

*Prof. Dr. Wolfgang Ludwig-Mayerhofer
Universität Siegen, Philosophische Fakultät – Soziologie
Wintersemester 2017/18*

Übung Statistik im Master-Studiengang (auch Statistik III – Multivariate Statistik)

Mi. 10:00–12:00 Uhr, AR-A 1007

Sprechstunde: Di. 15:00–16:00

Tel. (0271) 740 – 3046 (Frau Schatte-Utsch, Sekr.)

sekretariat.wlm@uni-siegen.de (Organisatorisches)

ludwig-mayerhofer@soziologie.uni-siegen.de (Inhaltliches)

Die Veranstaltung

Die Lehrveranstaltung soll einen ersten Einblick in die Fragestellungen und Vorgehensweisen multivariater Analyseverfahren geben. Das Augenmerk liegt auf sozialwissenschaftlichen Anwendungen und auf der praktischen Umsetzung der Verfahren mittels einschlägiger Statistik-Software – konkret: Stata.

Wir können im Rahmen der Lehrveranstaltung nur einen kleinen Teil des gesamten Gebietes der multivariaten Statistik kennenlernen. Im Kern erarbeiten wir uns einige wenige Varianten sogenannter Regressionsverfahren, in denen Einflüsse mehrerer unabhängiger Variablen auf eine abhängige Variable analysiert werden. Auch die Faktorenanalyse, die wir zwischendurch kurz besprechen, lässt sich von den Regressionverfahren her begreifen. Im einzelnen sollen folgende Verfahren besprochen werden:

- Lineare Regressionsanalyse;
- (Explorative) Faktorenanalyse und Itemanalyse (Verfahren der Datenreduktion);
- Logistische Regression für binäre und polytome abhängige Variablen;
- evtl. noch Anfangsgründe der Verlaufsdatenanalyse (Einflüsse auf eine abhängige Variable, die die Dauer bis zum Erreichen eines definierten Zielzustandes angibt).

Einen genauen Ablaufplan gibt es nicht, da es auch auf Ihr Arbeitstempo ankommt. Grundsätzlich können Sie davon ausgehen, dass wir auf alle Verfahren ca. drei bis vier Wochen verwenden; nur die Faktorenanalyse, wenngleich sie ein sehr komplexes Verfahren darstellt, möchte ich nach Möglichkeit kürzer (das heißt: ausschließlich anwendungsbezogen) abhandeln.

Leistungspunkte

sowie gegebenenfalls eine Note können Sie durch kontinuierliche Mitarbeit in der Lehrveranstaltung und durch die Bearbeitung von Hausaufgaben erwerben. Ob sie eine Studien- oder eine Prüfungsleistung oder beides erbringen können, hängt von Ihrer Studienordnung ab; die Details werden in der Veranstaltung besprochen. *Die Aufgaben sind in Zweier-, evtl. (nach Absprache) in Dreierteams zu bearbeiten!*

Die Aufgaben werden im Verlauf des Semesters – auch in Abhängigkeit davon, wie schnell wir vorankommen – ausgegeben und müssen jeweils bis zum Montag der kommenden Woche bis 17 Uhr per E-Mail an mich persönlich (also nicht Sekretariat!) geschickt werden. Bitte geben Sie Ihrer Datei jeweils einen Namen, der sich wie folgt zusammensetzt:

Nachname(n)_Aufgabengebiet_Aufgabe_Nummer(n).

Im Beispiel:

Karlstadt-Valentin_Lineare-Regression_Aufgabe_1-2.

Bitte schreiben Sie in die Betreffzeile Ihrer E-Mail „Statistik 3“.

Beachten Sie außerdem: Ich gehe per Voreinstellung davon aus, dass Sie sich in ihren Ausarbeitungen auf meine Präsentation oder auf die Ausführungen während der Lehrveranstaltungen beziehen. Diese können Sie wie Allgemeinwissen behandeln, müssen sie also nicht als Quelle ausweisen. Jede andere Quelle, an der Sie sich direkt oder indirekt orientieren, muss dagegen in der Arbeit nachgewiesen werden. Das gehört nicht nur zu den Standards wissenschaftlichen Arbeitens, sondern dient auch Ihrem Schutz! Denn manche Bücher oder Online-Quellen, die Studierende finden, erzählen ziemlich Unfug; und auch in guten Büchern werden u. U. manche Themen nicht differenziert genug dargestellt. Wenn Sie solche Quellen heranziehen und dies nicht belegen, muss ich davon ausgehen, dass die Aussagen auf Ihrem eigenen Mist gewachsen sind.

Der Verweis auf Quellen entbindet Sie aber nicht davon, in der Veranstaltung aufmerksam zu sein. Wenn ich anderer Auffassung bin als die Quellen, so sollten Sie nicht so tun, als gäbe es meine Auffassung überhaupt nicht; das mindeste wäre, beide Auffassungen gegeneinander abzuwägen.

Wichtig: Erfolgreiche Teilnahme an der Veranstaltung setzt erfahrungsgemäß ein beträchtliches Ausmaß an Eigenstudium voraus (siehe dazu die nachfolgenden Literaturhinweise). Der gesamte Workload ist *während* der Vorlesungszeit zu erbringen.

Daten und Software

Zur Bearbeitung der Hausaufgaben benötigen Sie u. a. den Datensatz ALLBUS 2002. Diesen können Sie auf dem Uni-Server finden, aber auch selbst bei der GESIS (www.gesis.org) herunterladen. (Die GESIS-Seiten wurden in letzter

Zeit mehrfach umstrukturiert, daher müssen Sie gegebenenfalls dort selbst suchen; der Datensatz hat die Nummer 3700).

Stata ist im Gebäude AR nur im Pool AR-A 1007 installiert. Ich unterschreibe gerne Formulare für einen (zeitlich begrenzten) elektronischen Zugangsschlüssel zu diesem Pool. Weitere Pools mit Stata befinden sich wahrscheinlich im Unteren Schloss, diese Pools sind jedoch (wie auch unser Pool) meist für Lehrveranstaltungen belegt.

Jahres- und Dauerlizenzen für Stata sind bei Dittrich & Partner zu (naja, einigermaßen) günstigen Preisen für Studierende erhältlich. Damit können Sie die Hausaufgaben auf Ihrem eigenen Rechner bearbeiten. Die Version „IC“ (Intercooled) ist für unsere Zwecke ausreichend. Sollten Sie an längerfristige Befassung mit der Materie denken, empfehle ich allerdings die Version SE.

Arbeitsmaterialien

1. Lehrbücher und Web-Materialien

- Backhaus, Klaus, Erichson, Bernd, Plinke, Wulff & Weiber, Rolf (2003): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Zehnte, neu bearbeitete und erweiterte Auflage (oder neuere Auflage). Berlin, Heidelberg, New York: Springer.
Website zum Buch: <http://www.multivariate.de/>
- Tabachnick, Barbara G. & Fidell, Linda S. (4. Aufl. 2001): *Using Multivariate Statistics*. Boston: Allyn & Bacon.
Website: <http://www.ablongman.com/tabachnick/stats/index.html>

Nur das letztere Buch deckt den Vorlesungsstoff weitgehend ab (einschließlich Verlaufsdatenanalyse); beide Bücher enthalten Kapitel über zahlreiche weitere Verfahren, für die die Vorlesungszeit nicht ausreicht. Beide Bücher finden Sie in meinem Semesterapparat.

Mehr anwendungsorientiert und voll mit ausgezeichneten Tipps ist folgendes Buch:

- Treiman, Donald J. (2009): *Quantitative Data Analysis. Doing Social Research to Test Ideas*. San Francisco: Jossey-Bass.

Zu Stata:

- Kohler, Ulrich & Kreuter, Frauke: *Datenanalyse mit Stata*, 4. Aufl. 2012, München: Oldenbourg.
- Mein Internet Guide to Stata, zugänglich über meine Uni-Webseite.

Beachten Sie auch mein Internet-Lexikon der Methoden der empirischen Sozialforschung (ILMES): <http://wlm.userweb.mwn.de/ilmes.htm>. Dort finden Sie vielleicht gelegentlich Erläuterungen, die Ihnen weiterhelfen, vor allem dann, wenn Ihre Vorkenntnisse nicht sehr gut sind.

2. Beachten Sie den Moodle-Kurs zu „Statistik III“. Dort finden sich einige einführende Texte, meine Präsentation sowie die Aufgaben.

3. Schließlich möchte ich Ihnen einige ausführlichere Texte als Hinweise für *spätere Vertiefung* empfehlen. Diese benötigen Sie also nicht zwingend im laufenden Semester, doch schadet es nicht, mal einen Blick hineinzuworfen. Soweit in der UB vorhanden, finden Sie die Bücher gleichfalls im Semesterapparat.

Zur linearen Regression:

- Fox, John (1997): Applied Regression Analysis, Linear Models, and Related Methods. Newbury Park: Sage.

Zur logistischen Regression und verwandten Verfahren:

- Agresti, Alan (2. Aufl. 2002): Categorical Data Analysis. New York: Wiley.
- Andrefß, Hans-Jürgen, Hagenaars, Jacques-A. & Kühnel, Steffen (1997): Analyse von Tabellen und kategorialen Daten. Berlin u. a.: Springer.

Zur Verlaufsdatenanalyse:

- Blossfeld, Hans-Peter, Golsch, Katrin & Rohwer, Götz (2007): Event History Analysis with Stata. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cleves, Mario, Gutierrez, Roberto G., Gould, William & Marchenko, Yulia V. (2010): An Introduction to Survival Analysis Using Stata (Third Edition). College Station, Texas: Stata Press.
- Collett, David (2. Aufl. 2003): Modelling Survival Data in Medical Research. London: CRC Press.

Weitere Hinweise finden Sie auf den Literatur-Seiten des ILMES.