{\partial c}\left(\sum_{i=1}^{n}\left((y_i - (a + bx_i + cx_i^2)\right)^2\right) = \sum_{i=1}^{n}\frac{\partial}{\partial c}\left(\left(y_i - (a + bx_i + cx_i^2)\right)^2\right) \\
&= \sum_{i=1}^{n} 2\left(y_i - (a + bx_i + cx_i^2)\right)\frac{\partial}{\partial c}\left(y_i - (a + bx_i + cx_i^2)\right) \\
&= -2\sum_{i=1}^{n}\left(y_i - (a + bx_i + cx_i^2)\right)\cdot x_i^2 \\
&= -2\left(\sum_{i=1}^{n} x_i^2 y_i - \sum_{i=1}^{n} ax_i^2 - \sum_{i=1}^{n} bx_i^3 - \sum_{i=1}^{n} cx_i^4\right) \\
\Rightarrow \qquad & \frac{1}{n}\sum_{i=1}^{n} x_i^2 y_i = a\frac{1}{n}\sum_{i=1}^{n} x_i^2 + b\frac{1}{n}\sum_{i=1}^{n} x_i^3 + c\frac{1}{n}\sum_{i=1}^{n} x_i^4.
\end{aligned}$$

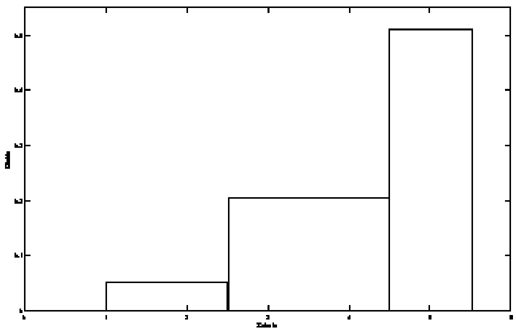
Dies sind die gewünschten Gleichungen.

Anmerkung für Studenten ab dem 3. Semester: Um zu zeigen, dass es sich in der Tat um ein Minimum handelt, muss man jetzt noch nachrechnen, dass die zugehörige Hessematrix positiv definit ist.

Lösung zur Aufgabe 12 (3 Punkte)

a) a$_1$) Die graphische Darstellung ist irreführend, da die Klassen, bzw. die Länge der Intervalle nicht alle gleich lang sind. Vergleicht man beispielsweise den Flächeninhalt der mittleren mit der rechten Klasse, so entsteht der Eindruck, dass die mittlere Klasse fast anderthalb mal so viele Datenpunkte enthält wie die rechte Klasse und das ist falsch!

a$_2$) Anzahl der Studenten in der Klasse 1, d.h. im Intervall $I_1(= [1, 2.5)) \approx 20$, in $I_2 \approx 105$ und in $I_3 \approx 130$. Somit ergibt sich folgendes Histogramm:



b) Beim Histogramm gibt der Flächeninhalt (FI) einer Klasse j den prozentualen Anteil der Datenpunkte (PAD) im zugrunde liegenden Intervall (I_j) an. Somit lässt sich die Anzahl der Datenpunkte in Klasse j wie folgt berechnen: 1. Möglichkeit:

$$\text{PAD in } I_j = \text{FI von } I_j$$
$$\Rightarrow \text{Anzahl der Datenpunkte in } I_j = \text{GD x PAD in } I_j$$

mit GD = Gesamtzahl der Datenpunkte 2. Möglichkeit

$$\text{Höhe von } I_j = \frac{n_j}{n \cdot \lambda(I_j)}$$
$$\Rightarrow \quad n_j = n \cdot \lambda(I_j) \cdot \text{Höhe von } I_j$$

mit n_j = Anzahl der Datenpunkte im j-ten Intervall und $\lambda(I_j)$ = Länge des j-ten Intervalls. Hier: $n_j \approx 182$

Lösung zur Aufgabe 13 (3 Punkte)

(a) Wird in der Abbildung die rechte obere mit der rechten unteren Graphik verglichen, dann ist festzustellen, dass der relative Anteil bei den Noten zwischen 4 und 6 in der Kontrollgruppe deutlich höher ist, als in der nicht ausgew./betracht. Stud., d.h. die betrachtete Fläche ist größer. Somit lässt sich die Aussage feststellen.

(b) Betrachtet man die Graphik "Noten Studiengruppe", dann ist der relative Anteil bei den Noten zwischen 4 und 6 deutlich niedriger als bei der Graphik "Noten Kontrollgruppe" (im gleichen Notenintervall). Daher lässt sich folgern, dass das Anbieten des Zusatzkurses zu einer Verringerung der Nichtbestehensquote bei den durch das Verfahren ausgewählten StudentInnen geführt hat.