



Prof. Dr. Antony Unwin, Dr. Ali Ünlü
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>



Stochastik IV – Multivariate statistische Verfahren

Übungsblatt 1

Bearbeitung: Dienstag 24. April 2007, 12.15 - 13.45 Uhr
Raum 2001 T

1. Betrachtet werden interaktive Optionen in Mondrian.
Erstellen Sie eine Tabelle, in der Sie eintragen, welche interaktiven Optionen (Spalten) in welchen Graphiken (Zeilen) anwendbar sind. (Die Reference Card – unter Menü ‘Help’ – wird sicherlich behilflich sein; ist sie aber vollständig?)
2. Betrachtet wird der Datensatz **EcoTestAutos**.
 - (a) Laden Sie den Datensatz in Mondrian.
 - (b) Stellen Sie die Daten in geeigneten Graphiken dar, so dass Sie eine Übersicht des Datensatzes bekommen. Welche Graphiken können hierfür eingesetzt werden?
 - (c) Erstellen Sie ein Histogramm für die Variable ‘FuelCons’. Welche Klassenbreite erscheint Ihnen am sinnvollsten, um die Verteilung gut darzustellen. Schalten Sie dann um zur Darstellung eines Spinogramms und vergleichen Sie die Untermengen für den Motorentyp Diesel. Wie würden Sie die Auto-Modelle mit Dieselmotoren, die hohe Werte der ‘FuelCons’ Variablen aufweisen, anhand der Eco-Variablen beschreiben?
3. Betrachtet wird der Datensatz **EcoTestAutos**.
 - (a) Stellen Sie die stetigen Variablen in einem ‘Parallel Coordinates Plot’ dar. Welche Werte würden Sie als Ausreisser ansehen?
 - (b) Gibt es im Datensatz zu den stetigen Variablen zweidimensionale Ausreisser, die keine eindimensionalen Ausreisser sind?
 - (c) Gibt es eine Assoziation zwischen den Variablen Typ und Schadstoffklasse? Mit welchem Plot würden Sie das untersuchen? Hilft es, die Reihenfolge der Kategorien zu ändern?
4. Betrachtet wird die Selektion, ein Grundelement interaktiver statistischer Graphik, anhand des Datensatzes **EcoTestAutos** in Mondrian und im mengentheoretischen Kontext.
 - (a) Zeigen Sie anhand eines Beispiels zweier Selektionssequenzen in Mondrian, dass für drei Selektionsmengen A , B und C im Allgemeinen $A \cup (B \cap C)$ und $(A \cup B) \cap C$ zu verschiedenen Auswahlen von statistischen Objekten führen. Welche Selektionswerkzeuge haben Sie benutzt?
 - (b) Geben Sie für den vorherigen Sachverhalt ein mengentheoretisches Beispiel an, formulieren Sie die XOR und NOT Selektionsmodi in mengentheoretischer Schreibweise und drücken Sie letztere nur unter Benutzung von \cap (AND), \cup (OR) und c (KOMPLEMENT) aus.
5.
 - (a) Wieviele ‘Mosaic Plots’ gibt es bei drei kategoriellen Variablen mit jeweils n_1, n_2 bzw. n_3 Kategorien, bei jeweils vorgegebenen fixen Reihenfolgen der Kategorien der einzelnen Variablen?
 - (b) Wieviele ‘Mosaic Plots’ gibt es, wenn die jeweiligen Reihenfolgen der Kategorien frei gewählt werden dürfen?