

# Fragenkatalog

## 2. Test, Multivariate Verfahren 2

### SBWL Tourismusanalyse und Freizeitmarketing, VK4

Achim Zeileis

Aus diesem Fragenkatalog werden die Theoriefragen im 2. Test des VK4 gestellt. Die Fragen sind auf Basis der LV-Unterlagen zu beantworten. Zusätzlich zu den Theoriefragen wird der Test Anwendungsbeispiele in R enthalten. In diesen Anwendungsaufgaben ist jeweils R-Output angegeben und muß anhand von Fragen interpretiert werden. Die R-Aufgaben sind immer an die Probleme angelehnt, die in den Übungen und Tutorien diskutiert wurden.

## 1 Lineare Modelle

- Welche Verfahren zur Schätzung von linearen Modellen gibt es? Was ist jeweils die grundsätzliche Idee?
- Gegeben seien  $n$  Beobachtungen  $y_1, \dots, y_n$ , für die ein Mittelwert  $\mu$  mit Hilfe der Methoden der kleinsten Quadrate bestimmt werden soll. Wie lautet die Fehlerquadratsumme  $RSS(\mu)$  und wie kann mit Hilfe dieser der Parameterschätzer bestimmt werden?
- Gegeben seien  $n$  unabhängige Beobachtungen  $y_1, \dots, y_n$  mit Dichte  $f(y|\beta)$ . Mit Hilfe von Maximum Likelihood soll ein Schätzer für  $\beta$  gefunden werden: Wie lautet die log-Likelihood  $\log L(\beta)$  und wie kann mit Hilfe dieser der Parameterschätzer bestimmt werden?
- Wie können qualitative Variablen in das lineare Modell aufgenommen werden und welche Rolle spielen dabei Kontraste? Illustrieren Sie das Problem anhand einer Einfach-Varianzanalyse.

## 2 Modellselektion

- Welche zwei grundsätzlich verschiedenen Methoden zum Vergleich zweier Modelle gibt es? Erläutern Sie kurz die Idee für jeweils einen Repräsentanten beider Methoden.
- Wie ist das AIC bzw. BIC definiert? Was ist die Motivation für diese Informationskriterien und wie können Sie zum Vergleich von Modellen eingesetzt werden?
- Welche Strategien zur schrittweisen Modellwahl gibt es? Erläutern Sie für jede dieser Strategien, wie dabei vorgegangen wird, um ein geeignetes Modell zu finden.

## 3 Verallgemeinerte lineare Modelle

- Wie lautet die Modellgleichung im GLM? Erklären Sie kurz die Rolle der einzelnen Symbole.
- Wie kann die Link-Funktion in der logistischen Regression interpretiert werden? Welche Veränderung in der abhängigen Variablen tritt ein, wenn sich ein Regressor  $x_j$  mit Koeffizient  $\beta_j$  um eine Einheit ändert und alle anderen Regressoren gleich bleiben?
- Wie sind die Pearson-Residuen im GLM definiert? Erläutern Sie kurz die Bedeutung der Symbole. Geben Sie außerdem die konkrete Formel fuer das normal/binomial/poisson-verteilte Modell an.