

Übung 2

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



1 Daten

Für diese Übung sollen wiederum die EU-SILC Daten (**E**uropean **S**urvey of **I**ncome and **L**iving **C**onditions) vom Jahr 2004 verwendet werden.

Jede/r Student/in soll wieder sein eigenes Subset an Daten analysieren (siehe Übung 1).

Die genaue Beschreibung der Variablen ist wie gehabt im File [EU-SILC rev 065_04.pdf](#) zu finden bzw. sind die *eigentlichen* Variablennamen als `comment(eusilc)` gespeichert. Verwenden Sie für diese Übung eine R-Version $\geq 2.5.0$. Kodieren Sie die Variablen wie in Übung 1.

2 Übungsbeispiele

5. Schätzen Sie Schiefe und Kurtosis (und auch deren approximierten standardisierten Schätzungen) für das Brutto-Einkommen (Variable `pek_g`) > 0 getrennt nach (Ihren) Bundesländern. Interpretieren Sie die Ergebnisse. Transformieren Sie die Brutto-Einkommen > 0 z.B. mit dem Logarithmus (Funktion `log()`) oder einer anderen Transformation zu einer symmetrischeren Verteilung und beschreiben Sie diese Verteilung.
6. Wählen Sie als Variable das Brutto-Einkommen und eine (sinnvolle) zweite Variable und zeichnen Sie einen Scatterplot (Streuungsdiagramm). Interpretieren Sie Ihre Graphik. Zeichnen Sie weiters Scatterplots der Brutto-Jahreseinkommen versus `pek_n` (Netto-Jahreseinkommen) mit Hilfe der Funktion `xypplot()` (aus dem Packet *lattice*). Splitten Sie die Plots anhand Ihrer Bundesländer. Stimmt alles mit den Steuerzahlern oder gibt es Unregelmäßigkeiten (Ausreißer) in einzelnen Bundesländern?
7. Zeichnen Sie eine Boxplot-Darstellung einiger „*stetig*“ skaliertes Variablen, wie z.B. `age`, `pek_g`, `pek_n` und noch einer weiteren Variablen. Achtung: Kodieren Sie zuerst Variable `age` zur Klasse `numeric`, wobei Sie zuvor der Ausprägung „80 und mehr Jahre“ den Wert 85 zuweisen sollten. (Kontrollieren Sie Ihr Ergebnis!)
8. Erzeugen Sie einen approximativ normalverteilten Zufallsvektor mit der Funktion `rnorm()`. Testen Sie, ob dieser Vektor in R laut *Shapiro-Wilk*-Test normalverteilt ist (mit Funktion `shapiro.test()`). Interpretieren Sie Ihr Testresultat.

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 23. Oktober 2007

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

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

name_UE_2.pdf

wobei 'name' für den Familiennamen steht.