



Übung im Frühjahrssemester 2012: Wenn die Anzahl zählt – Regressionsanalyse von Häufigkeits- bzw. Zähldaten

Veranstaltungsplan

Veranstaltungsleiter:

Sebastian Weingartner | weingartner@soziologie.uzh.ch
Sprechstunde: nach Vereinbarung per Email, Raum: AND 5.53

Zeit/Ort:

Mittwochs, 12.15h – 13.45h | wöchentlich vom 22.02.2012 bis zum 30.05.2012
Osterferien: 11.04.2012 entfällt
Raum: RAI-D021

Inhalt:

In den empirisch ausgerichteten Sozialwissenschaften ist die Analyse von Häufigkeits- bzw. Zähldaten ein verbreitetes Untersuchungsfeld. Es handelt sich dabei um Daten, die die Anzahl eines Ereignisses in einer bestimmten Periode abbilden, zum Beispiel die Anzahl der Telefonanrufe pro Tag, die Anzahl der Kinder pro Frau, die Zahl der von einer Bank vergebenen Kredite pro Jahr oder die Anzahl der Theaterbesuche einer Person in einem Monat. Aufgrund ihrer spezifischen Struktur (ausschließlich nicht-negative, ganzzahlige Werte) lassen sich solche Daten allerdings nicht durch ein lineares Regressionsmodell analysieren, sondern es bedarf spezieller Vorgehensweisen, die die Verteilung von Häufigkeits- bzw. Zähldaten berücksichtigen. Ziel der Veranstaltung ist es, diese Methoden (theoretisch) vorzustellen und anhand von Anwendungsbeispielen einzuüben. Dazu wird die Statistiksoftware STATA verwendet.

Nach einer kurzen Wiederholung linearer und binärer Regressionsmodelle wird zunächst die Poisson-Regression als Ausgangspunkt der Analyse von Häufigkeits- bzw. Zähldaten vorgestellt. Um spezielle Datenstrukturen, die auch mit der Poisson-Regression nicht adäquat gefasst werden können, zu berücksichtigen, werden darüber hinaus Negative-Binomial- und sog. Zero-Inflated- und Truncated-Modelle vorgestellt und an Beispielen erarbeitet.



Teilnahmebedingungen und Leistungsnachweis:

Studierende (ab dem 3. Fachsemester) im Bachelor-, Master- oder Lizentiatstudiengang können in dieser Übung teilnehmen. Bedingungen hierfür sind:

- regelmässige Teilnahme
- Vorbereitung auf Basis der angegebenen Literatur
- Beteiligung an der Kursdiskussion und -gestaltung

Studierende im Bachelor- oder Masterstudiengang können darüber hinaus einen Leistungsnachweis über **4 ECTS-Punkte** erwerben. Dazu ist die schriftliche Bearbeitung von zwei Übungsblättern erforderlich (siehe Hinweise).

Hinweise:

Die Bearbeitung der Übungsblätter umfasst vor allem die selbstständige statistische Berechnung vorgegebener Aufgaben, zum Teil aber auch (kurze) theoretische Fragestellungen. Ein Datensatz wird zur Verfügung gestellt. Die Ergebnisse bzw. Antworten sind in einem separaten Dokument zusammenzufassen. Die Vorgehensweise mit der Statistiksoftware STATA soll anhand von *do-files* dokumentiert werden, die zusammen mit dem Ergebnisdokument beim Veranstaltungsleiter abzugeben sind.

Spätester Abgabetermin für die Bearbeitung der Übungsblätter ist der **30.06.2012**, es wird jedoch empfohlen, die Bearbeitung im Zusammenhang mit den entsprechenden Veranstaltungsthemen möglichst zeitnah zu erledigen. Die schriftlichen Arbeiten (Ergebnisdokument und *do-files*) sollen als Dateien per Email abgegeben werden.

Sämtliche Veranstaltungsmaterialien und -literatur finden sie auf der OLAT-Plattform unter: <http://www.olat.uzh.ch> (Anmeldung erforderlich!)

STATA ist eine kostenpflichtige Software, in der Übung wird sie zu Verfügung gestellt. Für Studierende der Universität Zürich ist STATA in folgenden Räumen öffentlich zugänglich:

- Zentrum: RAI D005, KOL K01 (Turm)
- Irchel: Y11 E05
- SUZ: AND 5.47

STATA-Lizenzen für Studierende können bei Scientific Solutions (www.scientific-solutions.ch) erworben werden:

- STATA IC (max. 2047 Variablen): 225 Fr.
- STATA SE (max. 32767 Variablen): 525 Fr.



Ablaufplan

(T : Theoretische Literatur
S : Literatur zur Anwendung mit STATA)

1. Einführung

- 1.1. Grundlagen der Regressionsanalyse
- 1.2. Ein wenig über Schätztheorie
- 1.3. Ein wenig über Testtheorie

(T) Jann (2005): Kap. 5
(T) Fahrmeier et al. (2011): Kap. 9+10
(S) Long/Freese (2006): Kap. 3
(S) Cameron/Trivedi (2010): Kap. 12

2. Lineare Regression

- 2.1. Das lineare Regressionsmodell
- 2.2. Annahmen der linearen Regression
- 2.3. Interpretation der Effekte
- 2.4. Nicht-lineare lineare Regression

(T) Long (1997): Kap. 1+2
(S) Cameron/Trivedi (2010): Kap. 3

3. Binäre (logistische) Regression

- 3.1. Das lineare Wahrscheinlichkeitsmodell
- 3.2. Modellierung über eine latente Variable
- 3.3. Nicht-lineares Wahrscheinlichkeitsmodell

(T) Long (1997): Kap. 3
(S) Long/Freese (2006): Kap. 4
(S) Cameron/Trivedi (2010): Kap. 14

4. Regressionsmodelle für Häufigkeits- und Zähldaten

- 4.1. Häufigkeitsdaten und Poisson-Verteilung
- 4.2. Das Poisson-Regressionsmodell
 - Interpretation der Effekte
 - Regressionsdiagnostik
- 4.3. Das Negative-Binomial-Regressionsmodell
- 4.4. Überhäufige Null-Werte: Zero-Inflated-Modelle
- 4.5. Beschränkte Werte: Truncated-Modelle

(T) Long (1997): Kap. 8
(T) Cameron/Trivedi (1998): Kap. 1-3; 10
(S) Long/Freese (2006): Kap. 8
(S) Cameron/Trivedi (2010): Kap. 17

Ergänzung: Analyse von Häufigkeits- und Zähldaten durch Generalized Linear Models (GLM) (wird nicht in der Übung besprochen!)

(T) Cameron/Trivedi (1998): Kap. 4
(T) Simonoff (2003): Kap. 5



Literatur:

- Cameron, A. Colin; Trivedi, Pravin K. (2010): *Microeconomics Using Stata*. College Station: Stata Press
- Cameron, A. Colin; Trivedi, Pravin K. (1998): *Regression Analysis of Count Data*. Cambridge: Cambridge University Press
- Fahrmeier, Ludwig; Künstler, Rita; Pigeot, Iris; Tutz, Gerhard (2011): *Statistik. Der Weg zur Datenanalyse*. Berlin/Heidelberg: Springer
- Jann, Ben (2005): *Einführung in die Statistik*. München/Wien: Oldenbourg
- Long, J. Scott; Freese, Jeremy (2006): *Regression Models for Categorical Dependent Variables Using Stata*. College Station: Stata Press
- Long, J. Scott (1997): *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage
- Simonoff, Jeffrey S. (2003): *Analyzing Categorical Data*. New York: Springer

Weitere Literatur zur sozialwissenschaftlichen Datenanalyse:

- Agresti, Alan ; Finlay, Barbara (2008): *Statistical Methods for the Social Sciences*. Upper Saddle River: Prentice Hall
- Fox, John (2008): *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks: Sage
- Gujarati, Damodar N. (2004): *Basic Econometrics*. New York: McGraw-Hill
- Wooldridge, Jeffrey M. (2009): *Introductory Econometrics: A Modern Approach*. Mason: South Western

Weitere Literatur zum Umgang mit STATA:

- Acock, Alan C. (2010): *A Gentle Introduction to Stata*. College Station: Stata Press
- Kohler, Ulrich; Kreuter, Frauke (2008): *Datenanalyse mit Stata: Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung*. München: Oldenbourg
- Kohler, Ulrich; Kreuter, Frauke (2008): *Data Analysis Using Stata*. College Station: Stata Press
- Mitchell, Michael N. (2012): *A Visual Guide to Stata Graphics*. College Station: Stata Press