



Statistik III: Multivariate Stat. Verfahren

Übungsblatt 4

Bearbeitung: Do. 3.6.2004, 16.00 Uhr.

Alle Aufgaben können im Mac-Cip Pool bearbeitet werden.

1. Berechnen Sie die Hauptkomponenten für die 8 Fettsäuren der **Olive Oils** Daten, und betrachten Sie diese in `ggobi`. Untersuchen Sie nun die Daten der Hauptkomponenten wie in Aufgabe 2(a) auf Blatt 2. Was hat sich geändert, und was kann man über die Güte der Zuordnung sagen?
(Hinweis: Da `ggobi` nur numerische Daten lesen kann, die durch Leerzeichen getrennt sind, müssen beim Exportieren der Daten aus R mit `write.table` die Parameter entsprechend gesetzt werden. Die Datendatei muss auf `.dat` enden, die Variablenamen müssen zeilenweise in einer Datei mit der Endung `.col` eingetragen werden.)
2. Berechnen Sie die Hauptkomponenten der **Crabs** Daten.
 - (a) Wie viele Hauptkomponenten erachten Sie für notwendig, um die Variabilität der Daten zu beschreiben?
 - (b) Plotten Sie die Hauptkomponenten in einer Scatterplotmatrix, und versehen Sie dabei die 4 Gruppen mit vier verschiedenen Farben.
 - i. Wie lassen sich die Gruppen trennen – erklären Sie warum?
 - ii. Beschreiben Sie dabei die Unterschiede zwischen den Rohdaten und den normierten Daten in der PCA.
3. Führen Sie in R eine Faktoren Analyse für die **Zehnkampf Daten** durch.
 - (a) Wie viele Faktoren würden Sie a-priori erwarten?
 - (b) Erstellen Sie mit `factanal` ein Modell mit 2 bzw. 3 Faktoren, und vergleichen Sie jeweils die Residualmatrix.
 - (c) Stellen Sie für beide Modelle die Ladungen graphisch dar (ggf. rotiert und unrotiert) und interpretieren Sie die Ergebnisse. Bestätigen sich Ihre Erwartungen aus (a)?
4. Betrachtet werden die **Börsendaten** aus der Vorlesung.
Testen Sie mit dem Bartlett-Test nach der Vorlesung ob das 2-Faktormodell (MLE) signifikant ist, und formulieren Sie das Testergebnis. Ergibt sich je nach Korrekturfaktor n oder $(n - 1 - (2p + 4k + 5)/6)$ ein anderes Ergebnis?
5. Betrachtet werden die Daten **Swiss**, die im R base Package verfügbar sind.
 - (a) Bilden Sie die beiden Gruppen `soc` der "sozialen" Variablen aus `Agriculture`, `Education`, `Catholic` und `phy` der "physischen" Variablen `Fertility`, `Examination`, `Infant.Mortality`.
 - (b) Berechnen Sie die Korrelation des gesamten Daten. Wo finden sich die größten Korrelationen?
 - (c) Berechnen Sie nun die Kanonischen Korrelationen zwischen `soc` und `phy`. Wie interpretieren Sie das Ergebnis?
 - (d) Stellen Sie die Rohdaten und die Kanonischen Variablen jeweils in einer Scatterplot Matrix gegenüber.