

# Übung 3

Rudolf Dutter & Matthias Templ  
Computerstatistik (107.258) WS 2007  
22. Oktober 2007



## 1 Daten

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

## 2 Übungsbeispiele

Bsp 9 Die meisten Kenngrößen werden für die EU-SILC Daten auf Haushaltsebene berechnet und so auch Konfidenzintervalle für diverse Parameter bezüglich der Einkommenskomponenten.

- a) Für die Berechnung des 95 % Konfidenzintervalles für das verfügbare Haushaltseinkommen verwenden Sie die Information der Stichprobengewichte<sup>1</sup> und folgende Formel (mit Endlichkeitskorrekturfaktor  $\sqrt{1 - \frac{n}{N}}$  für die Standardabweichung):

$$\text{KI: } \left[ \bar{x} - \frac{t_{n-1;0.975} * s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}, \bar{x} + \frac{t_{n-1;0.975} * s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} \right],$$

wobei  $x$  das verfügbare Haushaltseinkommen (gesamtes Haushaltseinkommen dividiert durch die OECD-genormte Haushaltsgröße),  $s$  die Standardabweichung,  $n$  die Anzahl der Haushalte in der Stichprobe (der gesamten EU-SILC Daten, ca. 2260) und  $N$  (ca. 3.4 Mio.) die Anzahl aller Haushalte in Österreich ist.  $N$  wird berechnet durch

```
N <- sum(aggregate(eusilc[, "hgew"], list(eusilc[, "ln"]), unique)[,2])
```

Für die Berechnung der 0.975-Quantile der  $t$ -Verteilung verwenden Sie die Funktion `qt()`. Für die Schätzung des verfügbaren Haushaltseinkommens (unten bezeichnet mit `eqinc`) verwenden Sie Variable `ln` (Haushaltsnummer) und die „OECD Gewichte“<sup>2</sup> bzw. folgende R-Befehle:

```
totinc <- aggregate(eusilc[, "pek_g"], list(eusilc[, "ln"]), sum)[,2]
eqsize <- table(eusilc[, "ln"])
eqsize[eqsize > 1] <- 0.5 * (eqsize[eqsize > 1] - 1) + 1
eqinc <- as.numeric(totinc/eqsize)
```

- b) Der wahre Parameterwert wird (*im Mittel*) zu 95 % vom Konfidenzintervall überdeckt, wobei der wahre Parameterwert (fast) immer unbekannt ist. Ausnahmsweise ist hier das „wahre“ arithmetische Mittel der Einkommen der 6848736 Österreicher  $\geq 15$  Jahren aus der Steuerstatistik 2004 bekannt und beträgt 18174 Euro. Wird dieser „wahre“ Parameterwert von Ihrem geschätzten Konfidenzintervall überdeckt? Was könnten die Gründe für eine Nicht-Überdeckung sein? Ist das arithmetische Mittel überdies besonders geeignet zur Angabe eines „durchschnittlichen“ Einkommens oder gibt es in der Regel besser geeignete Lokationsschätzer für das mittlere Einkommen?

Bsp 10 *Risk-at-poverty*: Armutsgefährdet sind laut Definition jene Personen mit einem Haushaltseinkommen weniger als 60 % des verfügbaren Haushalts-Medianeinkommen (`eqinc`).

<sup>1</sup>Die Stichprobengewichte resultieren aus dem Stichprobendesign. Zumeist wurde hier zufällig aus bestimmten Gruppen/Schichten unterschiedlich viele Respondenten gezogen. Z.B. braucht man von den Wienern nicht so viele Respondenten in Relation zur Bevölkerung in Wien als Vorarlberger in Relation zu deren Bevölkerung um auf ein genügend genaue Hochrechnungen zu erreichen. Daher wird z.B. das Stichprobengewicht der Wiener Haushalte größer als das von den Vorarlbergern sein. Dies alles kann zu sehr viel komplexeren Berechnungsmethoden der Varianz führen als in dieser Vorlesung/Übung behandelt.

<sup>2</sup>1 für einen Erwachsenen im Haushalt, 0.5 für jeden weiteren Erwachsenen.

- a) Berechnen Sie den Anteil der armutsgefährdeten Personen von den Daten.
- b) Berechnen Sie diese Schwelle zur Armutsgefährdung für Personen aus unterschiedlichen Beschäftigungsformen (Variable P001000). Dazu benötigen Sie die Äquivalenzeinkommen aller Personen (`eqincP <- rep(eqinc, table(x[, 'ln']))`).
- c) Schätzen Sie die *risk-at-poverty rate* für ganz Österreich getrennt nach unterschiedlichen Beschäftigungsformen unter Berücksichtigung der Stichprobengewichte folgendermaßen:

$$(\text{Haushalte armutsgefährdet}) \quad y = \begin{cases} 1, & \text{wenn } eqinc < \text{Schwelle} \\ 0, & \text{sonst} \end{cases}$$

(Haushaltsgewichte) `w <- aggregate(eusilc[, 'hgew'], list(eusilc[, 'ln']), unique)[,2]`

$$\text{risk-at-poverty rate} = \frac{\sum_{i=1}^h w_i y_i}{\sum_{i=1}^h w_i}, \quad \text{mit } h = \text{Anzahl der Haushalte}$$

- d) Warum ist die Schwelle zur Armutsgefährdung für einzelne Beschäftigungsformen manchmal 0?

### 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 6. November 2007

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

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

**name\_UE\_3.pdf**

wobei 'name' für den Familiennamen steht.

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