



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



## Explorative Datenanalyse mit R und iPlots

### Übungsblatt 5

**Bearbeitung:** Mi. 15. Juni 2005, 10.15 Uhr.

Alle Aufgaben können im Cip Pool bearbeitet werden.

1. Erstellen Sie für die **Apotheken** Daten ein möglichst gutes Klassifikationsmodell. Verwenden Sie dabei sowohl logistische Regression als auch Klassifikationsbäume.  
Diskutieren Sie die Ergebnisse im Hinblick auf Klassifikationsgüte und Interpretierbarkeit.
2. Wir betrachten erneut die **Geburtsgewicht** Daten
  - (a) Erstellen Sie einen iPlot für das Geburtsgewicht in Abhängigkeit von der Schwangerschaftsdauer.
  - (b) Selektieren Sie die möglichen Ausreißer, und fragen Sie deren Id ab.
  - (c) Zeichnen Sie nun je für die Einflußfaktoren "smoke" und "parity" eine lowess-Glättung ein, welche die Ausreißer nicht berücksichtigt, und färben sie die 5% größten Residuen ein.

Welchen Einfluß haben die extremen Werte auf die Schätzung der Mittelwertfunktion?

3. Betrachten Sie nun die **Hurricane** Daten. Machen Sie sich mit dem zu untersuchenden Problem vertraut und untersuchen Sie die vorliegenden Variablen auf Abhängigkeiten.