

Blatt 6

Vorhersage

Aufgabe 6.1 Vorhersageintervall

Wir betrachten das lineare Modell

$$\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} X \\ x_0^t \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix}, \quad X \in \mathbf{R}^{n \times k}, \quad \beta \in \mathbf{R}^k, \quad x_0 \in \mathbf{R}^k, \quad \mathbf{Cov} \begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix} = \sigma^2 \begin{pmatrix} \mathbf{I}_n & 0_n \\ 0_n^t & 1 \end{pmatrix}.$$

Zeigen Sie, dass zu $\alpha \in (0, 1)$ durch

$$[x_0^t \hat{\beta} - \delta, x_0^t \hat{\beta} + \delta] \quad \text{mit} \quad \delta = t_{n-k, 1-\alpha/2} \sqrt{\widehat{\sigma^2} (x_0^t (X^t X)^{-1} x_0 + 1)}$$

ein $100(1 - \alpha)\%$ -Vorhersageintervall für y_0 gegeben ist.

Aufgabe 6.2 Bremsweg

Laden Sie den Datensatz `cars` aus `library(stats)` und machen Sie sich mit ihm vertraut. Modellieren Sie den Bremsweg s in Abhängigkeit der Geschwindigkeit v mittels eines GLM. Beachten Sie hierbei, dass ein quadratischer Zusammenhang besteht.

- Rechnen Sie die im Datensatz gegebenen Werte in die metrischen Einheiten km/h sowie m um und arbeiten Sie mit diesen Daten weiter
- Bestimmen Sie zum Signifikanzniveau $\alpha = 0.05$ ein punktwises sowie ein simultanes (Scheffé) Koinzidenzband für die Regressionsparabel. Berechnen Sie des Weiteren ein (punktweises) Prädiktionsband (vgl. Aufgabe 6.1) für den Bremsweg. Stellen Sie alles zusammen graphisch dar.
- Geben Sie die Verzögerungskonstante a in m/s^2 an.
- Nach einem Verkehrsunfall soll untersucht werden, ob der Fahrer sich an die Höchstgeschwindigkeit von $30 km/h$ gehalten hat. Der von den ermittelnden Beamten gemessene Bremsweg beträgt $28,5 m$. War der Fahrer zu schnell?

Aufgabe 6.3 Untersuchung der Verteilung

Laden Sie den Datensatz `Heights` aus `library(alr4)`. Bestimmen Sie für das lineare Modell $y = \beta_1 + x\beta_2 + \varepsilon$ die Residuen, wobei y die Körpergrößen der Töchter und x die Körpergrößen der Mütter enthält. Arbeiten Sie dazu mit einem Stichprobenumfang von 10, 20, 40 und 100 sowie dem gesamten Datensatz.

- Stellen Sie die Residuen in einem Histogramm und mittels eines Box-Whisker-Plots dar.
- Vergleichen Sie mittels eines Q-Q-Plots Verteilung der Residuen mit der Standardnormalverteilung, einer Normalverteilung mit Erwartungswert $\mu = 9$ und Standardabweichung $\sigma = 2$, einer t-Verteilung mit Freiheitsgrad 2, einer Binomialverteilung mit $p = 1/5$ und $n = 100$ sowie einer Exponentialverteilung mit Rate $\lambda = 1/10$.

(c) Treffen Sie eine Aussage zur Verteilung der Residuen!

Nächste Übung: Donnerstag 15. Juni 2017 um 13:00 Uhr im Sr C 115