

Einführung in die Stochastik

3. Übungsblatt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Mathematik
M. Kohler
A. Fromkorth
D. Furer

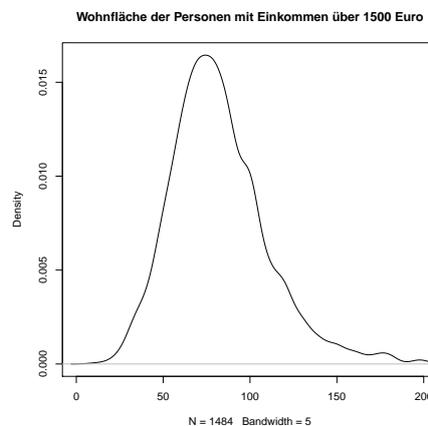
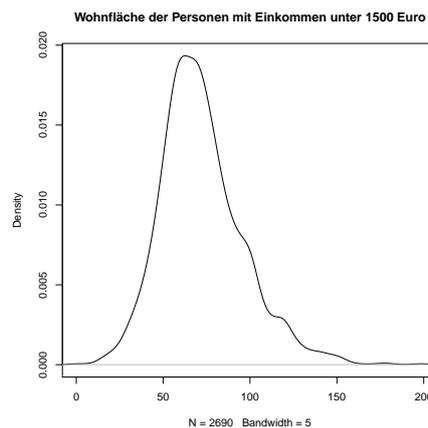
SS 2011
06.05.2011

Gruppen und Hausübung

Aufgabe 9

(4 Punkte)

Der Mikrozensus ist eine statistische Erhebung. Hierbei werden nach bestimmten Zufallskriterien Haushalte ausgewählt, die Daten zu unterschiedlichen Merkmalen liefern, wie z.B. Nettoeinkommen, Alter, Arbeitszeit, Wohnfläche. In den unten stehenden Abbildungen sind Kern-Dichteschätzer (mit Gauß-Kern) angewandt auf die Wohnfläche bei Personen mit einem Nettoeinkommen unter 1500 Euro und mit einem Nettoeinkommen von über 1500 Euro. Welche Aussagen lassen sich anhand dieser Grafiken treffen? Begründen Sie diese.



Hinweis: Betrachten Sie z.B. als Referenz die Stelle 100 m^2 .

Aufgabe 10

(4 Punkte)

In der folgenden Tabelle sind die Ausgaben pro Student (in Euro) und die Arbeitslosenquote (in Prozent) in den sechs neuen Bundesländern im Jahr 2001 angegeben.

	Ausgaben pro Student (in Euro)	Arbeitslosenquote (in Prozent)
Berlin	8100	17.9
Brandenburg	6600	18.8
M.-V.	8700	19.6
Sachsen	8700	19
Sachsen-Anhalt	9900	20.9
Thüringen	8800	16.5

- (a) Zeichnen Sie ein Streudiagramm (Scatterplot) der Daten, wobei sie als x -Wert die Ausgaben pro Student und als y -Wert die Arbeitslosenquote verwenden.
- (b) Bestimmen Sie mit Hilfe der in der Vorlesung hergeleiteten allgemeinen Formel die zugehörige Regressionsgerade und zeichnen Sie diese in das Streudiagramm aus a) ein.
- (c) Inwieweit ändert sich das Resultat in b), wenn man den zu Sachsen-Anhalt gehörenden Datenpunkt weglässt?

Aufgabe 11

(4 Punkte)

- (a) Seien $x_1, y_1, \dots, x_n, y_n \in \mathbb{R}$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ und $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Zeigen Sie:

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$$

und

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y}.$$

- (b) Berechnen Sie die Korrelation der Daten aus Aufgabe 10.
- (c) Was folgt aus b) für die Steigung der zugehörigen Regressionsgeraden?
- (d) Inwieweit ändert sich das Ergebnis aus b), wenn man vor Beginn der Berechnung der Korrelation die Ausgaben pro Student in Dollar und die Arbeitslosenquote in Promille umrechnet? Begründen Sie ihre Antwort.

Aufgabe 12

(4 Punkte)

Gegeben sei eine zweidimensionale Messreihe

$$(x_1, y_1), \dots, (x_n, y_n)$$

vom Umfang n . Anstelle einer Geraden (wie bei der linearen Regression) könnte man analog auch ein Polynom dritten Grades

$$y = a + b \cdot x + c \cdot x^2 + d \cdot x^3$$

durch Minimierung von

$$F(a, b, c, d) := \sum_{i=1}^n (y_i - (a + b \cdot x_i + c \cdot x_i^2 + d \cdot x_i^3))^2$$

an die Daten anpassen. Zeigen Sie (durch Nullsetzen geeigneter Ableitungen), dass die Werte a, b, c, d , für die $F(a, b, c, d)$ minimal wird, Lösungen des linearen Gleichungssystems

$$\begin{aligned} a + b \cdot \frac{1}{n} \sum_{i=1}^n x_i + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^2 + d \cdot \frac{1}{n} \sum_{i=1}^n x_i^3 &= \frac{1}{n} \sum_{i=1}^n y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^n x_i + b \cdot \frac{1}{n} \sum_{i=1}^n x_i^2 + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^3 + d \cdot \frac{1}{n} \sum_{i=1}^n x_i^4 &= \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^n x_i^2 + b \cdot \frac{1}{n} \sum_{i=1}^n x_i^3 + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^4 + d \cdot \frac{1}{n} \sum_{i=1}^n x_i^5 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \cdot y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^n x_i^3 + b \cdot \frac{1}{n} \sum_{i=1}^n x_i^4 + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^5 + d \cdot \frac{1}{n} \sum_{i=1}^n x_i^6 &= \frac{1}{n} \sum_{i=1}^n x_i^3 \cdot y_i \end{aligned}$$

sind.

Dieses Übungsblatt wird im Rahmen der Übungen am 09. bzw. 10.05.2011 besprochen. Ihre Ausarbeitungen geben Sie am 16. bzw. 17.05.2011 in Ihre Übungsgruppe ab. Sie erhalten diese am 23. bzw. 24.05.2011 korrigiert zurück.