



Einführung in die Statistische Software

Cornelia Oberhauser und Micha Schneider

Institut für Statistik

LMU München

Die Dozenten

Wer wir sind ...

Cornelia Oberhauser

- **2002-2007: Studium der Statistik an der LMU (Diplom)**
mit den Anwendungsgebieten Soziologie und Geographie
 - 2005-2007: studentische Hilfskraft am Institut für Statistik
- **Seit 2007: Wissenschaftliche Mitarbeiterin an der LMU**
 - am Lehrstuhl für Public Health und Versorgungsforschung
am Institut für Medizinische Informationsverarbeitung, Biometrie
und Epidemiologie (IBE)
 - Hauptaufgabe: Angewandte Forschung
 - am Institut für Statistik (Statistisches Beratungslabor)
 - Hauptaufgabe: Halten von überwiegend praxisorientierten
Lehrveranstaltungen
- **2014: Abschluss der Promotion zum Doktor der Humanbiologie**
an der Medizinischen Fakultät der LMU München

Micha Schneider

- **Studium der Statistik an der LMU (Diplom)**
mit den Anwendungsgebieten Politische Wissenschaft und Soziologie
 - Auslandsstudium am University College London
 - Eurostat, Phylax Financial Innovations, Teaching Assistant
- **Doktorand am Institut für Statistik (LMU)**
am Lehrstuhl für Angewandte Stochastik (Prof. em. Tutz)
 - Forschungsthema: Modellierung von diskreten Mischstrukturen
 - Lehre: R-Kurse, Statistik NF, Advanced Statistical Modelling
 - Mitarbeit beim Statistischen Beratungslabor und Münchner R-Kurse

Organisatorisches

Ziel der Lehrveranstaltung

- Theoretische Inhalte des Statistik-Studiums müssen praktisch umgesetzt werden.
- Dazu ist die Kenntnis statistischer Software unumgänglich.
- Überblick über gängige Softwarepakete
- Eigenständiges Anwenden von drei Softwarepaketen

Zielgruppe

- Primär: Studierende im Bachelor Statistik
- Aber auch: Alle anderen Teilnehmer der Veranstaltung Deskriptive Statistik

Ablauf

- Theorie:
 - 4 Vorlesungen
 - Ergänzende Materialien auf der Homepage
- Praxis:
 - Übungsblatt
 - Fragestunde
- Bewertung:
 - Hausübung

Theorie: 4 Vorlesungen

- Termine:
 - 31.10.16: Überblick zu Statistischer Software und Einführung in die Benutzung von R
 - 21.11.2016: Vertiefung von R
 - 28.11.2016: Einführung in SPSS
 - 05.12.2016: Einführung in SAS
- Für alle Interessierten! (nicht nur Bachelor Statistik)
- **Dringend empfohlen für alle: Vorlesungen zu R!**
(da R-Kenntnisse im Verlauf des Studiums benötigt und vorausgesetzt werden)
- Zusätzlich: Ergänzende Materialien auf der Homepage zum Selbststudium

Praxis: Übungsblatt und Fragestunde

- Übungsblatt – zu Übungszwecken
 - wird auf der Homepage verlinkt
 - **soll selbständig vor der Fragestunde gelöst werden**
 - wird nicht abgegeben und nicht korrigiert
 - Analysen in R zu Hause möglich (empfohlen für alle)
 - Analysen in SPSS und SAS im CIP-Pool möglich (primär für Bachelor Statistik)
 - offen für alle Interessierten!
(CIP-Pool-Kennung notwendig)
<http://www.statistik.lmu.de/studium/rechnerbetrieb/cip-netz.html>)
- Fragestunde
 - um individuelle Probleme beim Lösen des Übungsblattes zu besprechen
 - Wir gehen herum und helfen bei Problemen weiter.
 - Kein Frontalunterricht!
 - Es gibt keine Musterlösung – es wird nur bei aufgetretenen Problemen geholfen.

Mögliche Termine Fragestunde

- Termine: eher ungünstig, möglich, **präferiert**
- Mo 12.12.16: 8-10, **10-12, 14-16**, 16-18
- Di 13.12.16: 8-10, **10-12, 12-14**
- Mi 14.12.16: 12-14, 16-18
- Do 15.12.16: 10-12
- Fr 16.12.16: 8-10, 10-12
- Bitte zu Beginn des Time-Slots erscheinen, also auf jeden Fall in der ersten halben Stunde.
- Wenn um XX:30 keiner da ist, gehen wir wieder.

Bewertung: Hausübung

- Hausübung
 - wird auf der Homepage verlinkt
 - umfasst Analysen in R, SPSS und SAS + Interpretation
 - **soll selbständig bearbeitet werden**
 - **muss abgegeben werden (nur Bachelor-Statistik)**
 - **Jeder muss einzeln abgeben.**
 - Primär für Bachelor Statistik, aber auch offen für Interessierte.
- Bewertung
 - „Bestanden“ versus „Nicht bestanden“.
 - Detailliertere Rückmeldung nur auf Anfrage.
 - Bestehen dieser Hausübung ist Voraussetzung, um die Veranstaltung Deskriptive Statistik bestehen zu können. (nur Bachelor Statistik)

Überblick zu Statistischer Software

Anforderungen an statistische Software

- Korrekter Umgang mit den wichtigsten Skalenniveaus der Statistik (nominal, ordinal, metrisch)
- Erlauben von/Umgang mit fehlenden Werten
- Funktionen für deskriptive Analysen, Schätzungen und Modellierung
- Visualisierung von Daten
- Förderung der Reproduzierbarkeit von Analysen (Skripte, Programme)
- Abspeichern von Ergebnissen für Weiterverarbeitung

Vorgestellte Software

- R
- SPSS
- SAS

R: Stärken und Schwächen

- + Kostenlos (-> kann überall verwendet werden)
- + auf (fast) jedem Betriebssystem lauffähig
- + Sehr großer Funktionsumfang (Tausende Pakete)
- + Funktionen im Quelltext nachvollziehbar
- + Neue Methoden werden schnell implementiert
- + Guter Community Support
- Einarbeitung in die Syntax und die wichtigsten Funktionen zeitintensiv
- Stabilität und Qualität wenig genutzter Pakete oft fraglich und deutlich schlechter als die der in der Basis-Version enthaltenen Pakete
- Schwächen bei sehr großen Datensätzen

SPSS: Stärken und Schwächen

- + Leicht erlernbar
- + Einfaches, wenngleich nicht immer konsistentes GUI
- + Syntax kann für Reproduzierbarkeit zusammengeklickt werden
- + Umfangreiche Hilfe
- Stabilität nicht überragend (stürzt des öfteren ab!)
- Für sehr große Datenmengen nicht geeignet
- Teuer für den nichtakademischen Bereich
(Für Forschung und Lehre: 60 Euro pro Jahr über das LRZ)

SAS: Stärken und Schwächen

- + Stabil und schnell, auch für große Datensätze
- + Sehr gute Dokumentation und professioneller Support
- + Makros für Reproduzierbarkeit
- + Gesamtpakete für bestimmte Aufgaben

- Erlernung der Syntax zeitintensiv
- Benutzerschnittstellen nicht immer zeitgemäß
- Sehr teuer für den nichtakademischen Bereich
(Für Forschung und Lehre: 150 bzw. 440 Euro pro Jahr über das LRZ)

Weitere Software

STATA

- o beliebt in folgenden Bereichen:
Soziologie, Ökonomie, Politik- und Sozialwissenschaft
sowie Epidemiologie
- + gute Dokumentation
- + schnelle Algorithmen
- + preislich attraktiv im Vergleich zu SAS oder SPSS

Weitere Software

MATLAB

- + richtige Programmiersprache
- + gute Grafik
- + hervorragende Algorithmen
- o Preis hängt stark von der Zahl der gewünschten Module ab

Interessante Links

- Gute Einführung in die einzelnen Softwarepakete:
UCLA Institute for Digital Research and Education
<http://www.ats.ucla.edu/stat/>
- Überblick über die grundlegendsten R-Funktionen:
R Reference Card by Tom Short
<ftp://cran.r-project.org/pub/R/doc/contrib/Short-refcard.pdf>
- SPSS-Lizenz über das LRZ
<http://www.lrz.de/services/swbezug/lizenzen/spss/>
- SAS-Lizenz über das LRZ
<http://www.lrz.de/services/swbezug/lizenzen/sas/>