

# Übung 8

Rudolf Dutter & Matthias Templ  
Computerstatistik (107.258) WS 2007  
10. Dezember 2007



## 1 Daten

Für diese Übung werden wiederum die EU-SILC Daten aus dem Jahre 2004 verwendet.

## 2 Übungsbeispiele

(Modell 1): Schätzen und testen Sie ein lineares Modell der Variablen Brutto-Einkommen (abhängige Variable, `pek_g`) mit den (unabhängigen) Variablen, wie sie in der folgenden Tabelle aufgeführt sind.

Nummer	Name	Variable	Anzahl der Möglichkeiten
9	Geschlecht	<code>sex</code>	2
12	Gegenwaertige Beschaeftigungssituation	<code>P001000</code>	9
22	Zahl der Beschaeftigten in letzter Betriebsstaette	<code>P011000</code>	15
57	Gesundheitszustand	<code>P102000</code>	7
70	Familienstand	<code>P114000</code>	6
74	Hoechster Bildungsabschluss	<code>P118000</code>	11
79	Zufriedenheit mit persönlichen Einkommen	<code>P120040</code>	9
121	Bundesland	<code>bundesld</code>	9
126	Geburtsjahr	<code>r002000</code>	82

Für die weitere Bearbeitung sollten die Variablen mit den Nummern 22, 74 und 126 in maximal 10 sinnvolle Gruppen zusammengefasst werden. Z.B. bei „Zahl der Beschäftigten in letzter Betriebsstätte“ alle Werte mit Anzahl = 1 bis 10 zu einer Gruppe zusammenfassen und „Keine Angabe“ und „Nicht zutreffend (P008000  $\leq 0$  oder P008000  $\geq 91$ )“ weglassen. Bei „Höchster Bildungsabschluss“ sollten max. 4 Klassen gebildet werden: „Keinen Abschluss“, „Pflichtschulabschluss“, „Abschluss einer Mittelschule“, „Universitätsabschluss“. Die Geburtsjahre sollten auch in nur 3 Klassen geteilt werden.<sup>1</sup>

Bsp 21 Finden Sie ein entsprechendes Modell der obigen Vorgabe, so allgemein wie möglich und testen Sie die einzelnen Hypothesen.

**Lösung:** Ein vollständiges Modell mit Wechselwirkungen wird aus rechentechnischen Schwierigkeiten nicht möglich sein. Versuchen Sie deshalb zunächst ein Modell mit allen Variablen, aber ohne Wechselwirkungen. Interpretieren Sie Ihr Ergebnis!

Bsp 22 Verwerfen Sie jene Variablen, deren geschätzter Einfluss einen p-Wert größer als 1% aufweist, schätzen Sie das gleiche, jedoch reduzierte Modell und versuchen Sie nun auch die Wechselwirkungen. Interpretieren Sie Ihr Ergebnis!

## Bemerkung wegen Bsp. 19 und 21

Die Reihenfolge der angegebenen Faktoren beeinflusst das Ergebnis: Dies stimmt nicht bei balancierten Versuchsplänen (d.h. in jeder Faktorstufenkombination gibt es gleich viele Beobachtungen). Bei nicht-balancierten Versuchsplänen wie in unserem Fall funktioniert die Orthogonalzerlegung nicht exakt und die Resultate sind

<sup>1</sup>Für die Umkodierung dieser Variablen gibt es mehrere Möglichkeiten. Eine Möglichkeit ist z.B. die Variable P118000 (der Klasse factor) in die Klasse `integer` „umzuwandeln“ (mit `as.integer()`; Bemerkung: intern werden Faktoren ohnehin als integers gespeichert.) und die Funktion `ifelse()` anzuwenden.

nur approximativ (siehe auch `help(aov)`). Macht man aus unserem Versuchsplan einen balancierten, indem man die Anzahl der Beobachtungen auf das Minimum der Wiederholungen pro 'Zelle' beschränkt (was programmier-technisch nicht ganz einfach ist - Fleißaufgabe), so können natürlich die Faktoren in der Reihenfolge beliebig vertauscht werden.

### 3 Abgabe

Bitte senden Sie die Ausarbeitung in Form eines *pdf*-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Listing des Programmkodes (Funktion) an Ihren jeweiligen Gruppenleiter bis 8. Januar 2008

r.dutter@tuwien.ac.at  
oder  
m.templ@tuwien.ac.at

Bitte den Namen des *pdf*-File wieder folgendermaßen:

**name\_UE\_8.pdf**

wobei 'name' für den Familiennamen steht.

Bitte beachten Sie, dass Sie zusätzlich Ihre gelösten Beispiele für die Übung ankreuzen müssen.