

Vorlesung: Statistik II für Studierende der Soziologie und Nebenfachstudierende

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2013

Besonderer Dank gilt Prof Augustin, der mir das Material zur Verfügung gestellt hat

Termine und Informationen

Homepage:

http://www.statistik.lmu.de/institut/ag/statsoz_neu/lehre/2013_SoSe/Stat2Soz_13/index.html

Vorlesung:

Prof. Helmut Küchenhoff

Mi. 12:00 bis 14:00 Uhr Hauptgebäude M 018

Do. 12:00 bis 14:00 Uhr Hauptgebäude M 018

Übung 1	Di.	12:15 bis 13:45 Uhr	HGB-E 004	Hauptgebäude
Übung 2	Di.	14:15 bis 15:45 Uhr	HGB-E 004	Hauptgebäude
Tutorium	Mo.	12:15 bis 13:45 Uhr	HGB-M 010	Hauptgebäude

L.Fahrmeir, R.Künstler, I.Pigeot, G.Tutz:
Statistik - Der Weg zur Datenanalyse
Springer-Verlag, 7. Auflage

Helmut Küchenhoff et al. (2006):
Statistik für Kommunikationswissenschaftler
2., überarbeitete Auflage
UVK Verlagsgesellschaft mbH, Konstanz



- 0 Einführung
- 1 Wahrscheinlichkeitsrechnung
- 2 Zufallsvariablen und ihre Verteilung
- 3 Statistische Inferenz
- 4 Hypothesentests
- 5 Regression



- 0 Einführung
- 1 Wahrscheinlichkeitsrechnung
- 2 Zufallsvariablen und ihre Verteilung
- 3 Statistische Inferenz
- 4 Hypothesentests
- 5 Regression

Deskriptive Statistik (Statistik I):

- Beschreibung von Daten (Grundgesamtheit oder Stichprobe).
- Keine Verallgemeinerung von einer Stichprobe auf die zugehörige Grundgesamtheit angestrebt.

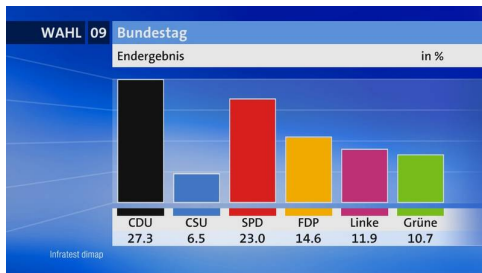
Induktive Statistik (Statistik II):

- Induktion: Schluss vom Teil auf das Ganze, von vielen Einzelbeobachtungen auf allgemeine Gesetze
- Schluss von einer Stichprobe auf Eigenschaften der Grundgesamtheit (*Inferenz*)

Beispiel 1: Bundestagswahl 2009

Prognose 18:00 Infratest Dimap (ARD)

CDU/CSU	SPD	FDP	Linke	Grüne	Sonstige
33,5	22,5	15	12,5	10,5	6



Basis: Nachwahlbefragung 100 000 Wahlberechtigte nach Verlassen der Wahllokale

Infos unter

<http://www.tagesschau.de/wahl/umfragen/infratest102.html>

Beispiel 1: Bundestagswahl 2009

- Grundgesamtheit: Alle Wähler der Bundestagswahl
- Stichprobe: 100 000 Wähler
- Gesucht: Information über *alle* Wähler, also die *Grundgesamtheit*

Stichprobe dient zum Lernen über die Grundgesamtheit

Beispiel 2: Kardiale Notfälle

Studie mit Medizinischer Klinik in Großhadern

Prospektive Analyse kardialer Notfälle

01.05. - 31.07.2006

Retrospektive Analyse

01.05. - 3.07.2003

01.05. - 3.07.2005

Analyse der Protokolle von 24 Notarztstandorten aus München sowie näherer und weiterer Umgebung

Einschlusskriterien:

ST-Hebungsinfarkt, Nicht-ST-Hebungsinfarkt/instab. Angina

Herzrhythmusstörungen mit ausgeprägter/lebensbedrohlicher Symptomatik

Herzrhythmusstörungen mit geringer Symptomatik

Beispiel 2: Kardiale Notfälle

Ziele und Methoden:

- Beschreibung des Zusammenhangs zwischen emotionalem Stress und Auftreten von Herzinfarkten
- Berücksichtigung von Störgrößen
- Schluss von Beobachtung auf allgemeines Gesetz
- Bewertung des Risikos

Beispiel 3: Lebenszufriedenheit und Alter

- Gibt es eine Midlife Crisis?
- Analysen von Panel-Daten zur subjektiven Lebenszufriedenheit mit
- semiparametrischen Regressionsmodellen
- In Zusammenarbeit mit Sonja Greven, Andrea Wiencierz, Christoph Wunder

Beispiel 3: Lebenszufriedenheit und Alter

Datengrundlage

- Daten stammen aus den Haushaltsstichproben A (Westdeutsche) und C (Ostdeutsche) des Sozio-Ökonomischen Panels (SOEP)
- für die ausgewählten Modellvariablen liegen Beobachtungen aus den Jahren 1992, 1994 bis 2006 vor
- durchschnittliche Anzahl von Beobachtungen pro Person: 7.77
- in die Modellberechnungen gingen 102 708 vollständige Beobachtungen von 13 224 Individuen ein
- Anzahl Beobachtungen pro Jahr:

1992	1994	1995	1996	1997	1998	1999
8 145	7 720	7 943	7 606	8 052	7 550	7 403
2000	2001	2002	2003	2004	2005	2006
7 628	7 092	7 068	7 000	6 876	6 543	6 082

Beispiel 3: Lebenszufriedenheit und Alter

Methode: Multiples Lineares Regressionsmodell

- Zielgröße: Subjektive Lebenszufriedenheit
- Einflussgrößen: Alter Gesundheit, Gehalt usw.
- Hauptfrage: Wie hängen Lebenszufriedenheit und Alter zusammen
- Alterseffekt wird nicht parametrisch modelliert

Beispiel 3: Lebenszufriedenheit und Alter

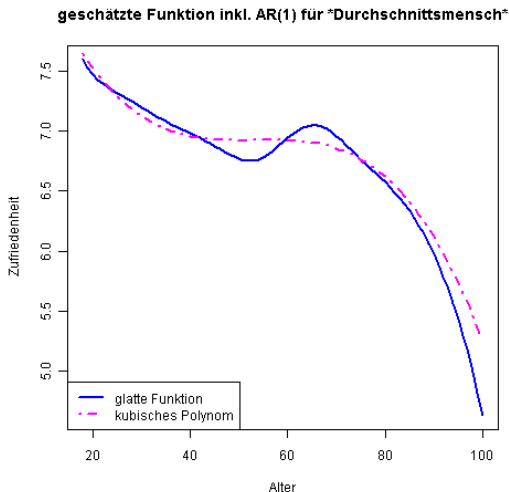
Ergebnisse für das Regressionsmodell

Variable	Coefficient	Standard error
Sex: female	0.074	(0.015)
Disability status: disabled	-0.452	(0.014)
Nights stayed in hospital	-0.012	(0.000)
Years of education	0.034	(0.002)
Log of net household income	0.492	(0.010)
Log of household size	-0.194	(0.012)
German	0.053	(0.020)
Full time employed	0.079	(0.011)
Part time employed	0.019	(0.012)
Unemployed	-0.597	(0.014)
Single	-0.174	(0.017)
Divorced	-0.137	(0.018)
Widowed	-0.196	(0.023)
West-Germany	0.511	(0.017)

Aus: WUNDER, C., WIENCIERZ, A., SCHWARZE, J. and KÜCHENHOFF, H.(2011). Well-Being over the Life Span: Semiparametric Evidence from British and German Longitudinal Data. *Review of Economics and Statistics*. Accepted for publication.

Beispiel 3: Lebenszufriedenheit und Alter

Ergebnis für Alterseffekt



Midlife-Crisis nur bei glatter Funktion erkennbar.

Beispiel 3: Lebenszufriedenheit und Alter

Ziele und Methoden

- Zusammenhänge analysieren
- Komplexe Einflüsse
- flexibles Modell

Beispiel 4: Mineralwasserstudie

Studie in Zusammenarbeit mit Prof. Adam (LMU)

Fragestellung: Schmeckt mit Sauerstoff angereichertes Mineralwasser besser als gewöhnliches Mineralwasser ?

- DoppelBlindstudie
- KontrollGruppe: zweimal das gleiche Wasser ohne O_2
- VerumGruppe: Beim zweiten Mal mit O_2 angereichertes Mineralwasser

Ergebnis (Clausnitzer et al., 2004):

Placebo: 76% gaben an, dass das zweite Wasser anders schmeckt

Verum : 89 % gaben an, dass das zweite Wasser anders schmeckt

Signifikanter Effekt → Zulassung von



- **Punktschätzung:**

- Zum Beispiel: Wie groß ist der Anteil der schwarz-gelb-Wähler unter allen Wahlberechtigten?
- Wie erhält man aus der Stichprobe gute Schätzwerte für Charakteristika („Parameter“) der Grundgesamtheit?
- Wann ist ein Schätzverfahren gut/besser als ein anderes?

- **Bereichsschätzung:**

- Typischerweise stimmt der Punktschätzer nicht mit dem wahren Wert überein.
- Realistischer: Gib einen Bereich an, „der den wahren Anteil der schwarz-gelb-Wähler mit hoher Wahrscheinlichkeit enthält“.
- Ungenauere Aussagen, dafür aber zuverlässiger.

- Überprüfe aus substanzwissenschaftlicher Theorie abgeleitete Hypothesen über die Grundgesamtheit anhand der Daten.
- Zum Beispiel: Verdienen Männer wirklich mehr als Frauen?
- Ist der Unterschied „signifikant“?

Regressionsmodelle incl. Varianzanalyse

- Modelle zur Beschreibung des Einflusses von Variablen
- Zum Beispiel: Wie hängt die Lebenszufriedenheit vom Alter ab ?

- Zentrales Problem der induktiven Statistik: *Jeder* Induktionsschluss ist potentiell fehlerbehaftet
- Beispielsweise ist der wahre Anteil der Wähler einer Partei in der Grundgesamtheit nicht exakt mithilfe der Stichprobe vorhersagbar.
- Entscheidende Idee: Kontrolle des Fehlers mit Hilfe der Wahrscheinlichkeitsrechnung.
 - Stichprobenziehung *zufällig*
 - Verwende Methoden der Wahrscheinlichkeitsrechnung zur Quantifizierung/Kontrolle des Fehlers

- Verwende statistische Modelle mit zufälligen Komponenten
- „Gesetze“ bzw. Aussagen enthalten stochastischen Aspekt

Kap. 1: Wahrscheinlichkeitsrechnung

Kap. 2: Wie nutzt man Wahrscheinlichkeitsüberlegungen für die Statistik?

Induktive Statistik