

Aufgaben zur Vorlesung

“Werkzeuge der empirischen Forschung”

Aufgabe 11 (Fortsetzung), h) (3 P.) 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!

Aufg. 12) (Toxaemia in der Schwangerschaft)

- (2 P.) Erzeugen Sie eindimensionale Häufigkeitstabellen für die Variablen 'hyperten' und 'protein', jeweils nach den verschiedenen Ausprägungen der Variablen 'social' und 'smoking'.
- (2 P.) Erzeugen Sie eine gemeinsame Häufigkeitstabelle für die Variablen 'hyperten' und 'protein', und testen Sie diese auf Unabhängigkeit. Hinweis: Für den Unabhängigkeitstest verwenden Sie die Option `Chisq`.
- (+2 P.) Bestehen Abhängigkeiten zwischen Rauchgewohnheiten und sozialem Status einerseits und Bluthochdruck und Proteinuria andererseits?

Aufg. 13) (Robustheit der Korrelationskoeffizienten)

Seien jeweils die folgenden $n = 5$ Beobachtungspaare (X, Y) gegeben,

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

- (2 P.) Bestimmen Sie die Pearson-, Spearman-, und Kendall- Korrelationskoeffizienten, jeweils für die Werte $X_n = k = 5, 10, 100, 1000$
- (2 P.) Zeigen Sie für den den Pearson-Korrelationskoeffizienten den Grenzwert $\lim_{k \rightarrow \infty} r_{XY} = \frac{1}{\sqrt{2}}$.
- (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 $\lim_{k' \rightarrow -\infty} r_{XY}$.
- (1 P) Welche Grenzwerte gelten für die Spearman-, und Kendall- Korrelationskoeffizienten?
- (2 P.) Welche Schlussfolgerungen ziehen Sie für die Robustheit der drei Korrelationskoeffizienten?

Aufg. 14) (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.