Statistik I

für Studierende der Soziologie und Nebenfachstudierende

Gernot Müller, WS 2011/12

Herrn Prof. Dr. Thomas Augustin danke ich für die Überlassung der Materialen seiner gleichnamigen LV im WS 2009/10, auf denen einige Teile dieser Lehrveranstaltungsunterlagen beruhen.

Organisatorisches

- Vorlesung: Gernot Müller
  Dienstag: 8-10 c.t. THE B 052
  Donnerstag: 14-16 c.t. HGB M 118

- Übung (für Studierende der Soziologie): Gero Walter
  Montag: 16-18 c.t. THE C 123

- Tutorium (für Studierende der Medieninformatik): Felix Klug
  Montag: 8-10 c.t. HGB M 109

- http://www.statistik.lmu.de/institut/ag/statsoz_neu/lehre/2011_WiSe/Stat1Soz_1112/
- Folien werden i.d.R. ca. 24 Stunden vor der Vorlesung über die Homepage zur Verfügung gestellt
- Folien sind kein Skriptum, sie sollen Ihnen das Mitschreiben erleichtern!
- Prüfung am Ende des Semesters: Klausur
- Probeklausur am 22.12.2011

Literatur

Bamberg, G. & **Baur**, F. (2009): **Statistik**. R. Oldenburg Verlag, München, Wien.

Benninghaus, H. (2007): **Deskriptive Statistik: Eine Einführung für Sozialwissenschaftler (Studienskripten zur Soziologie)**. VS Verlag für Sozialwissenschaften, Wiesbaden.

Fahrmeier, L. & **Künstler**, R. & **Pigeot**, I. & **Tutz**, G. (2009): **Statistik - Der Weg zur Datenanalyse**. Springer Verlag, Berlin, Heidelberg, New York.

Fahrmeier, L. & **Künstler**, R. & **Pigeot**, I. & **Tutz**, G. & **Caputo**, A. & **Lang**, S. (2008): **Arbeitsbuch Statistik**. Springer Verlag, Berlin, Heidelberg, New York.

Jann, B. (2005): **Einführung in die Statistik**. R. Oldenburg Verlag, München, Wien.

Janssen, J. & **Laatz**, W. (2009): **Statistische Datenanalyse mit SPSS: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests**. Springer Verlag, Berlin, Heidelberg, New York.

Rasch, B. & **Friese**, M. & **Hofmann**, W.J. & **Naumann**, E. (2009): **Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler**. Springer Verlag, Berlin, Heidelberg, New York.

Rohwer, G. & **Pötter**, U. (2001): **Grundzüge der sozialwissenschaftlichen Statistik**. Juventa (Grundlagentexte Soziologie). Weinheim, München.

Rohwer, G. & **Pötter**, U. (2002): **Wahrscheinlichkeit. Begriff und Rhetorik in der Sozialforschung**. Juventa (Grundlagentexte Soziologie). Weinheim, München.

Toutenburg, H. & **Heumann**, C. (2009): **Deskriptive Statistik**. Springer Verlag, Berlin, Heidelberg, New York.

Toutenburg, H. (2008): **Induktive Statistik**. Springer Verlag, Berlin, Heidelberg, New York.

Toutenburg, H. & **Schomaker**, M. & **Wißmann**, M. & **Heumann**, C. (2009): **Arbeitsbuch zur deskriptiven und induktive Statistik**. Springer Verlag, Berlin, Heidelberg, New York.

Wagschal, U. (2010): **Statistik für Politikwissenschaftler**. R. Oldenburg Verlag, München, Wien.

1 Einführung

1.1 Was ist Statistik ?

Gabler's Wirtschaftslexikon, 2011:

„[Statistik ist ein] umfassendes methodisch-quantitatives Instrumentarium zur Charakterisierung und Auswertung empirischer Befunde bei gleichartigen Einheiten (Massenphänomenen) mit universellen Einsatzmöglichkeiten in Politik, Wirtschaft und Gesellschaft und allen Geistes-, Sozial- und Naturwissenschaften einschließlich Medizin und Technik [...]. Ergebnisse statistischer Untersuchungen werden ebenfalls als Statistik bezeichnet. “

- Methodenlehre, Wissenschaft
- Informationen in Form von Tabellen und Diagrammen

Gabler's Wirtschaftslexikon, 2011:

„Statistische Methoden sind darauf ausgerichtet, unter gleichen Rahmenbedingungen häufig beobachtbare Vorgänge mithilfe von Zahlen allgemein zu charakterisieren. Typischerweise sind statistische Untersuchungen auf die Beschreibung (Deskription) oder Erkundung (Exploration) einer geeignet abgegrenzten Grundgesamtheit von Einheiten [...] gerichtet und insbesondere darauf, Schlussfolgerungen über die Strukturen in einer Grundgesamtheit (Induktion; Inferenz) zu ermöglichen. "

Warum Statistik?

Ziel der Statistik ist meist:

- eine genauere Beschreibung der Wirklichkeit
- das Aufdecken von Zusammenhängen und Strukturen ($\rightarrow$ statistische Modelle)
- Unterstützung einer Entscheidungsfindung, oft auch auf Grundlage einer Prognose von **zukünftigen** Entwicklungen.

Problem bei Prognosen: Was ist, wenn sich die Rahmenbedingungen verändern?

Achtung:
Viele Bestätigungen steigern nicht die Allgemeingültigkeit einer Annahme
(Truthahn-Beispiel von Bertrand Russell)

Ausmaß des Mangels an quantitativ qualifizierten Absolventen sozialwissenschaftlicher Studiengänge

- Untersuchung von Rainer Schnell (Jetzt: Universität Duisburg-Essen)
- Information aus dem Arbeitgeberinformationssystem (AIS): Daten über größten Teil der bundesweit arbeitslos gemeldeten Personen.
- Analyse von 1745 arbeitslosen Soziolog(inn)en auf
  - Beherrschung Statistik-Software (z.B. SPSS),
  - Spezielle Statistik-Kenntnisse,
  - Erfahrung bei der Durchführung quantitativer empirischer Projekte,
  - Erfahrung bei der Durchführung qualitativer empirischer Projekte.

Qualifikationsprofile der am 1.6.01 arbeitslos gemeldeten Soziologen:

N	Prozent	SPSS	Statistik	Quantitativ	Qualitativ
1	0.06	1	1	1	1
1	0.06	1	1	0	1
2	0.11	0	1	0	1
3	0.17	0	0	1	1
7	0.40	1	0	0	1
13	0.74	1	0	1	0
18	1.03	1	1	1	0
26	1.49	0	1	1	0
28	1.60	1	1	0	0
34	1.95	0	0	0	1
80	4.58	1	0	0	0
93	5.33	0	1	0	0
97	5.56	0	0	1	0
1342	76.91	0	0	0	0

Münchner Mietspiegel

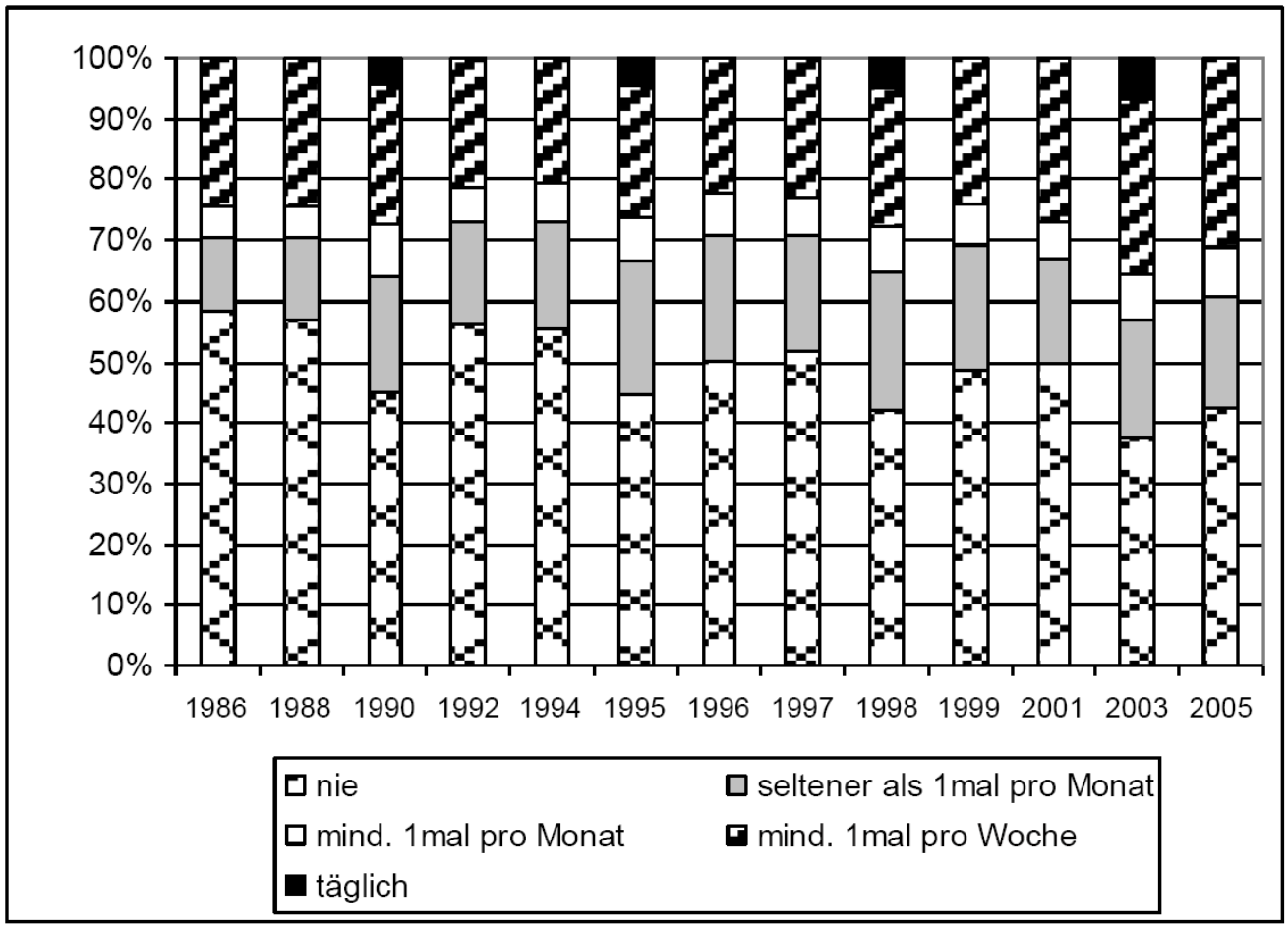
- Mietspiegel bieten Mietern und Vermietern eine Übersicht zu den sogenannten „ortsüblichen Vergleichsmieten“.
- Ortsüblichen Vergleichsmiete: „die üblichen Entgelte, die in der Gemeinde X für nicht preisgebundenen Wohnraum vergleichbarer Art, Größe, Beschaffenheit und Lage in den letzten vier Jahren vereinbart (. . .) oder geändert worden sind.“
- Statistische Fragestellung: Wie beeinflussen Merkmale einer Wohnung (Wohnfläche, Baujahr, Küchenausstattung, etc.) die Nettomiete (pro Quadratmeter)?
- Den aktuellen Mietspiegel für München finden Sie unter

http://www.mietspiegel-muenchen.de

Sozio-ökonomisches Panel (SOEP)

- Seit 1984 durchgeführte Befragung von deutschen Haushalten
- 2007 waren etwa 12.000 (repräsentativ ausgewählte) Haushalte mit mehr als 20.000 Befragungspersonen beteiligt
- Themenschwerpunkte: Haushaltszusammensetzung, Erwerbs- und Familienbiographie, Erwerbsbeteiligung und berufliche Mobilität, Einkommensverläufe, Gesundheit und Lebenszufriedenheit
- Besonderheiten:
  - Die gleichen Personen werden wiederholt befragt (Panelstudie)
  - Befragung auf Haushaltsebene
  - Freiwillige Teilnahme
  - Werden gegen Aufwandsentschädigung Forschern zur Verfügung gestellt

Abbildung 1: Häufigkeit sportlicher Aktivität in den Jahren 1986 bis 2005

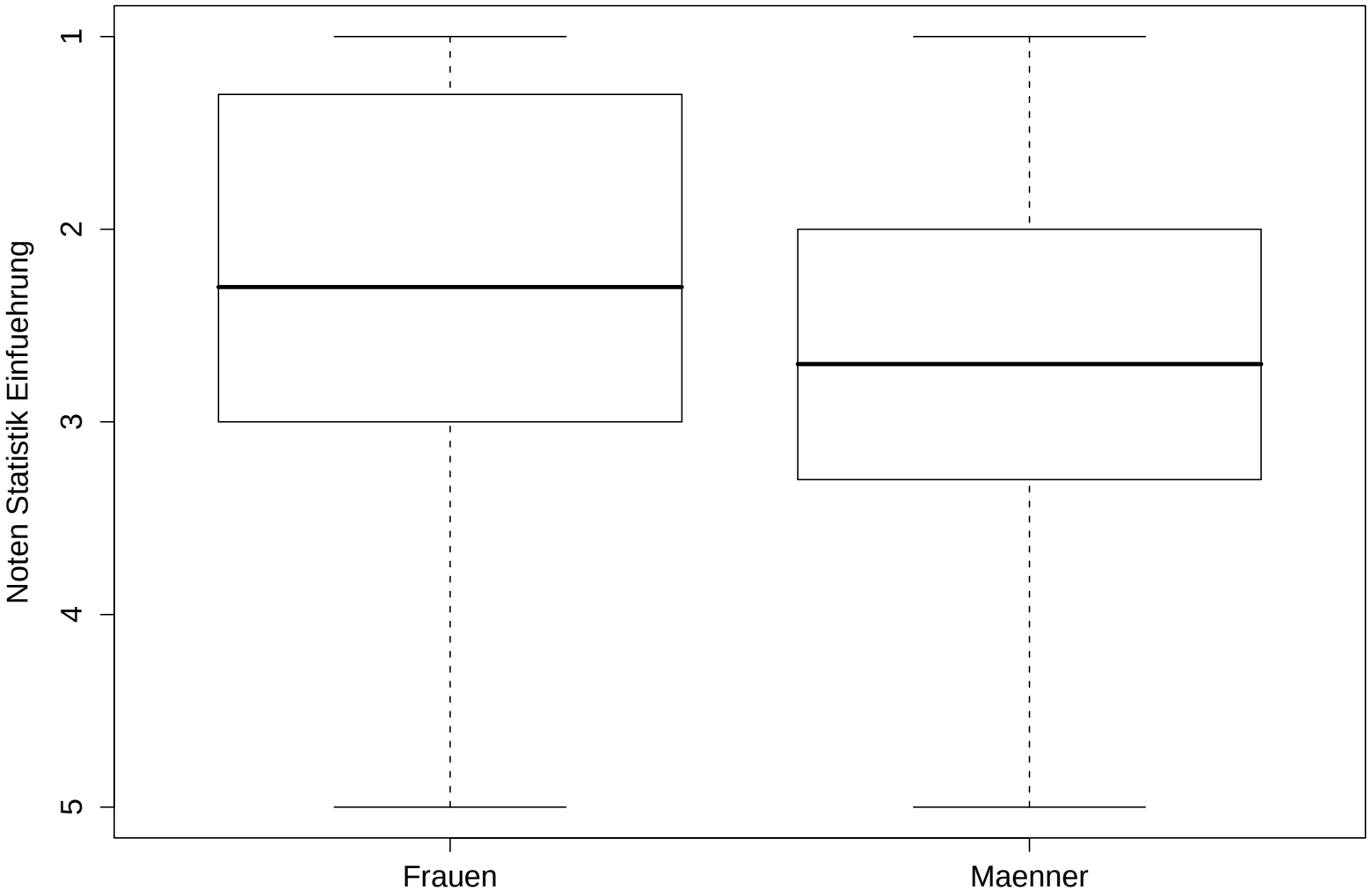


Quelle: SOEP 1986, 1988, 1990, 1992, 1994, 1995, 1996, 1997, 1998, 1999, 2001, 2003, 2005, eigene Berechnungen

Wer hat Angst vor Statistik?

- Umfrage in Statistik Einführungs-Vorlesungen für Soziologen, Psychologen, BWLer etc. im WS 06/07.
- Ergebnisse:
  - weibliche Studierende
  - Studierende, die in der Schule schon Angst vor Mathematik hatten,
  - Studierende, die sich falsch auf Klausuren vorbereiten,

  haben an der Uni eher Angst vor Statistik
  - Studierende, die versuchen den Stoff auswendig zu lernen, haben mehr Angst als Studierende, die den Stoff anhand vieler Aufgaben einüben.


Noten Statistik Einfuehrung
1
2
3
4
5
Frauen
Maenner

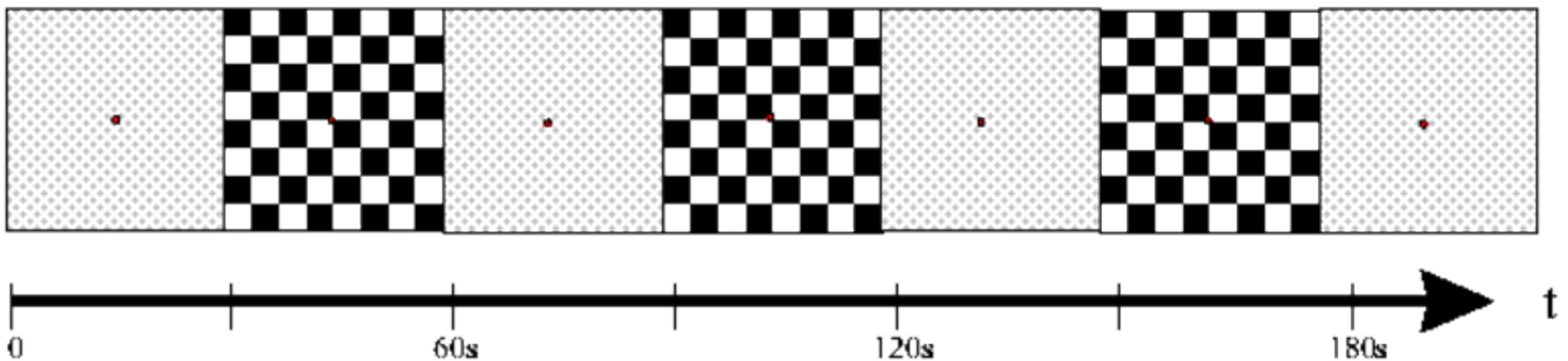
weitere Beispiele

- Waldschadensdaten
- Auswirkungen von Luftverschmutzung / Sendemasten
- Ernährungsgewohnheiten und Herzkreislauferkrankungen
- Strahlenbelastung (z.B. am Arbeitsplatz) und Krebs
- KfZ-Unfälle
- Statistische Genetik
- Wirksamkeit eines Medikaments
- Qualitätskontrolle: Wann läuft ein Prozess aus dem Ruder?
- Geschlechtsunterschiede bei innerbetrieblicher Mobilität
- Gehirnkartierung

Human Brain Mapping (Gehirnkartierung)

- Ziel: Identifikation von Regionen im Gehirn, die an der Erfüllung bestimmter Aufgaben beteiligt sind (z.B. das Sehzentrum).

- Experiment mit visuellem Stimulus:
  - Abwechselnd Phasen mit und ohne Stimulus.
  - Dauer einer Phase jeweils 30 Sekunden.
  - Die Gehirnaktivität wird alle drei Sekunden an 128×128×7 Voxeln gemessen.

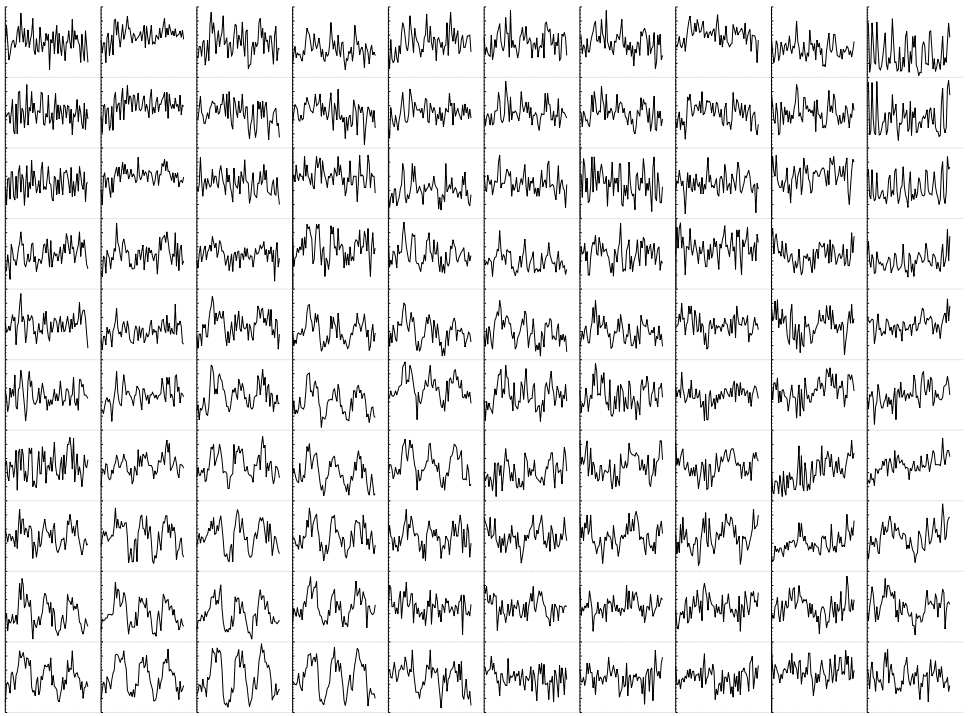


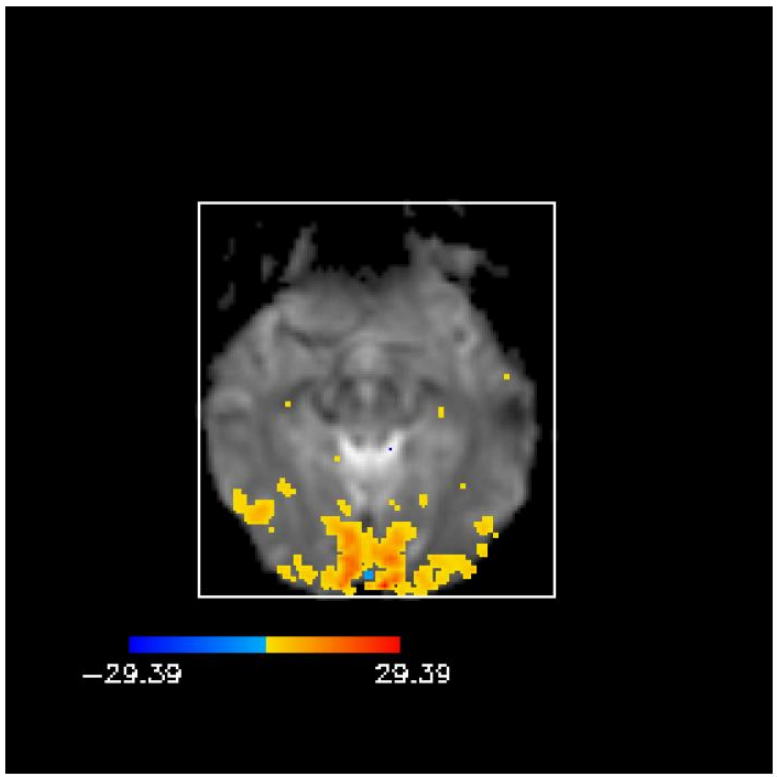
- Aktivierung wird durch funktionelle Magnetresonanztomografie (fMRT) gemessen.

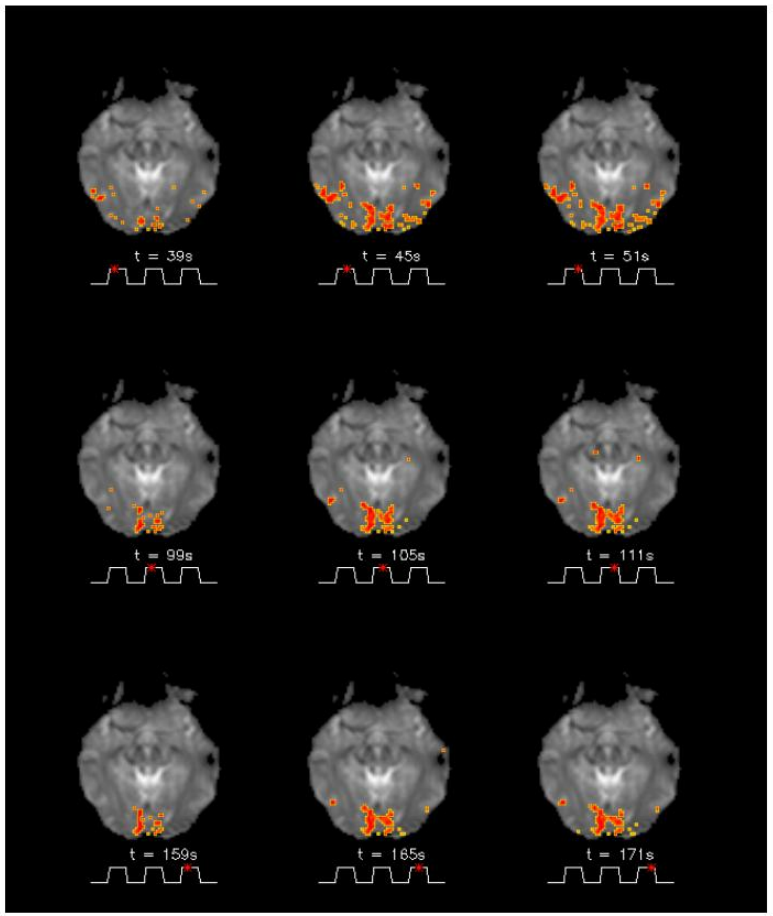
- Die Messungen werden durch zufällige Fehler überlagert (Patient bewegt sich, ist unkonzentriert, Messungenauigkeit, . . .).

- Rolle der Statistik: Trennung von Signal und Rauschen.

- Aktivierung an einigen Voxeln:




−29.39
29.39


t = 39s
t = 45s
t = 51s
t = 99s
t = 105s
t = 111s
t = 159s
t = 165s
t = 171s

1.2 Hauptgebiete der Statistik als Methodenlehre

Deskriptive / explorative Analyse

- Analyse der Daten der *konkret vorliegenden* Gesamtheit (*keine Verallgemeinerung beabsichtigt*).
- Deskription = Beschreibung (durch Tabellen, Kennzahlen, . . .)
- Informationsgewinn durch Verdichtung (Wald vor lauter Bäumen sehen)
- Aufspüren von Zusammenhängen, Hypothesen*generierung* (keine Prüfung!)
- Data Mining

Induktive Statistik = Statistische Inferenz

- Schluss von einer Stichprobe auf die dahinterstehende Grundgesamtheit

- Die Ergebnisse der Stichprobe sind nur Mittel zum Zweck des verallgemeinernden Schlusses

- Schlüsse vom Teil auf das Ganze sind zwangsläufig potentiell fehlerhaft. Man kann diesen Fehler („Inferenzfehler“) nicht ausschalten (Induktionsproblem), $aber$ unter Umständen kontrollieren.
  „Trick“: Ziehe die Stichprobe zufällig (Wahrscheinlichkeitsauswahl), dann kann man die Wahrscheinlichkeit von groben Fehlschlüssen berechnen.

Zur Abschätzung des Inferenzfehlers dient die

Wahrscheinlichkeitsrechnung

- Mathematische Theorie zur Beschreibung unsicherer / zufälliger Phänomene.
- unverzichtbare Voraussetzung für induktive Statistik
- Statistische Modellierung (z.B. Modelle sozialer Mobilität).

Methodologie der Datengewinnung

- Stichprobendesigns:
  Wie gewinnt man geeignete Stichproben?

- Konzipierung des Erhebungsinstruments: Wie erhebt man was?
  Operationalisierung komplexer Konstrukte (z.B. Integrationsfähigkeit), Gestaltung des Instruments (z.B. Techniken der Fragebogenerstellung)

- Datenproduzenten
  - Amtliche Statistik (durch statistische Ämter (Europa, Bund, Länder, teilweise Städte), Unterstützung politischer und wirtschaftlicher Entscheidungen, auf gesetzlicher Grundlage (informationelle Selbstbestimmung, dafür aber meist Auskunftspflicht), meist rein deskriptive Analyse)
  - freiwillige Umfragen auf Stichprobenbasis (wissenschaftlich, kommerziell, teilweise auch von Behörden/Städten)

Quellen der Unsicherheit beim statistischen Schließen

- Kausalität vs. Zusammenhang
- Messfehler und Messungenauigkeit
- Stichproben

1.3 Überblick über die Lehrveranstaltung

I. Descriptive und explorative Statistik

- Grundbegriffe
- Beschreibung eindimensionaler Merkmale
  z.B Verteilung von Noten, Erwerbseinkommen, der Parteienpräferenz, der formalen Bildung
- Beschreibung mehrdimensionaler Merkmale
  Wie stark hängt/hängen
  - die Noten von den Mathematikkentnissen ab
  - das Erwerbseinkommen vom Geschlecht ab
  - die Gewaltbereitschaft von der Parteienpräferenz ab
  - die formale Bildung von der Schichtzugehörigkeit der Eltern ab?

II. Wahrscheinlichkeitsrechnung = Mathematische Modellierung und Analyse des Zufälligen / Unsicheren

III. Induktive Statistik

- Punktschätzung: Wie groß ist der Anteil der Rot-Grün-Wähler in der Grundgesamtheit?
- Intervallschätzung: Innerhalb welcher Grenzen ist der wahre Wert mit hoher Sicherheit?
- Hypothesentest: Kann man aus der Stichprobe schließen, dass Frauen weniger verdienen als Männer?
- Ausblick auf komplexere statistische Verfahren

Statistik I : Deskriptive und explorative Statistik

Gliederung der Vorlesung :

1. Einführung.
2. Häufigkeitsverteilungen.
3. Lage- und Streuungsmaße.
4. Konzentrationsmaße.
5. Analyse von Zusammenhängen.
6. Regression.

1.4 Grundbegriffe

1.4.1 Statistische Einheiten und Gesamtheiten

Statistische Einheiten: Objekte an denen interessierende Größen erhoben werden.

Grundgesamtheit: Die Menge aller für eine bestimmte Fragestellung relevanten statistischen Einheiten heißt *Grundgesamtheit* (Universum, Population). Die Grundgesamtheit muss durch sachliche, räumliche und zeitliche Kriterien exakt festgelegt sein. Die Abgrenzungskriterien richten sich nach dem Untersuchungsziel.

Vollerhebung : Eine Erhebung der interessierende Größen von allen Objekte der Grundgesamtheit nennt man Vollerhebung.

Stichprobe: Ist eine Vollerhebung nicht sinnvoll, nicht möglich, oder zu teuer, dann wird nur eine Auswahl der Grundgesamheit, eine sogenannte Stichprobe untersucht.

Gesamtheit: Im Rahmen der deskriptiven Statistik wird keine Verallgemeinerung der aus der Stichprobe gewonnenen Ergebnisse auf die Grundgesamtheit angestrebt. Es ist also nicht nötig, zu unterscheiden, ob die zu analysierenden Daten aus einer Stichprobe stammen oder bereits die Grundgesamtheit darstellen. Wir sprechen dann einfach von einer $Gesamtheit$ von statistischen Einheiten, die analysiert werden sollen.

Notation:

- In einer Gesamtheit mit n Elementen ($Stichprobenumfang$, $bzw.$ $Umfang$ der $Gesamtheit$) werden die statistischen Einheiten mit $\omega_1, \omega_2, \ldots, \omega_n$ bezeichnet.
- Gesamtheit $\Omega = \{\omega_1, \ldots, \omega_n\}$.
- Bezieht man sich auf ein festes, aber beliebiges Element der Grundgesamtheit, so schreibt man meist ω (ohne Index).

1.4.2 Merkmale und Merkmalsausprägungen

Merkmale: Inhaltlich interessant sind nicht die Einheiten an sich, sondern bestimmte Eigenschaften oder $Merkmale$ der Einheiten (Variablen).

Merkmalsausprägung: Der Wert eines Merkmals für eine konkret vorliegende statistische Einheit.

Wertebereich: Alle prinzipiell möglichen Ausprägungen eines Merkmals.

Notation: Merkmale werden typischerweise mit Großbuchstaben bezeichnet (X, Y, Z, etc.), Ausprägungen mit dem zugehörigen Kleinbuchstaben (x, y, z). Der Wertebereich wird mit W bezeichnet.

Formal ist jedes Merkmal eine Funktion: jedem Element der Gesamtheit wird ein Wert zugeordnet.

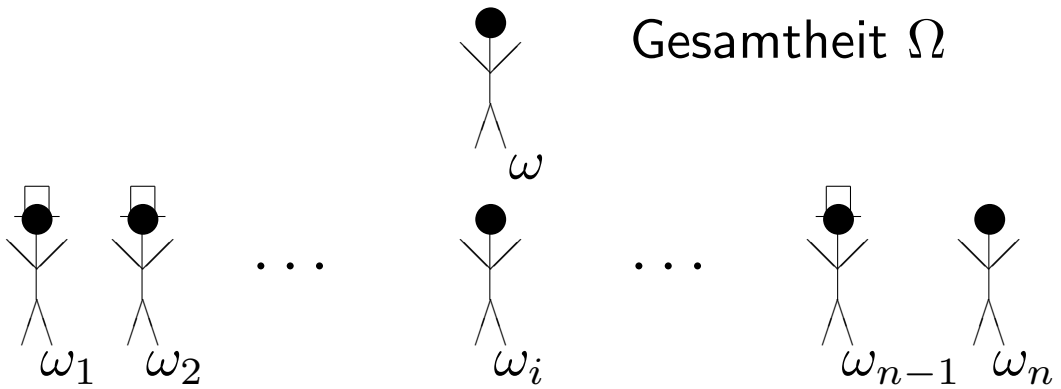
Schreibweisen:

- $X(\omega)$ Merkmalsausprägung der Einheit $\omega \in \Omega$
- $X(\omega) = x$
- $X(\omega_i) = x_i$

Die Elemente von W werden wir später mit $a_1, \ldots, a_k$ bezeichnen.

Teilmengen der Gesamtheit: Oft braucht man auch die Menge aller Einheiten, bei denen X einen bestimmten Wert, üblicherweise mit x bezeichnet, annimmt.

$$\{ \quad \omega \in \Omega \quad | \quad X(\omega) \quad = \quad x \quad \}$$


Gesamtheit Ω
ω
ω_1
ω_2
ω_i
ω_{n-1}
ω_n

Beispiel: Mietspiegel

- Statistische Einheiten
- Grundgesamtheit: Nicht preisgebundene Mietwohnungen, für die in den letzten vier Jahren die Miete neu vereinbart oder geändert wurde.
- Die Gesamtheit der statistischen Einheiten ist eine Stichprobe von ca. 3000 Wohnungen der Grundgesamtheit.
- Merkmale und Wertebereich
- Merkmalsausprägungen für die 713-te Wohnung im Datensatz
- Teilmengen (mit Merkmalen $X =$ Nettomiete pro QM, $Y =$ Wohnfläche):

1.4.3 Merkmalstypen

Eine adäquate statistische Analyse hängt entscheidend davon ab, welche Eigenschaften die möglichen Merkmalsausprägungen haben.

Merkmale werden unterschieden

- nach der *Zahl der Ausprägungen*: diskret, quasi-stetig, stetig
- nach der *Struktur des Wertebereiches*: nominal, ordinal, Intervallskala, Verhältnisskala
- nach *Art der Ausprägungen*: qualitativ, quantitativ

Diskrete und stetige Merkmale

- *Diskret*: Das Merkmal kann nur endlich viele (oder abzählbar viele) Ausprägungen annehmen.

- *Stetig*: Das Merkmal kann (im Prinzip) alle Werte in einem Intervall annehmen (überabzählbar viele verschiedene Ausprägungen).

- *Quasi-stetig*: Zwischenform. Jede Messung hat nur endliche Genauigkeit, ist also eigentlich diskret, kann aber als stetig behandelt werden.

- *Kategorisierung*: Oft werden stetige Daten auch absichtlich diskretisiert, nämlich bei Gruppenbildung (gruppieren, klassieren, kategorisieren).

Skalenniveau

Das Skalenniveau eines Merkmals ergibt sich aus der Struktur des Wertebereiches.

- *Nominalskala*: Ein Merkmal heißt nominalskaliert, wenn die Ausprägungen Namen oder Kategorien sind, die keine natürliche Ordnung haben.

  Häufig werden (etwa zur Datenanalyse am PC) den Ausprägungen Zahlen zugeordnet. Diese Zahlen sind aber nur Stellvertreter ohne inhaltliche Bedeutung; ihre Zuordnung kann - solange sie eindeutig ist - völlig willkürlich erfolgen.

- *Ordinalskala*: Ein Merkmal heißt ordinalskaliert, wenn sich die Ausprägungen ordnen lassen.

  Man kann beliebige Zahlen zuordnen, solange diese die Ordnung erhalten. Die Abstände der Merkmalsausprägungen lassen sich nicht sinnvoll interpretieren.

- *Intervallskala*: Ein Merkmal heißt intervallskaliert, wenn die Abstände der Merkmalsausprägungen sinnvoll interpretiert werden können.
- *Verhältnisskala / Ratioskala*: Ein Merkmal heißt verhältnisskaliert, wenn es intervallskaliert ist und zusätzlich ein sinnvoll interpretierbarer Nullpunkt existiert.

Verhältnisskala und Intervallskala werden oft zur *Kardinalskala* zusammengefasst. Ein kardinalskaliertes Merkmal wird auch als *metrisch* bezeichnet. Metrische Merkmale sind oft stetig oder quasi-stetig (z.B. Größe, Einkommen), können aber auch diskret sein (z.B. Anzahlen).

Sinnvolle Operationen

Je nach Skalenniveau sind unterschiedliche Operationen sinnvoll:

Skala	Häufigkeiten	Größenvergleich	Differenz	Quotienten bilden
Nominalskala				
Ordinalskala				
Intervallskala				
Verhältnisskala				

Alle Operationen, die auf einer Nominalskala sinnvoll sind, sind auch auf der Ordinalskala sinnvoll, aber nicht umgekehrt!

Das Skalenniveau eines Merkmals bestimmt, welche statistischen Verfahren sinnvoll angewendet werden können.

Zulässige Transformationen

Mathematisch exakt charakterisiert man Skalen über die Transformationen, die man durchführen darf, ohne die inhaltliche Struktur zu zerstören, d.h. vor und nach der Transformation sollen die für die jeweilige Skala grundlegenden Operationen jeweils dieselben inhaltliche Ergebnisse liefern.

Transformation („Umrechnung“):

	Transformation
Nominalskala	eineindeutige
Ordinalskala	streng monotone
Intervallskala	linear affine $(a + bX;\ b > 0)$
Verhältnisskala	lineare $(bX;\ b > 0)$

Damit bleiben auf der Intervallskala Verhältnisse von Differenzen gleich, und auf der Verhältnisskala Verhältnisse.

Beispiel: Linear affine Transformation:

$$y = a + bx$$

Verhältnis der Differenz $y_1 - y_2$ zu $y_3 - y_4$:

$$\frac{y_1 - y_2}{y_3 - y_4} =$$

Qualitative und quantitative Merkmale

- Qualitativ: Das Merkmal beschreibt eine Eigenschaft / eine Qualität und kein Ausmaß. Das Merkmal besitzt nur endlich viele Ausprägungen und ist nominal- oder ordinalskaliert.
- Quantitativ: Das Merkmal gibt messbar ein Ausmaß wieder. Das Merkmal ist sinnvoll in Zahlen messbar und intervall- oder verhältnisskaliert.

1.4.4 Erhebungsformen

Experiment vs. Beobachtungsdaten

- Experiment: Die Daten werden gezielt erzeugt. Insbesondere können die interessierenden Größen direkt beeinflusst werden.

- Beobachtungsdaten: Die Daten sind prinzipiell bereits vorhanden und müssen nur noch „beobachtet“ werden.
  Kontrolle von Störgrößen ist in der Analyse notwendig!

Vollerhebung vs. Stichprobe

- Vollerhebung: Alle statistischen Einheiten der Grundgesamtheit werden untersucht.
- Stichprobe: Nur ein Teil der Grundgesamtheit wird untersucht. Dieser soll möglichst repräsentativ für die Grundgesamtheit sein.
- Gründe für Stichproben:
  - Geringerer Aufwand.
  - Vollerhebung nicht möglich (z.B. in der Qualitätskontrolle).

Auswahltechniken

- Einfache Zufallsstichprobe,
- Klumpenstichprobe,
- Geschichtete Stichprobe.

Studientypen

- Querschnittsstudie: An einer Menge von Einheiten werden zu einem Zeitpunkt mehrere Merkmale erhoben.
- Zeitreihe: Ein Merkmal wird wiederholt zu verschiedenen Zeitpunkten erhoben.
- Longitudinal- / Paneldaten: An einer festen Menge von statistischen Einheiten werden wiederholt (die gleichen) Variablen erhoben.

Analysearten

- Primärerhebung / -analyse:
  Daten werden im Rahmen des Forschungsprojekts erhoben und analysiert.
- Sekundäranalyse:
  Analyse von im Rahmen anderer Forschungsprojekte erhobener Daten.
- Tertiäranalyse:
  Analyse von aggregierten (zusammengefassten) Daten.
- Metaanalyse:
  Sekundäranalyse oder Tertiäranalyse (= Metaanalyse im engeren Sinn) von mehreren Studien.