

Aufgabe 1:

Betrachten Sie die Münchner Feinstaubmessungen aus Blatt 7/Aufgabe 2.

- Passen Sie ein allgemeines lineares Modell auf der Basis des Modells aus Blatt 7/Aufgabe 2d an. Verwenden Sie zur Schätzung die Funktion `gls` aus dem Package `nlme` und gehen Sie von einem autokorrelierten Störterm erster Ordnung aus. Gehen Sie außerdem davon aus, dass die Störterme verschiedener Wochen nicht korreliert sind, um die Rechenzeit des Modells zu verringern.
- Sind die Modell-Annahmen bzgl. der Autokorrelation erfüllt?
- Ist der Umweltzonen-Effekt signifikant von Null verschieden? Unterscheidet sich der Wochenverlauf der Feinstaubkonzentration vor der Einführung der Umweltzone signifikant vom Verlauf nach Einführung der Umweltzone?

Aufgabe 2:

Der Datensatz `zufriedenheit.txt` enthält (simulierte) Daten zur Lebenszufriedenheit von 137 Personen aus den Jahren 1986 bis 2006, wobei für jede Person Angaben aus mehreren aufeinanderfolgenden Jahren vorliegen. Neben der Zielgröße *wellbeing* enthält der Datensatz die Einflussgrößen Alter (*age*, auch quadratisch und kubisch), Jahr (*year*), einen Indikator für Westdeutschland (*west*) und einen Indikator für Frauen (*female*).

- Stellen Sie die Daten graphisch dar. Zeigen Sie auch (beispielhaft für die ersten neun Personen des Datensatzes), welche Gestalt die individuellen Verläufe haben.
- Rechnen Sie ein lineares Modell mit sämtlichen Einflussgrößen, wobei das Alter mit einem Regressionsspline 3. Ordnung ($g = 3$) und $K = 3$ äquidistanten Knoten modelliert werden soll (vgl. Blatt 7, Aufgabe 2).
- Die Residuen verschiedener Personen streuen unterschiedlich stark. Vergleichen Sie folgende drei Modellierungsvarianten, um der Heteroskedastie gerecht zu werden: i) gewichtete KQ-Schätzung, ii) verallgemeinerte KQ-Methode, iii) REML-Schätzung. Verwenden Sie die empirische Kovarianzmatrix der Residuen des Modells aus Teilaufgabe b) in (i) und (ii). Vergleichen Sie diese Gewichte mit den simultan geschätzten in (iii) und die Parameterschätzer der Modelle.
- Berücksichtigen Sie nun, dass zu jeder Person mehrere Messungen vorliegen. Verwenden Sie zur Schätzung wieder die Funktion `gls` aus dem Package `nlme`. Gehen Sie für die Fehler von einem Autoregressiven Prozess 1. Ordnung aus. Beschreiben Sie auch die Struktur der Varianz-Kovarianz-Matrix, die Sie für die Störterme annehmen.
- Bisher wurden nur die personen-spezifische Autokorrelation und Streuung der Residuen im Modell berücksichtigt. Allerdings hängt auch der Mittelwert der Residuen von der Person ab. Veranschaulichen Sie sich dies grafisch. Dafür wird nun ein lineares gemischtes Modell mit der function `lme` geschätzt. Geben Sie für dieses Modell die Modellgleichung und die Annahmen an und interpretieren Sie die geschätzten Parameter.