



Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse, Dr. Martin Theus
Institut für Mathematik
Universität Augsburg



Explorative Datenanalyse mit R und iPlots

Übungsblatt 4

Bearbeitung: Mi. 1. Juni 2005, 10.15 Uhr.

Alle Aufgaben können im Cip Pool bearbeitet werden.

1. Erstellen Sie für die *Mietspiegel* Daten ein möglichst gutes, statistisches Modell. Überprüfen Sie, in wie weit dieses Modell die Fragestellungen aus der Übung beantworten kann. Verwenden Sie dabei die Variable *Baujahr* sowohl als geeigneten Faktor, als auch in der rohen Form.
Was kann durch das Modell mehr, bzw. besser beantwortet werden, als durch die bisherigen, explorativen Analysen?
2. Erstellen Sie für die folgenden Probleme reguläre Ausdrücke:
 - (a) In einem Datensatz, der in Tabulator getrennten Spalten vorliegt, soll die dritte Spalte gelöscht werden.
 - (b) Ein Datensatz liegt in deutscher, kaufmännischer Notation vor, d.h. der Dezimalpunkt ist “;”, und das Tausender Trennzeichen ist “.”. Konvertieren Sie diesen Datensatz in wissenschaftliche Notation.
 - (c) In einer Datei sind durch einen Fehler einige Zeilen dupliziert worden. Schreiben Sie einen Ausdruck, der identische Zeilen auf je ein Vorkommen reduziert.

Bemerkung: Zur Lösung nötige Editoren, welche reguläre Ausdrücke beherrschen, finden Sie auf der Internetseite zur Vorlesung.

3. Betrachten Sie nun die *Apotheken* Daten, und machen Sie sich mit dem Hintergrund des Problems vertraut. Formulieren Sie potentielle Fragestellungen, die sich aus diesem Datensatz ergeben.