

Aufgabe 1:

Betrachten Sie wieder den Datensatz `teengamb` (Blatt 3/Aufgabe 1) und das Modell aus Blatt 6/Aufgabe 4a.

- (a) In der Vorlesung haben Sie Konfidenzellipse kennengelernt, welche, ähnlich zu Konfidenzintervallen im univariaten Fall, zur Darstellung der gemeinsamen Unsicherheit mehrerer Parameter verwendet werden können. Erstellen Sie eine Graphik, welche die Konfidenzellipse für die Parameter der Variablen `status` und `verbal` zeigt.
Hinweis: Betrachten Sie die Funktion `?confidenceEllipse` aus dem `car` package.
- (b) Zeichnen Sie weiterhin die univariaten Konfidenzintervalle sowie die dazugehörigen Bonferroni-Konfidenzintervalle für jeden Parameter ein; tragen Sie auch den Ursprung ab. Kann man aus der Graphik Aussagen hinsichtlich der Hypothese $\beta_{status} = \beta_{verbal} = 0$ ableiten?
- (c) Vergleichen Sie die naiven p-Werte mit den nach Bonferroni und dem Maximum-Prinzip adjustierten p-Werten!

Aufgabe 2:

Die Datei `uwz.Rdata` enthält stündliche Daten über die PM_{10} -Konzentrationen, die an zwei Stationen in München gemessen wurden (Prinzregentenstraße und Johanneskirchen).

- (a) Untersuchen Sie den Effekt der Einführung der Umweltzone auf die (log-transformierte) PM_{10} -Konzentration an der Prinzregentenstraße (Y_{Prin}). Nehmen Sie die Feinstaub-Messungen der Station Johanneskirchen, die außerhalb der Umweltzone liegt mit in das Modell auf, um für die städtische Hintergrundkonzentration zu korrigieren. Geben Sie auch die Modellgleichung an.
- (b) Die Variable `cosWD` gibt an, inwiefern der Wind aus der Richtung der Referenzstation in Johanneskirchen kommt. Untersuchen Sie, inwieweit der Effekt der Referenzstation mit der Windrichtung zusammenhängt.
- (c) Stellen Sie die Wochenverläufe der log-transformierten Messungen an der Prinzregentenstraße graphisch dar. Verwenden Sie hierzu die Variable `stwo`, die die Nummerierung der Stunden innerhalb einer Woche anzeigt.
- (d) Rechnen Sie eine polynomiale Regression 1., 2., ..., bis 8. Grades für die (log-transformierte) PM_{10} -Konzentration an der Prinzregentenstraße über die Zeit innerhalb einer Woche (t)

$$Y_{Prin} = \sum_{i=0}^g \beta_i t^i + \varepsilon \quad \text{für } g = 1, \dots, 8 \quad (1)$$

und stellen Sie die Anpassung an die Daten jeweils graphisch dar. Was schließen Sie daraus?

- (e) Betrachten Sie jetzt ein Modell mit Regressionssplines g -ten Grades

$$Y_{Prin} = \sum_{i=0}^g \beta_i t^i + \sum_{k=1}^K \gamma_k (t - x_k)_+^g + \varepsilon \quad (2)$$

mit K Knotenpunkten $x_1 < \dots < x_K$ im Wertebereich von t , wobei $z_+ = \max(z, 0)$ und $z_+^g = (z_+)^g$.

- i) Interpretieren Sie die β - und γ -Koeffizienten.
 - ii) Halten Sie $K = 3$ fest und wählen Sie drei äquidistante Knotenpunkte. Schätzen und zeichnen Sie die Modelle für $g = 1, \dots, 8$. Diskutieren Sie das Ergebnis.
 - iii) Halten Sie nun $g = 3$ fest. Schätzen und zeichnen Sie nun die Modelle für $K = 3, 5, 10, 30$ mit jeweils äquidistanten Knoten (d.h. die Knoten teilen das Intervall vom Minimum bis zum Maximum von t in $K + 1$ gleichlange Teilintervalle). Diskutieren Sie das Ergebnis.
- (f) Berücksichtigen Sie nun den zeitlichen Verlauf im Modell aus Teilaufgabe a), indem Sie die Wochenstunde geeignet modellieren. Geben Sie für dieses Modell auch die Modellgleichung an. Ist der Effekt über die Woche hinweg konstant? Schätzen Sie auch den Gesamt-Effekt der Einführung der Umweltzone.
- (g) Überprüfen Sie die Modellannahmen bzgl. der Autokorrelation des Störterms graphisch sowie mit einem geeigneten Test.