

## Aufgaben zur Vorlesung „Werkzeuge der empirischen Forschung“

**Aufg. 14)** (2 P.) (Banknoten, Datei **banknote.dat**)

Erzeugen Sie genau einen Plot, bei dem die Werte von **laenge** in Abhängigkeit von den Werten von **links** dargestellt werden. Die Datenpunkte und die Regressionsgeraden sollen für beide Gruppen verschieden dargestellt werden. Geben Sie die Regressionsgleichungen an (Kommentar und Ausgabe).

**Aufg. 15)** (3 P.) (Sterblichkeit und Wasserhärte)

Führen Sie mit den Variablen **mortal** und **calcium** eine **lineare**, eine **kubische** und eine **glättende Spline**-Regression durch. Geben Sie für die lineare und kubische Regression die Regressionsgleichungen an und plotten Sie anschließend die 3 Regressionskurven jeweils zusammen mit den Beobachtungen (Sie können wählen, ob Sie einen oder drei Plots machen). **Hinweis:** Verwenden Sie  $I(x^3)+I(x^2)$  . . . als Formel für die kubische Regression.

**Aufg. 16)** (3 P.) (Old Faithful Geysir, Datei **geyser.dat**)

Die Datei **geyser.dat** enthält 300 Wartezeiten zwischen aufeinanderfolgenden Ausbrüchen des Old Faithful Geysir im Yellowstone Nationalpark (in min). Führen Sie nichtparametrische Dichteschätzungen für die Wartezeiten durch. Probieren Sie verschiedene Glättungsparameter.

**Aufg. 17)** (3 P.) (Toxaemia in der Schwangerschaft)

- a) (1 P.) Erzeugen Sie eindimensionale Häufigkeitstabellen für die Variablen **hyperten** und **proteinu**, jeweils nach den verschiedenen Ausprägungen der Variablen **social** und **smoking**.
- b) (2 P.) Erzeugen Sie eine gemeinsame Häufigkeitstabelle für die Variablen **hyperten** und **proteinu** und testen Sie diese auf Unabhängigkeit (**Hinweis:** **chisq.test**).

**Aufg. 18)** (9 P.) (Über die Robustheit der Korrelationskoeffizienten, auf Papier)

Es seien jeweils die folgenden  $n = 5$  Beobachtungspaare  $(X, Y)$  gegeben:

X	1	2	3	4	k
Y	1	2	3	4	5

- a) (3 P.) Bestimmen Sie die Pearson-, Spearman-, und Kendall-Korrelationskoeffizienten, jeweils für die Werte  $X_n = k = 5, 10, 100, 1000$  mit der Hand. Hinweis: Überprüfen Sie Ihre Ergebnisse mit Hilfe von R.
- b) (1 P.) Zeigen Sie für den Grenzwert des Pearson-Korrelationskoeffizienten:  
 $\lim_{k \rightarrow \infty} r_{XY} = \frac{1}{\sqrt{2}}$ .
- c) (3 P.) Seien  $(X_i, Y_i)$  beliebig, sowie  $X_{(n)} = k$  und  $X_{(1)} = k'$ . Berechnen Sie die Grenzwerte  $\lim_{k \rightarrow \infty} r_{XY}$  und  $\lim_{k' \rightarrow -\infty} r_{XY}$ .
- d) (1 P.) Welche Grenzwerte ( $k \rightarrow \infty, k' \rightarrow -\infty$ ) gelten für die Spearman-, und Kendall-Korrelationskoeffizienten?
- e) (1 P.) Welche Schlussfolgerungen ziehen Sie für die Robustheit der drei Korrelationskoeffizienten?