Lennart Oelschläger ✉
14-tägig, Mo, 16:15 – 17:45, T2-213

Aufgabenblatt 3

Aufgabe 1 (Multiple lineare Regression)

a) Wie wird das multiple lineare Regressionsmodell definiert, und wie können die Modellparameter geschätzt werden?

b) Welche Eigenschaften hat der Kleinste-Quadrate Schätzer, und welche Voraussetzungen müssen dafür erfüllt sein?

c) Bitte prognostizieren Sie den y Wert für $x_1 = 6$ und $x_2 = -1$, gegeben die Daten

$$
\begin{aligned}
x_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \\
x_2 &= \begin{pmatrix} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 & 5 & 5 \end{pmatrix}, \\
y &= \begin{pmatrix} 1 & 3 & 1 & 5 & 2 & -3 & -3 & -1 & -2 & -2 \end{pmatrix}.
\end{aligned}
$$

Aufgabe 2 (Trendmodellierung durch polynomiale Regression)

a) Sie finden im Lernraum der PÜ den Datensatz `hermannslauf_frauen.csv` mit den Hermannslaufbestzeiten der Frauen. Bitte erstellen Sie eine Grafik der Bestzeiten in Minuten. Können Sie Elemente des Komponentenmodells erkennen und interpretieren?

b) Bitte passen Sie polynomiale Trendmodelle verschiedenen Grades mittels linearer Regression an die Daten an. Wie würden Sie den Polynomgrad wählen und warum?

c) Im Jahr 2005 wurde die Laufstrecke um 500 Meter verlängert. Modellieren Sie an dieser Stelle einen Strukturbruch in der Zeitreihe und testen Sie auf statistische Signifikanz.

Aufgabe 3 (Trendbereinigung mittels variate-difference Methode)

Betrachten Sie die folgenden drei Zeitreihen für $t = 1, \dots, T$ mit Restkomponente u_t:

- $a_t = 3 + u_t$
- $b_t = a_t + 0.4t$
- $c_t = b_t + 0.3t^2$

a) Bitte berechnen Sie jeweils die erste und zweite Differenz der drei Zeitreihen.

b) Die Restkomponente u_t sei eine standardnormalverteilte und unabhängige Zufallsvariable. Welchen Erwartungswert und Varianz haben die Zeitreihen sowie ihre erste und zweite Differenz für gegebenes t jeweils?

c) Simulieren Sie die Zeitreihe c_t bis $T = 100$ und berechnen Sie die erste und zweite Differenz. Visualisieren Sie anschließend das Ergebnis.