

Übung 10

Rudolf Dutter & Matthias Templ
Computerstatistik (107.258) WS 2007
16. Jänner 2008



1 Daten

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

2 Übungsbeispiele

Für bestimmte Fragestellungen ist es wichtig einzelne Einkommenskomponenten möglichst gut durch ein Modell zu beschreiben.

Bsp 26 Verwenden Sie zur Schätzung des Unselbständigeneinkommen (`py010n`) nur jene Beobachtungen mit mehr als 0 Euro Unselbständigeneinkommen und nur jene Beobachtungen, welche in Variable `py010n_i` als “nicht imputiert“ ausgewiesen sind.

Schätzen Sie ein lineares Regressionsmodell mit `py010n` als abhängige und `pek_g` (Brutto-Einkommen) als unabhängige Variable.

Interpretieren Sie Ihre Ergebnis.

Bsp 27 Schätzen Sie, mit den ausgewählten Beobachtungen aus Beispiel 26, ein lineares Regressionsmodell mit `log(py010n)` als abhängige Variable und `log(pek_g+0.01)` (logarithmiertes Brutto-Einkommen), `age`, `P118000` (Bildung), `P033000` (Zahl der erwerbstätigen Jahre), `P014000` (Berufliche Funktion) und `sex` als unabhängige Variablen. Berücksichtigen Sie Wechselwirkungen bezüglich der Variablen `P014000` bezüglich allen oben aufgezählten unabhängigen Variablen.

Dabei sollten Sie folgende Variablen zuerst umkodieren:

- `age`, wie in Übung 2. Als Klasse `numeric` betrachtet.
- `P11800`, wie in Übung 8.
- `P033000`, “Nicht zutreffend (`P001000 >= 3`)“ als 0. Als Klasse `numeric`.
- `P014000`, mit Ausprägungen `Beamter`, `Lehrling`, `Angestellter`, `Selbstständig`, `Sonstige`.

Bsp 28 Plotten Sie die geschätzten standardisierten Residuen $\hat{\epsilon}$ (siehe Übung 9) gegen die geschätzten Werte \hat{y} aus Beispiel 27. Dazu können Sie auch die Standard-Plot-Methode für Objekte der Klasse “`lm`“ verwenden (`plot(lmobj)`).

Interpretieren Sie Ihr Ergebnis. Ist in diesem Plot bereits ersichtlich, ob Modellvoraussetzungen verletzt sind? Welche Modellvoraussetzungen scheinen erfüllt zu sein?

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 Programmcodes (Funktion) an Ihren jeweiligen Gruppenleiter bis 22. Januar 2008

r.dutter@tuwien.ac.at

oder
m.templ@tuwien.ac.at

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

name_UE_10.pdf

wobei 'name' für den Familiennamen steht.

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