

# Multivariate Methoden: Übungsblatt 2

## Aufgabe 1

Der Datensatz BOSTON.DTA (Quelle: *Härdle, W./L. Simar: Applied Multivariate Analysis*) enthält die durchschnittlichen Preise von Wohneigentum (= *VALUEOFHOME*) in 506 Distrikten der Stadt Boston zusammen mit einer Reihe von möglicherweise mit diesen Preisen zusammenhängenden Größen (*CRIMRATE* = Kriminalitätsrate, *RESIDENTIALZONE* = Gebietsanteil von Wohnsiedlungen, *NONRETAIL* = Gebietsanteil von Gewerbebetrieben ohne Einzelhandel, *RIVER* = Indikator für Flußnähe, *NOCONCENTRATION* = Stickoxid-Konzentration, *NOOFROOMS* = durchschnittliche Zimmeranzahl, *AGE* = Anteil Wohnungen mit Baujahr vor 1940, *DISTANCE* = gewichtete Entfernungen zu fünf Bostoner Wirtschaftszentren, *ACCESSTOHIGHWAY* = Verkehrsanbindung zum Highway, *TAXRATE* = Steuersatz Grundsteuer, *PUPILTEACHER* = Verhältnis Schüler/Lehrer, *BLACKPEOPLE* = Maß für Bevölkerungsanteil von Farbigen in nicht überwiegend von Farbigen bewohnten Distrikten, *LOWERSTATUS* = Bevölkerungsanteil unterprivilegierter Schichten).

- a) Führen Sie eine Kleinste-Quadrate Regression zur Erklärung der Wohneigentumspreise durch die genannten Faktoren durch. Interpretieren Sie die Ergebnisse. Welche Regressoren sind signifikant von Null verschieden? Angenommen, die Kriminalitätsrate in einem Distrikt würde durch verstärkte Verbrechensbekämpfung plötzlich von 10 Prozent pro Einwohner auf 0 Prozent sinken. Lässt sich aus der Regression vorhersagen, wie dies die durchschnittlichen Wohneigentumspreise beeinflussen würde?
- b) Führen Sie einen F-Test der Hypothese durch, dass alle Regressionskoeffizienten außer der Konstanten gleich Null sind.
- c) Testen Sie die Hypothese, dass die Koeffizienten der Regressoren *NONRETAIL* und *AGE* gemeinsam gleich Null sind. Führen Sie einen F-Test und einen  $\chi^2$ -Test durch. Verwenden Sie auch den STATA Befehl zur Durchführung eines F-Tests `test`.

## Aufgabe 2

Führen Sie eine kanonische Korrelationsanalyse für den Datensatz CARS.DTA aus, um den Zusammenhang zwischen den Kundenbewertungen der Eigenschaften *PRICE* (=Preis) und *VALUE* (=Wertbeständigkeit) auf der einen, und den Bewertungen der Eigenschaften *ECONOMY* (= Sparsamkeit), *SERVICE* (= Service), *DESIGN* (= Design), *SPORT* (= Sportlichkeit), *SAFETY* (= Sicherheit) und *EASYINESS* (= Handhabbarkeit) auf der anderen Seite zu untersuchen. In wie weit kann man sagen, dass die eine Gruppe von Variablen ähnliche Information enthält wie die andere Gruppe? Wie hängt die Sportlichkeit eines Automodells mit der Wahrnehmung des Preises und der Wertbeständigkeit zusammen? (Hinweis: Verwenden Sie den Befehl `canon`.)