

Leuphana Universität Lüneburg

---

---

# Statistik I

## Deskription

---

---

Fakultät W

Wirtschaftswissenschaften

Professur 'Statistik und Freie Berufe'

Univ.-Prof. Dr. Joachim Merz

Skriptum zur Vorlesung

Zwölfte Auflage 2014

Impressum: Statistik I - Deskription, Skriptum zur Vorlesung,  
herausgegeben von der Leuphana Universität Lüneburg,  
Fakultät W - Wirtschaftswissenschaften  
Univ.-Prof. Dr. Joachim Merz, Forschungsinstitut Freie Berufe,  
Professur 'Statistik und Freie Berufe'.  
Campus, Scharnhorststraße 1, Gebäude 4, 21335 Lüneburg  
E-Mail: [ffb@uni.leuphana.de](mailto:ffb@uni.leuphana.de)  
[www.leuphana.de/ffb](http://www.leuphana.de/ffb)

Gedruckt auf 100 % Altpapier, chlorfrei gebleicht.  
Copyright © 2014

## **Vorwort**

Statistik I - Deskription ist Thema dieses Skriptums zu meiner gleichlautenden Vorlesung an der Universität Lüneburg.

Das Skriptum soll vorlesungsbegleitend helfen, den Blick auf das Wesentliche, auf das Verständnis der Methoden und ihrer Anwendungen zu erleichtern. Im Rahmen einer anwendungsorientierten Statistik stehen bei der Auswahl des Stoffes Beispiele und Bezüge aus den Wirtschafts- und Sozialwissenschaften im Vordergrund. Die jeweils aktuellen Tabellen und Abbildungen sollen zugleich Größenordnungen wichtiger gesellschaftlicher Bereiche vermitteln.

Ich empfehle, den Stoff mit der angegebenen Literatur zu vertiefen: Manchmal hilft ein anderer Blickwinkel, die Dinge besser zu begreifen. Das Verstehen, das verständige Umgehen mit Statistik als ein wesentlicher Baustein, Theorie mit Empirie zu verbinden, ist mir ein wichtiges Anliegen.

Für den problemorientierten Einstieg und den Umgang mit dem Computer als Hilfsmittel werden SPSS (Statistical Package for the Social Sciences) sowie ET (Econometrics Toolkit) verwendet. SPSS ist das wohl weitverbreitetste Statistikpaket und wie ET menuegesteuert leicht zu bedienen; ET ist zudem gute Grundlage für weitere Programmanwendungen (bspw. für LIMDEP) im Rahmen einer empirischen Wirtschaftsforschung.

Hinsichtlich der neuen redaktionellen Überarbeitung des Skriptums gilt mein herzlicher Dank Frau Lena Krüger und Herrn Dipl.-Kfm. Tim Rathjen.

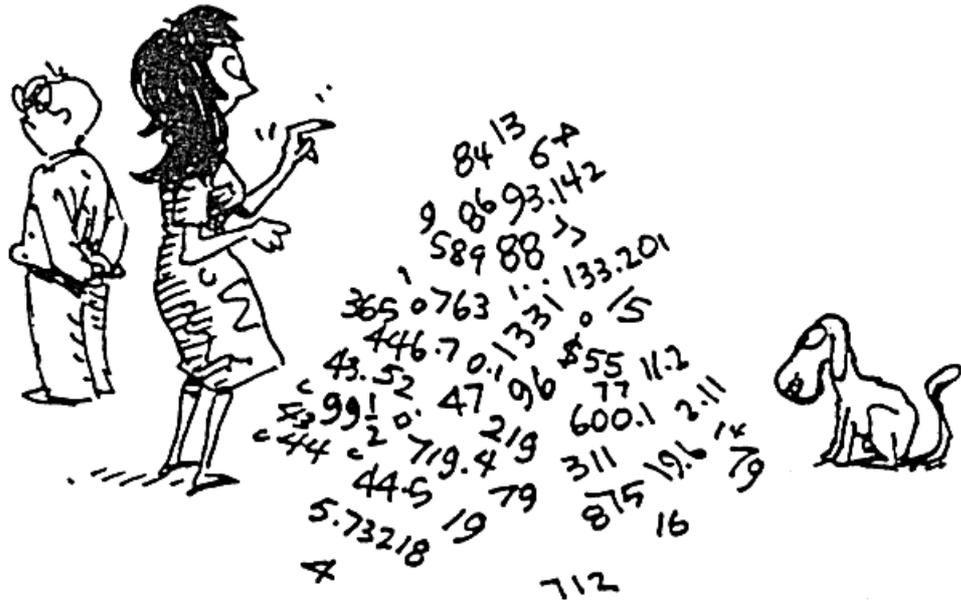
Nicht zu vergessen: Studium und späterer Beruf sollen auch Spaß machen. Die Cartoons im Skriptum sind entsprechende Lockerungsübungen.

Viel Spaß und Erfolg!

Lüneburg, im August 2014

Univ.-Prof. Dr. Joachim Merz

UM...  
WELL... IT'S...  
THEY'RE...  
AHEM?  
COUGH...



Statistics is ...

... to compress information

... to quantify uncertainty

Gonick, Smith (1993)

**STATISTIK I - DESKRIPTION**

**THEMENBEREICHE**

**I EINFÜHRUNG UND ALLGEMEINE GRUNDLAGEN**

**II WIRTSCHAFTS- UND SOZIALSTATISTIK**

**III STATISTISCHE ANALYSE EINES EINZELNEN MERKMALS**

**IV STATISTISCHE ANALYSE MEHRERER MERKMALE**

**V VERHÄLTNIS- UND INDEXZAHLEN**

**VI VOLL- UND TEILERHEBUNGEN**

**VII ZEITREIHEN- UND QUERSCHNITTSANALYSE**

**VIII COMPUTERPROGRAMME ZUR STATISTISCHEN  
AUSWERTUNG UND ANALYSE**

**FORMELSAMMLUNG**

**LITERATUR**

THIS IS NOT JUST A MATTER OF ORDERING SOUP! STATISTICS ALSO INVOLVES MATTERS OF LIFE AND DEATH...

HEY—HAVE YOU EVER HAD THE SOUP HERE ON AN OFF NIGHT?

FOR EXAMPLE, IN 1986, THE SPACE SHUTTLE CHALLENGER EXPLODED, KILLING SEVEN ASTRONAUTS. THE DECISION TO LAUNCH IN 29-DEGREE WEATHER HAD BEEN MADE WITHOUT DOING A SIMPLE ANALYSIS OF PERFORMANCE DATA AT LOW TEMPERATURE.

OH... THAT PART OF THE CURVE!!

A MORE POSITIVE EXAMPLE IS THE SALK POLIO VACCINE. IN 1954, VACCINE TRIALS WERE PERFORMED ON SOME 400,000 CHILDREN, WITH STRICT CONTROLS TO ELIMINATE BIASED RESULTS. GOOD STATISTICAL ANALYSIS OF THE RESULTS FIRMLY ESTABLISHED THE VACCINE'S EFFECTIVENESS, AND TODAY POLIO IS ALMOST UNKNOWN.

MOST OF US LIVE COMFORTABLY WITH SOME LEVEL OF UNCERTAINTY.

I'LL HAVE THE SOUP, PLEASE.

AARRH... COULD YOU JUST BRING ME A CALCULATOR?

WHAT MAKES STATISTICS UNIQUE IS ITS ABILITY TO QUANTIFY UNCERTAINTY, TO MAKE IT PRECISE. THIS ALLOWS STATISTICIANS TO MAKE CATEGORICAL STATEMENTS, WITH COMPLETE ASSURANCE—ABOUT THEIR LEVEL OF UNCERTAINTY!

GOOD CHOICE! I'M 95% CONFIDENT THAT TONIGHT'S SOUP HAS PROBABILITY BETWEEN 73% AND 77% OF BEING REALLY DELICIOUS!

Univ.-Prof. Dr. Joachim Merz

# STATISTIK I – DESKRIPTION

## GLIEDERUNG

<b>I</b>	<b>Einführung und allgemeine Grundlagen.....</b>	<b>3</b>
1	Einführende Beispiele .....	3
1.1	Anwendungsorientierte Statistik: Mikroanalyse der individuellen Wirkungen der Steuerreform 1990 - Mikroökonomische Theorie, Mikrodatenbasis, Mikroökonomie und Mikrosimulation.....	3
1.2	Statistik und EDV: Deskription und Inferenz am Beispiel des Programmpakets SPSS (Statistical Package for the Social Sciences) .....	4
1.3	Problemorientierte Statistik: Studien zur aktuellen und zukünftigen Situation der Erde – Club of Rome und Intergovernmental Panel on Climate Change (IPCC) .....	5
1.4	Anwendungsorientierte Statistik: Fragen zur Wohnsituation aus dem Sozio-Ökonomischen-Panel (SOEP), Leben in Deutschland, Befragung 2007 zur sozialen Lage der Haushalte .....	9
2	Begriff, Aufgaben und Entwicklung der Statistik .....	10
2.1	Begriff und Aufgaben der Statistik .....	10
2.2	Zur geschichtlichen Entwicklung .....	11
3	Träger der Wirtschaftsstatistiken und statistische Quellen .....	13
3.1	Amtliche Statistik.....	13
3.2	Nichtamtliche Statistik.....	17
3.3	Internationale Organisationen.....	18
3.4	Neue Aufgaben und Quellen der Wirtschaftsstatistik im vereinten Deutschland.....	18
4	Das Adäquationsproblem und einige wissenschaftstheoretische Bemerkungen.....	19
4.1	Wissenschaftstheoretische Grundlagen: Zur Struktur und Anwendung wissenschaftlicher Theorien .....	19
4.2	Das Adäquationsproblem: Allgemeine Problemstellung und statistische Operationalisierung .....	21
5	Sachgerechte Interpretation: 'How (not) to lie with statistics'.....	22
5.1	Some pitfalls .....	22
5.2	How not to lie with statistics.....	23
6	Statistische Einheiten und statistische Massen.....	23
6.1	Statistische Einheiten .....	23
6.2	Statistische Massen .....	24
7	Merkmale, Merkmalsausprägungen und Meßskalen.....	24
7.1	Merkmale und Merkmalsausprägungen.....	24

7.2	Meßskalen und ihre Eigenschaften .....	25
7.3	Diskrete und stetige Merkmale .....	26
7.4	Quantitative und qualitative Merkmale .....	26
8	Statistische Untersuchungen: Erhebung, Aufbereitung und Analyse .....	27
8.1	Vorgehensweise bei statistischen Untersuchungen .....	27
8.2	Erhebung: Erhebungsarten und Erhebungstechnik .....	28
8.3	Aufbereitung und Analyse .....	29
9	Tabellarische und grafische Darstellung .....	30
9.1	Zur Präsentation von Informationen .....	30
9.2	Tabellenaufbau und grafische Darstellung .....	33
10	Datenschutz und Datensicherheit .....	35
<b>II</b>	<b>Wirtschafts- und Sozialstatistik.....</b>	<b>37</b>
1	Zur Aufgabe und Bedeutung der Wirtschafts- und Sozialstatistik.....	37
2	Bevölkerung .....	40
2.1	Aufgaben und Abgrenzung .....	42
2.2	Statistik des Bevölkerungsbestandes .....	42
2.3	Statistik der Bevölkerungsbewegung.....	57
2.4	Bevölkerungsprognosen: Demografische Vorausschätzungen und computergestützte mikroanalytische Simulationsverfahren .....	64
3	Produktion .....	69
3.1	Aufgabe, Abgrenzung und Messung.....	69
3.2	Das Erhebungssystem der Produktionsstatistik .....	70
3.3	Indizes der industriellen Produktion und Produktionsmessung.....	73
3.4	Produktivitätsmessung .....	75
4	Dienstleistung .....	76
4.1	Allgemeine Bedeutung.....	76
4.2	Handel, Gastgewerbe, Kreditinstitute .....	76
4.3	Verkehr, Freie Berufe .....	79
5	Erwerbstätigkeit.....	82
5.1	Wichtige arbeitsmarktstatistische Erhebungen .....	82
5.2	Konzepte zur Erfassung der Erwerbstätigkeit.....	90
5.3	Indikatoren und Maßzahlen .....	92
6	Einkommens- und Verbrauchsstatistik, Preise .....	96
6.1	Einkommens- und Verbrauchsstatistik .....	96
6.2	Preisstatistik .....	99
7	Umwelt .....	101
8	Volkswirtschaftliche Gesamtrechnung, Soziale Indikatoren .....	106
8.1	Volkswirtschaftliche Gesamtrechnung (VGR) .....	106
8.2	Soziale Indikatoren .....	109
<b>III</b>	<b>Statistische Analyse eines einzelnen Merkmals.....</b>	<b>118</b>
1	Eindimensionale Häufigkeitsverteilungen und ihre Darstellung.....	118
1.1	Häufigkeitsverteilung nominalskaliert (qualitativer) Merkmale .....	118
1.2	Häufigkeitsverteilung metrisch skaliert, diskreter Merkmale.....	124
1.3	Häufigkeitsverteilung metrisch skaliert (quantitativer) stetiger Merkmale .....	127
1.4	Computergestützte grafische Darstellung .....	130
2	Lageparameter .....	132
2.1	Häufigster Wert (Modus) .....	132

2.2	Median (Zentralwert) .....	132
2.3	Arithmetisches Mittel.....	136
2.4	Geometrisches Mittel .....	141
2.5	Harmonisches Mittel .....	141
3	Streuungsmaße .....	142
3.1	Spannweite .....	143
3.2	Quartilsabweichung und p-Quantile .....	144
3.3	Mittlere absolute Abweichung.....	148
3.4	Mittlere quadratische Abweichung: Varianz und Standardabweichung .....	149
3.5	Variationskoeffizient.....	153
3.6	Konzept der Momente, Schiefe und Exzeß.....	155
4	Konzentration einer Verteilung .....	162
4.1	Konzentration.....	162
4.2	Lorenzkurve und Gini-Koeffizient .....	165
<b>IV</b>	<b>Statistische Analyse mehrerer Merkmale.....</b>	<b>171</b>
1	Zweidimensionale Häufigkeitsverteilungen und ihre Darstellung .....	171
1.1	Allgemeine Grundbegriffe und Darstellungsweisen.....	171
1.2	Randverteilungen .....	173
1.3	Bedingte Verteilungen .....	174
2	Korrelationsrechnung .....	177
2.1	Zusammenhangsmaße.....	177
2.2	Korrelation zwischen nominal skalierten Merkmalen: Kontingenzanalyse und Kontingenzkoeffizient.....	177
2.3	Korrelation zwischen ordinal-skalierten Merkmalen: Rangkorrelationskoeffizient nach Spearman .....	179
2.4	Korrelation zwischen metrisch-skalierten Merkmalen: Bravais-Pearson-Korrelationskoeffizient .....	180
3	Regressionsrechnung .....	186
3.1	Lineare Regression.....	187
3.2	Nichtlineare Regression .....	198
4	Zur Analyse drei- und mehrdimensionaler Häufigkeitsverteilungen .....	199
<b>V</b>	<b>Verhältnis- und Indexzahlen.....</b>	<b>201</b>
1	Verhältniszahlen .....	201
1.1	Gliederungszahlen.....	201
1.2	Beziehungszahlen .....	203
1.3	Meßziffern.....	203
1.4	Meßziffernreihe.....	204
2	Indexzahlen .....	205
2.1	Preisindizes: Laspeyres und Paasche .....	205
2.2	Mengenindizes .....	209
2.3	Umsatzindizes (Wertindizes).....	210
3	Umbasierung, Verknüpfung und Preisbereinigung von Indizes .....	210
<b>VI</b>	<b>Voll- und Teilerhebungen.....</b>	<b>215</b>
1	Allgemeine Bemerkungen und Beispiele aus der amtlichen und nichtamtlichen Statistik .....	215
2	Auswahlverfahren .....	215

3	Stichprobenerhebungen .....	216
3.1	Einfaches Stichprobenverfahren .....	216
3.2	Geschichtete Stichprobenverfahren .....	216
3.3	Klumpenstichprobe .....	217
3.4	Mehrstufige Stichprobenverfahren .....	217
4	Das ADM - Stichprobenverfahren: Ein Beispiel aus der privaten Umfrageforschung .....	217
5	Hochrechnung von Teilerhebungen .....	218
<b>VII</b>	<b>Zeitreihen- und Querschnittsanalyse .....</b>	<b>221</b>
1	Charakteristika und Komponenten einer Zeitreihe.....	221
2	Bestimmung der Trend- und Zykluskomponente.....	222
2.1	Methode der kleinsten Quadrate .....	222
2.2	Methode der gleitenden Durchschnitte .....	225
2.3	Exponentielle Glättung .....	226
3	Bestimmung der Saisonkomponente .....	230
3.1	Saisonbereinigung bei konstanter Saisonfigur.....	230
3.2	Saisonbereinigung bei variabler Saisonfigur .....	236
4	Spektralanalyse: Zeitreihenanalyse im Frequenzbereich .....	237
4.1	Spektralanalyse .....	237
4.2	Kreuzspektralanalyse .....	240
5	Querschnittsanalyse .....	243
5.1	Charakteristika und Beispiele .....	243
5.2	Analysemethoden.....	243
6	Panelanalyse .....	244
6.1	Panel- und Längsschnittsanalyse .....	244
6.2	Analysemethoden.....	244
<b>VIII</b>	<b>Computerprogramme zur statistischen Auswertung und Analyse .....</b>	<b>245</b>
1	Statistikauswertungen mit Tabellenkalkulationsprogrammen .....	245
2	SPSS und SAS.....	245
3	ET, LIMDEP, TSP und RATS .....	247
	<b>Formelsammlung.....</b>	<b>249</b>
	<b>Literatur .....</b>	<b>262</b>

## **Statistik – Warum ist sie so wichtig?!**

### **Deskriptive Statistik, Wahrscheinlichkeitsrechnung und induktive Statistik**

Moderne Statistik ist **Informationskomprimierung**. Dazu zählen in erster Linie Ansätze, mit denen eine Vielzahl von Informationen auf zentrale **Indikatoren** und **Kennzahlen** verdichtet werden können (**deskriptive Statistik**). Sind aus Kosten- und anderen Vereinfachungsgründen Erkenntnisse aus **Stichproben** für eine **übergeordnete Grundgesamtheit** zu gewinnen, dann ist es notwendig, etwas über die Signifikanz der Stichprobenergebnisse auszusagen (Wahrscheinlichkeitsrechnung, **schließende Statistik**).

Statistische Informationen (Graphen, Kennzahlen etc.) sind wesentliche Bausteine zur **Untermauerung von Argumenten**, sei es im **beruflichen, politischen** aber auch **privaten Bereich**. Zeitungen und andere gesellschaftlich wichtige Medien bleiben ohne das Verständnis für Zahlen und Graphen der deskriptiven Statistik unverständlich. Vorkenntnisse der deskriptiven Statistik (Mittelwerte, Streuung etc.) sind zudem eine notwendige Voraussetzung für eine darauf aufbauende schließende Statistik (Hypothesentest etc.)

Die ehemalige Bundesministerin für Bildung und Forschung, Edelgard Bulmahn hat in ihrem Vorwort des Gutachtens der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001) die Wichtigkeit von statistischen Informationen betont:

*„Gute politische Entscheidungen brauchen als Grundlage aussagekräftige statistische Informationen zur Situation und Entwicklung von Wirtschaft und Gesellschaft. Nur dann können Sozial- und Wirtschaftswissenschaften treffsichere Analysen erstellen und damit die Handlungsgrundlage für die Politik verbessern.“*

Mit entsprechenden Erhebungen mit einer Fülle von Einzeldaten (Mikrodaten) werden die Informationen für zielkonforme Analysen gewonnen. Das Ziel ist es, aus der Vielzahl der Daten dann wesentliche Informationen wie Trend und durchschlagende Phänomene zu gewinnen. Vor einer multivariaten Analyse, einer Analyse mit konkurrierenden Erklärungsfaktoren steht die zusammenfassende und komprimierende Beschreibung der Situation aus der Gesamtheit der Daten: die Deskription, oder beschreibende Statistik, die im Vordergrund dieses Skriptums steht.

#### **Statistik I - Deskription:**

Beschreibende Statistik mit Verfahren zur Aufbereitung statistischer Daten bezogen auf die beobachteten Werte (Informationsaufbereitung und -verdichtung). Umfasst die Darstellung eines Datenmaterials in Form von Kennzahlen, Tabellen und Grafen.

#### **Statistik II - Wahrscheinlichkeitsrechnung und induktive Statistik:**

Dient der Überprüfung allgemeingültiger Theorien. Informationsbewertung durch Inferenzstatistik (schließende Statistik): Wahrscheinlichkeitsaussagen über die Vereinbarkeit der in den Daten erfassten Realität (Empirie) mit den aus einer Theorie abgeleiteten Hypothesen.

Die Wahrscheinlichkeitsrechnung ist notwendig, um von Teilerhebungen (Stichproben, 'sample') auf eine Grundgesamtheit zu schließen (induktive Statistik).

## **Zum Aufbau von Statistik I - Deskription**

### Einführung und allgemeine Grundlagen

- Beispiele, Begriff und Aufgaben
- Träger der Wirtschaftsstatistik und statistische Quellen
- Adäquationsproblem und sachgerechte Interpretation
- Statistische Einheiten, Massen, Merkmale und Meßskalen

### Wirtschafts- und Sozialstatistik

- Aufgabe und Bedeutung
- Bevölkerungs-, Produktions- und Arbeitsmarktstatistik
- Einkommens- und Verbrauchsstatistik und andere Statistiken
- Volkswirtschaftliche Gesamtrechnungen und Soziale Indikatoren

### Statistische Analyse ein- und mehrdimensionaler Daten

- Eindimensionale Häufigkeitsverteilungen und Lageparameter
- Streuungs- und Verteilungsmaße
- Zweidimensionale Häufigkeitsverteilungen
- Korrelation und Regression

### Verhältnis- und Indexpzahlen, Voll- und Teilerhebungen

- Preis- und Mengenindizes
- Stichprobenerhebungen

### Zeitreihen- und Querschnittsanalyse

- Zeitreihenanalyse: Trend, Zyklus und Saison
- Querschnitts- und Panelanalyse

### Computerprogramme

# I Einführung und allgemeine Grundlagen



*Beispiele, Begriffe, Aufgaben und Quellen sowie statistische Einheiten, Massen und Skalen als Grundlage für die deskriptive Statistik*

## 1 Einführende Beispiele

### 1.1 Anwendungsorientierte Statistik: Mikroanalyse der individuellen Wirkungen der Steuerreform 1990 - Mikroökonomische Theorie, Mikrodatenbasis, Mikroökometrie und Mikrosimulation

#### Mikroökonomisches Modell Multipler Markt- und Nichtmarktmäßiger Aktivitäten Privater Haushalte

- Steuern und Transfers
- Sozioökonomische Charakteristika
- Mikroökonomisches Modell optimaler Zeitallokation

#### Mikrodaten und Merge

##### Berechnung individueller Steuervariablen

- Steuerschuld
- Grenzsteuersätze
- Sozio-ökonomisches Panel 1. Welle 1984
- ESt-/LSt-Statistik 1983
- Steuerrecht 1983/1990

##### Erweiterte Mikrodatenbasis

- Sfb 3 - Nebenerwerbstätigkeitsumfrage 1984 (BfLR)

##### Merge

- Steuervariablen
- Regionale Wirtschafts- und Arbeitsmarktdaten

#### Mikroökometrisches Modell und Schätzung

- 3 stufiges selektionskorrigiertes Modell multiplen Arbeits(Aktivitäts)angebots
- Eigenarbeit, Nebenerwerb/Schwarzarbeit und Haupterwerb
  - Partizipation
  - Löhne/Einkommen
  - Zeitallokation

#### Mikrosimulation der Steuerreform 1990 für die Jahre 1990 und 2000

##### Dynamische Mikrosimulation demografischer Entwicklungen

##### Hochrechnung der Mikrodaten nach dem Prinzip des minimalen Informationsverlustes (MIL)

- Simultane Hochrechnung mit der demografischen Situation 1990
- Simultane Hochrechnung mit der demografischen Situation 2000

##### Mikrosimulation der Steuerreform 1990 für 1990 und 2000

- Mikrosimulation mit dem Statischen Sfb3-Mikrosimulationsmodell MICSIM
- Zeitallokationseffekte auf individuelle multiple markt- und nichtmarktmäßige Aktivitäten

Abb. I.1: Mikrosimulation der Steuerreform 1990: Struktur des Analysesystems

Quelle: Merz, J. (1991a)

## 1.2 Statistik und EDV: Deskription und Inferenz am Beispiel des Programmpakets SPSS (Statistical Package for the Social Sciences)

### SPSS – Funktionen:

- Datenhandling, Datenbearbeitung
- Analysemodule
- Grafiken
- Utilities

[www.spss.com](http://www.spss.com)

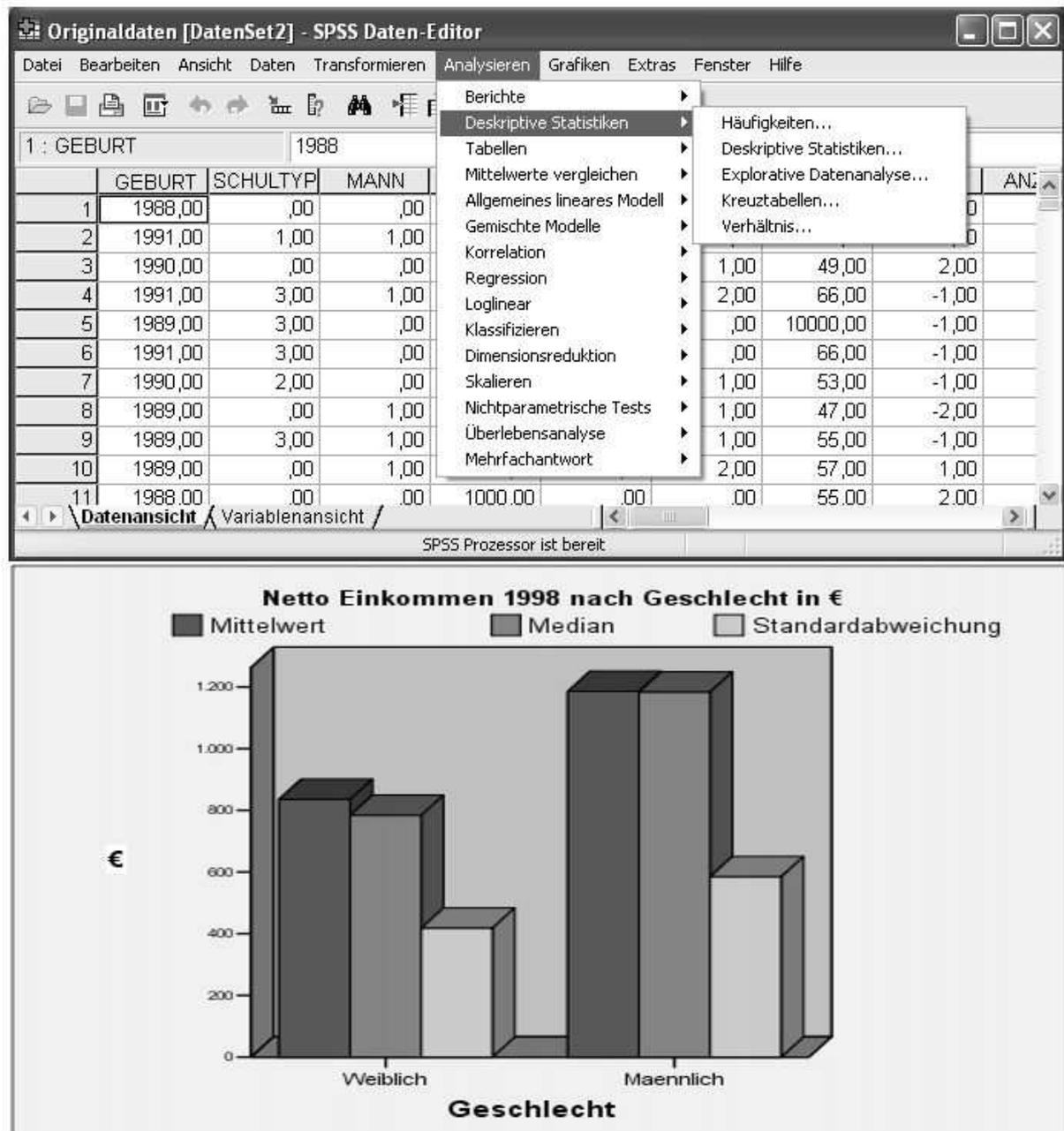


Abb. I.2 SPSS, Screenshot Datenansicht und Beispiel Grafikoutput

### 1.3 Problemorientierte Statistik: Studien zur aktuellen und zukünftigen Situation der Erde – Club of Rome und Intergovernmental Panel on Climate Change (IPCC)

Club of Rome: Grenzen des Wachstums mit ersten Studien aus den 70er Jahren

<http://www.clubofrome.de/>

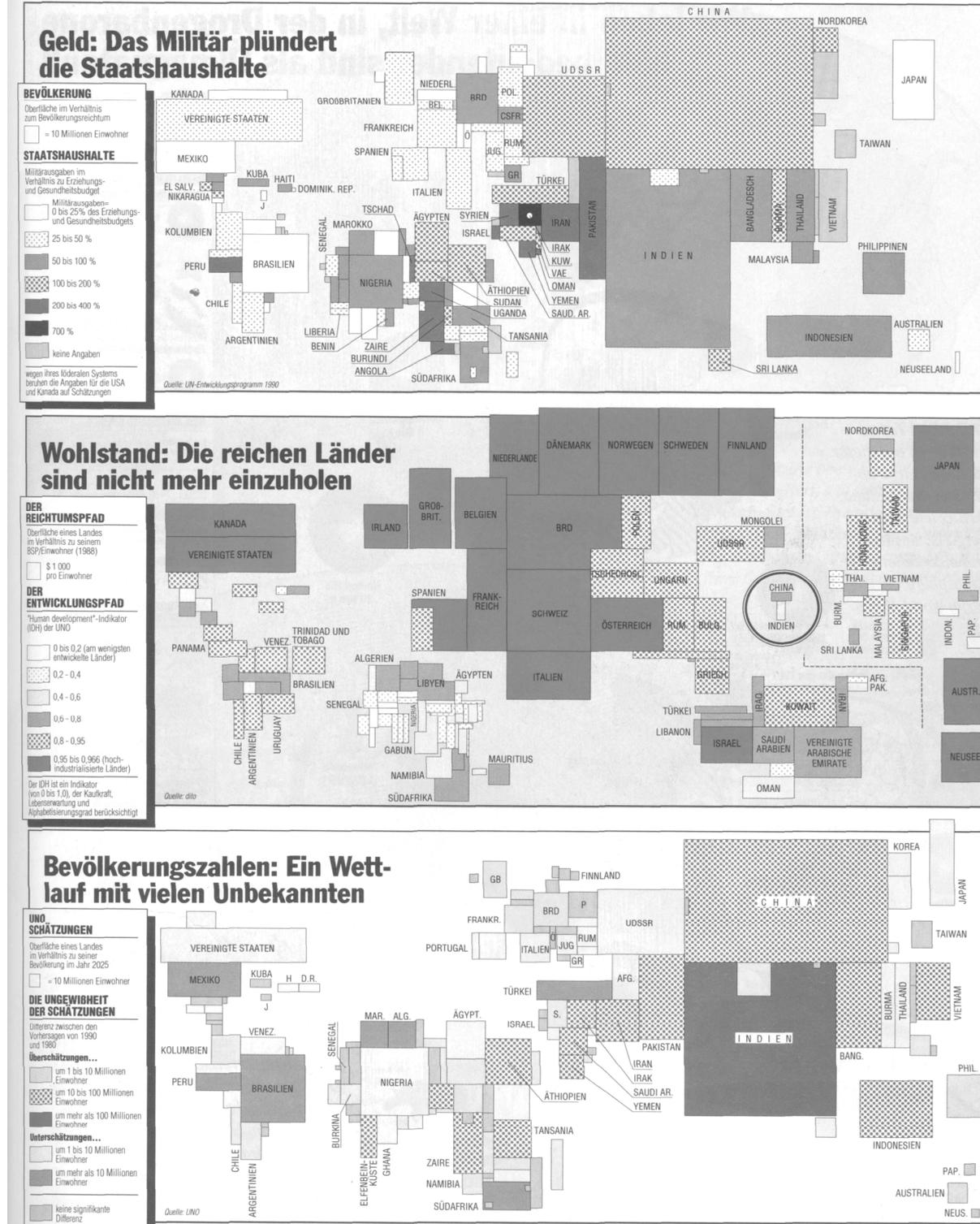


Abb. I.3: Blick in die Zukunft: Militär, Wohlstand, Bevölkerung

Quelle: Club of Rome (1991), S. 31

# „Wir leben in einer Welt, in der Drogenbarone bald bedeutender sind als Ölmagnaten“

Alexander King, Ehrenpräsident

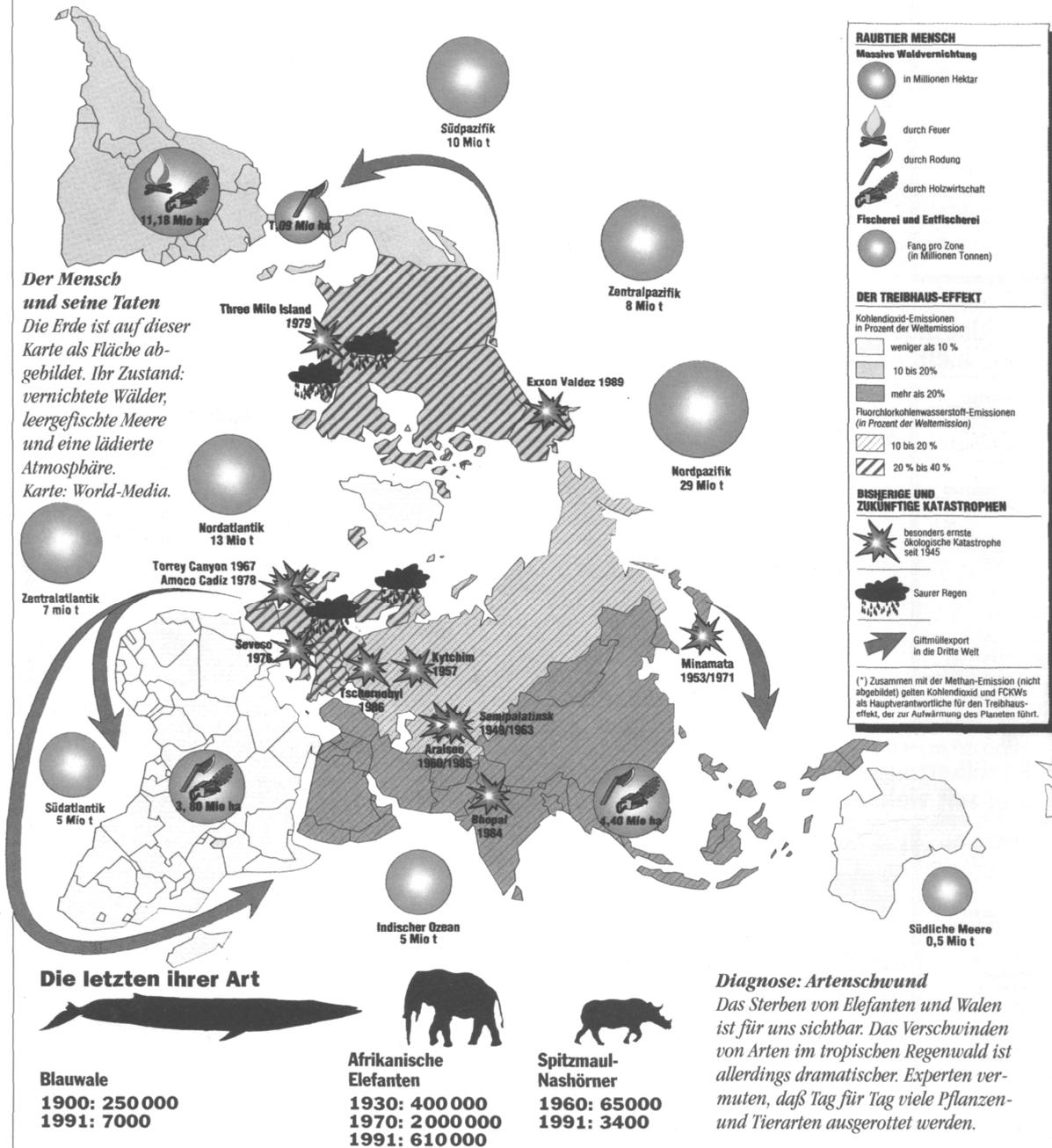


Abb. I.4: Blick in die Zukunft: Der Mensch und seine Taten  
Quelle: Club of Rome (1991), S. 32

Tab. I.1: Weltweites Wachstum in ausgewählten Sektoren

	1970		1990	
Weltbevölkerung	3,6	Mrd	5,3	Mrd
Kraftfahrzeuge gefahrte Kilometer/Jahr (nur OECD-Länder)	250,0	Mio	560,0	Mio
PKW	2.584,0	Mrd	4.489,0	Mrd
LKW	666,0	Mrd	1.536,0	Mrd
Ölverbrauch/Jahr	17,0	Mrd Barrel	24,0	Mrd
Kohleverbrauch/Jahr	2,3	Mrd Tonnen	5,2	Mrd
Kapazität E-Werke	1,1	Mrd Kilowatt	2,6	Mrd
Strom aus Kernkraft/Jahr	79,0	Mrd Terawatt-Std.	1.884,0	Mrd
Getränkeverbrauch/Jahr nicht alkoholisch/Jahr	23,0	Mrd Liter	58,0	Mrd
Bierverbrauch/Jahr	19,0	Mrd Liter	29,0	Mrd
Aluminium für Getränkebehälter	72.700,0	Mrd Tonnen	1.251.900,0	Mrd
Müll aus Gemeinden/Jahr (nur OECD-Länder)	302,0	Mio Tonnen	420,0	Mio

Quelle: Meadows et al. (1992), S. 27

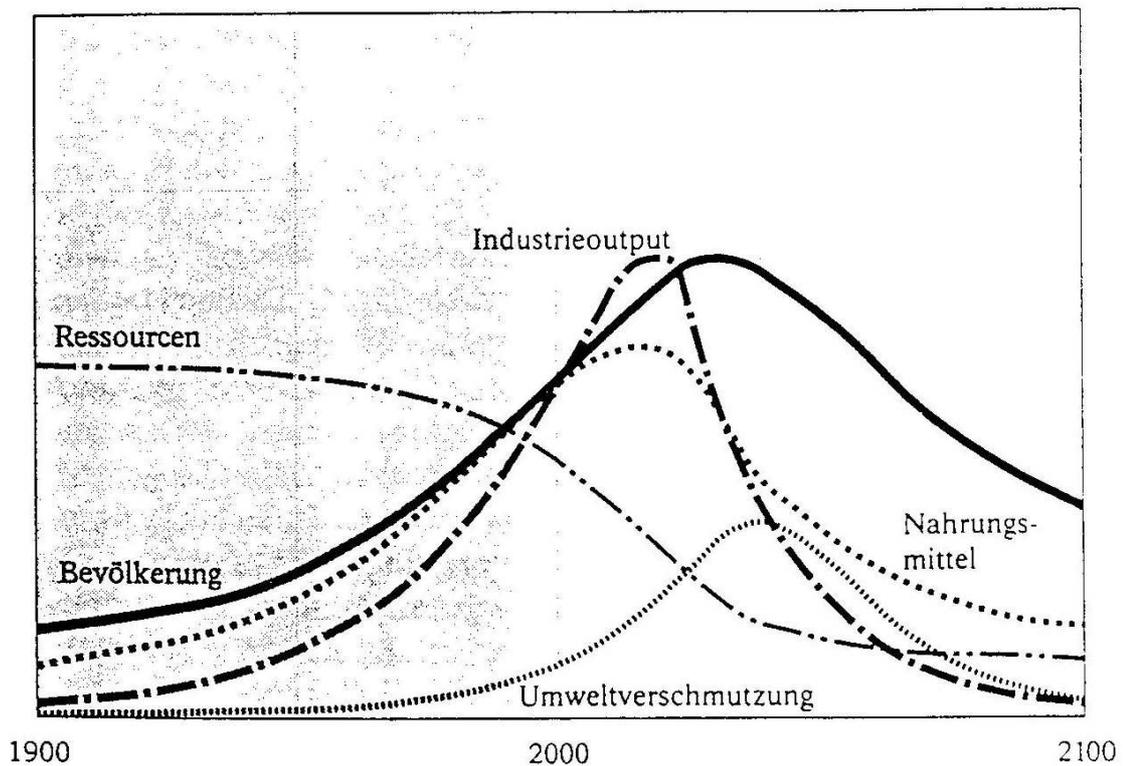


Abb. I.5: Szenario 1: 'Standardlauf' von Grenzen des Wachstums

Quelle: Meadows et al. (1992), S. 166

**Intergovernmental Panel on Climate Change (IPCC) <http://www.ipcc.ch/>**  
Weltklimabericht der UNEP und WMO

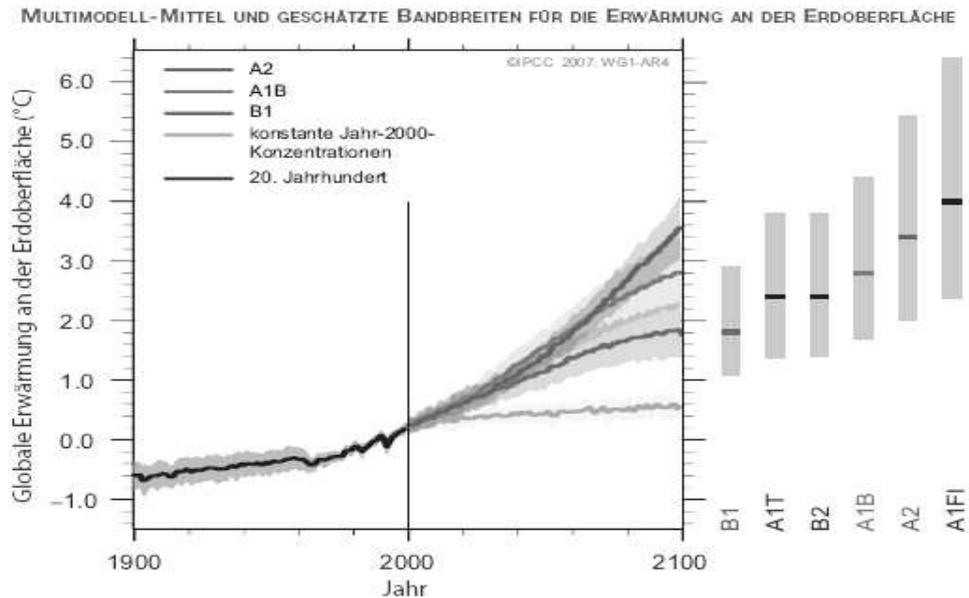


Abb. I.6: Erwärmung der Erdoberfläche

Quelle: <http://www.bmu.de/klimaschutz/downloads/doc/39255.php> Stand August 2007

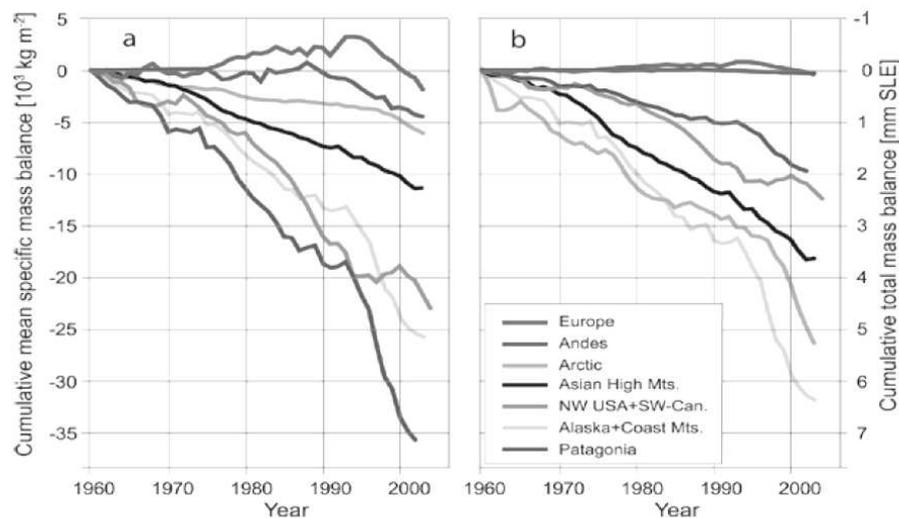


Abb. I.7: Veränderung der Gletschermassen

Quelle: Climate Change and Water, Technical Paper of the Intergovernmental Panel on Climate Change, S. 20

# 1.4 Anwendungsorientierte Statistik: Fragen zur Wohnsituation aus dem Sozio-Ökonomischen-Panel (SOEP), Leben in Deutschland, Befragung 2007 zur sozialen Lage der Haushalte

## Auszug zu Wohnungsfragen aus dem Haushaltsfragebogen

5. Wie würden Sie die Wohngegend hier beschreiben?
- Ein reines Wohngebiet mit überwiegend Altbauten .....
  - Ein reines Wohngebiet mit überwiegend Neubauten .....
  - Ein Mischgebiet mit Wohnungen und Geschäften bzw. Gewerbebetrieben .....
  - Ein Geschäftszentrum (Läden, Banken, Verwaltungen) mit wenigen Wohnungen .....
  - Ein Gewerbe- bzw. Industriegebiet mit wenigen Wohnungen .....
10. Wie groß ist die Wohnfläche dieser Wohnung insgesamt? .....  qm
11. Und wie viele Räume hat Ihre Wohnung?  
☞ Gemeint sind Räume ab 6 qm, ohne Küche und ohne Bad. .....  Räume
12. Wie beurteilen Sie insgesamt die Größe Ihrer Wohnung?  
 Ist sie für Ihren Haushalt ...
- viel zu klein .....
  - etwas zu klein .....
  - gerade richtig .....
  - etwas zu groß .....
  - viel zu groß? .....
14. Wie ist Ihre Wohnung ausgestattet?  
 Gehört zu Ihrer Wohnung ...
- |   | Ja                       | Nein                     |
|---|--------------------------|--------------------------|
| - Küche .....                               | <input type="checkbox"/> | <input type="checkbox"/> |
| - Bad / Dusche innerhalb der Wohnung .....  | <input type="checkbox"/> | <input type="checkbox"/> |
| - Fließend Warmwasser / Boiler .....        | <input type="checkbox"/> | <input type="checkbox"/> |
| - WC innerhalb der Wohnung .....            | <input type="checkbox"/> | <input type="checkbox"/> |
| - Zentralheizung oder Etagenheizung .....   | <input type="checkbox"/> | <input type="checkbox"/> |
| - Balkon / Terrasse .....                   | <input type="checkbox"/> | <input type="checkbox"/> |
| - Keller / Abstellräume .....               | <input type="checkbox"/> | <input type="checkbox"/> |
| - Eigener Garten / Gartenbenutzung .....    | <input type="checkbox"/> | <input type="checkbox"/> |
| - Alarmanlage .....                         | <input type="checkbox"/> | <input type="checkbox"/> |
| - Klimaanlage .....                         | <input type="checkbox"/> | <input type="checkbox"/> |
| - Sonnenkollektor, Solarenergieanlage ..... | <input type="checkbox"/> | <input type="checkbox"/> |
15. Haben Sie oder Ihr Vermieter seit Anfang 2006 an dieser Wohnung eine oder mehrere der folgenden Modernisierungen vorgenommen?
- Eine Küche eingebaut .....
  - Bad, Dusche oder WC innerhalb der Wohnung eingebaut .....
  - Zentralheizung oder Etagenheizung eingebaut .....
  - Neue Fenster eingebaut .....
  - Sonstige größere Maßnahmen.....
- Nein, nichts davon ....
- Sie springen auf Frage 18!
16. Erfolgte diese Modernisierung auf Kosten des Vermieters oder auf Ihre eigenen Kosten?
- Auf Kosten des Vermieters .....  ➔ Sie springen auf Frage 18!
  - Auf eigene Kosten .....
  - Teils / teils.....
22. Wie hoch ist derzeit die monatliche Miete?
- EURO      Zahle keine Miete .....  ➔ Sie springen auf Frage 37!



**Beispiele:**

---

*Problem:* Erfassung der Arbeitslosensituation in Niedersachsen (Arbeitsamtsbezirke) und in den fünf neuen Bundesländern

*Fragen:* Durchschnittliche Dauer der Arbeitslosigkeit?  
Anteil der Altersgruppen?  
Regionale Differenzierung?  
Einfluß der Berufsqualifikation?  
Vergleichbarkeit der Informationen?

*Fazit:* Beschreibend → Deskriptive Statistik

*Problem:* Qualitätskontrolle im Produktionsbereich eines Unternehmens

*Fragen:* Annahme oder Ablehnung des 'Loses'?  
Eingriff in den Produktionsprozeß?

*Fazit:* Operationale Funktion (Entscheidungshilfe) → Schließende, Induktive Statistik

*Problem:* Wirtschaftspolitische Behauptung: Transfer des Staates (z.B. Arbeitslosengeld) verlängert signifikant die Arbeitslosigkeitsdauer

*Fragen:* Wie kann dies operationalisiert werden?  
Gibt es zwei 'identische Gruppen' mit und ohne Transfers (Soziale Experimente: USA z.B. New Jersey Income Maintenance Experiment)?  
Wie ist das Verhalten zu quantifizieren?  
Daten: Querschnitt (Umfrage), Panel?

*Fazit:* Deskription und Inferenz im Rahmen einer empirischen Wirtschaftsforschung;  
Raum-sachliche, Raum-zeitliche Begrenzung der Aussage

---

## 2.2 Zur geschichtlichen Entwicklung

Geschichtlich: **Erhebungsstatistiken**

Erhebung über staatskundliche Phänomene:

- Bevölkerung, Ackerfläche, Goldbestand (Ägypten 2500 v. Chr. etc.)

Neuere Zeit: **Wahrscheinlichkeitsrechnung**

Blaise PASCAL (1623-1662), Pierre de FERMAT (1601-1665)

de Moivre, Laplace: Frankreich

Bernoulli, Euler: Schweiz

Gauss: Deutschland

Kolmogorov, Tschebyscheff, Markoff: Rußland

19.-20. Jahrhundert Induktive Statistik

### Statistik ist für viele Bereiche von Bedeutung:

Wirtschaft:	Unternehmen (Marketing), Private Haushalte (Einkommen und Konsum), Staat (Wirtschafts- und Sozialpolitik) etc.
Soziologie:	Gruppenverhalten (Sozioökonomie)
Medizin:	Rauchgewohnheiten → Lungenkrebs
Psychologie:	Lernerfolg
Physik/Mathematik:	Atome, Unschärfbereiche, Zufall
Biologie:	Mendelsche Gesetze
Umwelt:	Verschmutzungsgrade
etc.	

Benutzergruppen statistischer Informationen im Überblick finden sich in Abb. I.6.

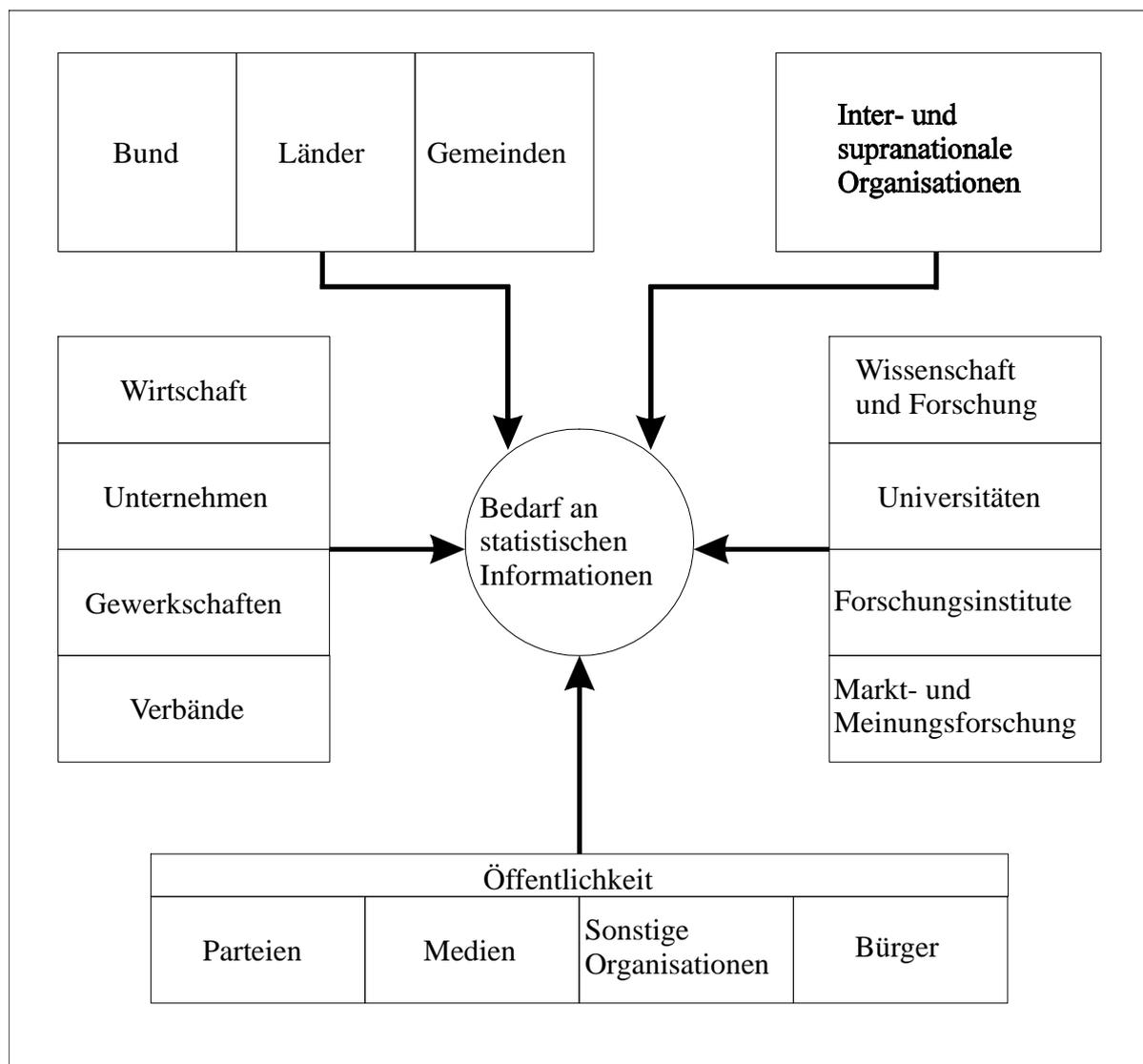


Abb. I.6: Benutzergruppen statistischer Informationen  
Quelle: Statistisches Bundesamt (1989)

## 3 Träger der Wirtschaftsstatistiken und statistische Quellen

### 3.1 Amtliche Statistik

Legale Basis: Gesetze, Rechtsverordnungen (Bundesstatistikgesetz: BStatG 1987)

#### Statistische Ämter

Statistisches Bundesamt, Statistische Landesämter, Statistische Ämter der Städte, Gemeinden und Kommunen, Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

Das Statistische Bundesamt gibt Informationsbroschüren über seine Aufgabe, Aufbau und Arbeitsweise heraus, so z.B.:

- Statistisch gesehen

Diese Broschüre und weiteres Material sind kostenlos erhältlich bei:

Statistisches Bundesamt  
Gustav-Stresemann-Ring 11  
65189 Wiesbaden  
Tel.: 0611/75-2405  
Fax: 0611/75-3330  
[www.destatis.de](http://www.destatis.de)  
info@destatis.de

Allgemeine Aufgabenbeschreibung: Statistisches Bundesamt (Hrsg.), Das Arbeitsgebiet der Bundesstatistik, Kohlhammer Verlag, Mainz 1988

Zum Ablauf von Bundesstatistiken vgl. Abb. I.7.

#### Veröffentlichungen des Statistischen Bundesamtes

- Statistisches Jahrbuch für die Bundesrepublik Deutschland
- Wirtschaft und Statistik (monatlich)
- Fachserien 1-19

Zur Übersicht des Veröffentlichungssystems des Statistischen Bundesamtes vgl. Abb. I.8.

**GENESIS-Online:** Datenbankzugriff auf das statistische Informationssystem des Bundes

#### Beispiele:

---

Statistiken aus Befragungen:

Volkszählung (1970/71, 1987), Mikrozensus (jährlich), Einkommens- und Verbrauchsstichprobe (EVS), Zeitbudgeterhebung (1991/92 und 2001/02)...

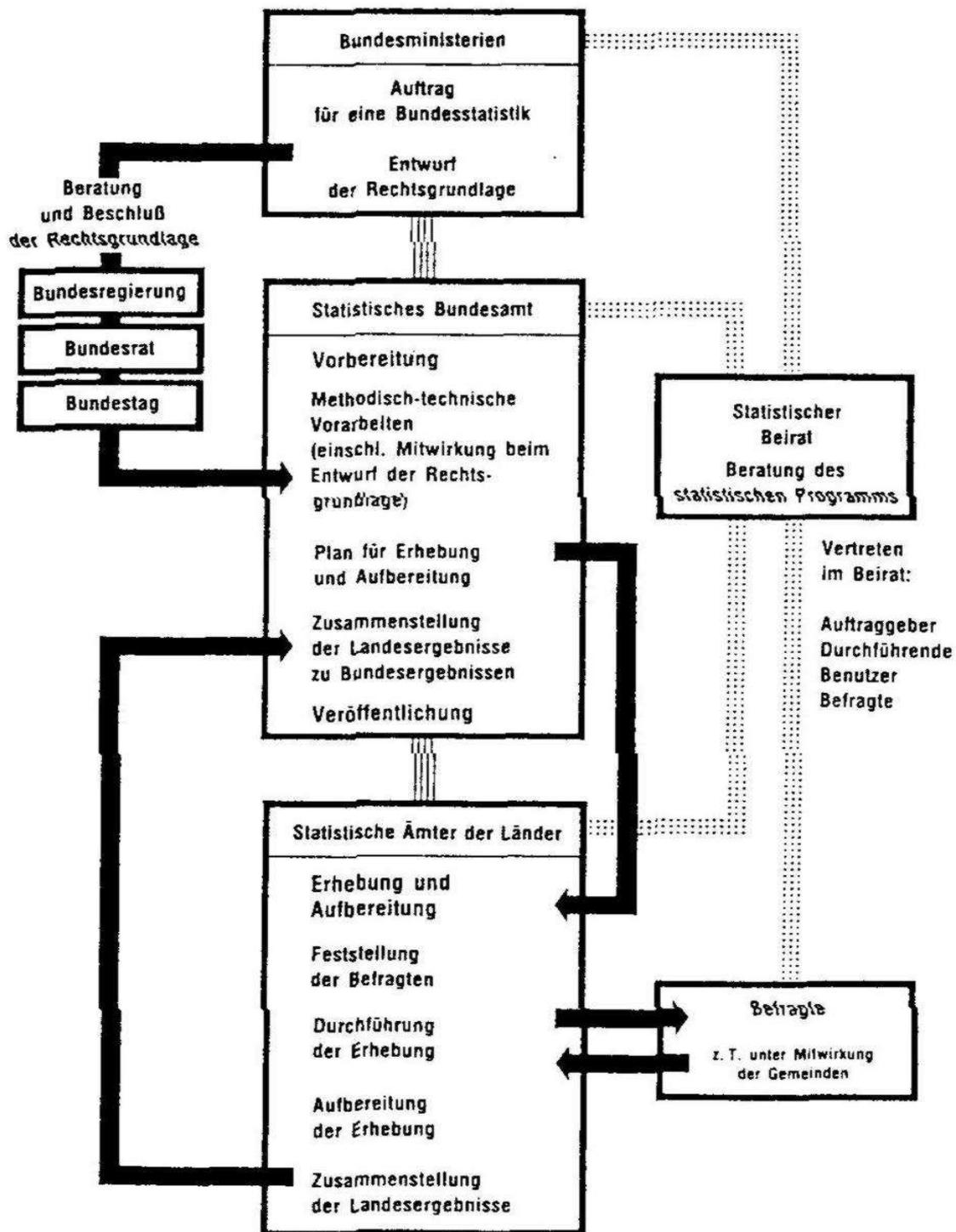
---

**Ressortstatistik**

- Deutsche Bundesbank [www.bundesbank.de](http://www.bundesbank.de)  
Veröffentlichungen: - Monatsberichte
  - Statistische Beihefte zu den Monatsberichten
    - Reihe Bankenstatistik
    - Reihe Kapitalmarktstatistik
    - Reihe Zahlungsbilanzstatistik
    - Reihe Saisonbereinigte Wirtschaftszahlen
    - Reihe Devisenkursstatistik
  
- Bundesagentur für Arbeit [www.arbeitsagentur.de](http://www.arbeitsagentur.de)  
Veröffentlichungen: - Amtliche Nachrichten der BA (ANBA)
  - Monatlicher Arbeitsmarktbericht
  - Jahresbericht
  
- IAB: Institut für Arbeitsmarkt- und Berufsforschung [www.iab.de](http://www.iab.de)  
Veröffentlichungen: - Mitteilungen aus der Arbeitsmarkt- und Berufsforschung

Weitere Hinweise: Rinne (1994), Kunz (1987), Grohmann (1986a), v.d. Lippe (1996)

## Ablauf von Bundesstatistiken



Vereinfachte Darstellung. Bei zentral durchgeführten Statistiken übernimmt das Statistische Bundesamt auch die Erhebung und Aufbereitung.

Abb. I.7: Ablauf von Bundesstatistiken

Quelle: Statistisches Bundesamt (1988), S. 47

<b>Zusammenfassende Veröffentlichungen</b>			
Allgemeine Querschnittsveröffentlichungen	Thematische Querschnittsveröffentlichungen	Veröffentlichungen zu Organisations- und Methodenfragen	Kurzbroschüren

<b>Fachserien</b>
1 Bevölkerung und Erwerbstätigkeit
2 Unternehmen und Arbeitsstätten
3 Land- und Forstwirtschaft, Fischerei
4 Produzierendes Gewerbe
5 Bautätigkeit und Wohnungen
6 Handel, Gastgewerbe, Reiseverkehr
7 Außenhandel
8 Verkehr
9 Geld und Kredit
10 Rechtspflege
11 Bildung und Kultur
12 Gesundheitswesen
13 Sozialleistungen
14 Finanzen und Steuern
15 Wirtschaftsrechnungen
16 Löhne und Gehälter
17 Preise
18 Volkswirtschaftliche Gesamtrechnungen
19 Umweltschutz

<b>Systematische Verzeichnisse</b>				
Unternehmens- und Betriebs-systematiken	Güter-systematiken	Personen-systematiken	Regional-systematiken	Sonstige Systematiken

**Karten**

**Statistik des Auslandes**

**Fremdsprachige Veröffentlichungen**

Abb. I.8: Veröffentlichungssystem des Statistischen Bundesamtes

Quelle: Statistisches Bundesamt (1989)

## 3.2 Nichtamtliche Statistik

- Wirtschaftsverbände, Berufsorganisationen
- Industrie- und Handelskammer (IHK), Kammern
- Markt- und Meinungsforschungsinstitute (Infratest, Marplan, Emnid, Allensbach...)
- Arbeitnehmer- und Arbeitgeberorganisationen
- Wirtschaftsforschungsinstitute

von Interessenverbänden:

- IW - Institut der Deutschen Wirtschaft (Köln), [www.iwkoeln.de](http://www.iwkoeln.de)
- WSI - Wirtschafts- und Sozialwissenschaftliches Institut (Düsseldorf), [www.wsi.de](http://www.wsi.de)

gemeinnützig und unabhängig:

- DIW - Deutsches Institut für Wirtschaftsforschung (Berlin), [www.diw.de](http://www.diw.de), z.B. Vierteljährliche Volkswirtschaftliche Gesamtrechnung
- ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften (Kiel), [www.zbw.eu](http://www.zbw.eu), Leibniz-Informationszentrum Wirtschaft
- Ifo - 'Information und Forschung' - Ifo Institut für Wirtschaftsforschung (München), [www.ifo.de](http://www.ifo.de), CESifo Economic Studies (vierteljährlich), Ifo-Geschäftsklimaindizes
- IfW - Institut für Weltwirtschaft an der Universität Kiel (Kiel), [www.uni-kiel.de/ifw](http://www.uni-kiel.de/ifw)
- RWI - Rheinisch-Westfälisches Institut für Wirtschaftsforschung (Essen), [www.rwi-essen.de](http://www.rwi-essen.de)
- IWH - Institut für Wirtschaftsforschung Halle (Halle a.d. Saale), [www.iwh-halle.de](http://www.iwh-halle.de)

Jahresgutachten von DIW, Ifo, IfW, RWI und IWH

- GESIS, Leibniz-Institut für Sozialwissenschaften, Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS), [www.gesis.org/das-institut/](http://www.gesis.org/das-institut/)

Wissenschaftliche Abteilungen:

- Survey Design and Methodology (SDM)
- Dauerbeobachtung der Gesellschaft (DBG): German Microdata Lab – GML, Zentrum für Sozialindikatorenforschung – Zsi, Survey Programme: Allgemeine Bevölkerungsumfrage ALLBUS, International Social Science Programme ISSP, Comparative Study of Electoral Systems CSES, German Longitudinal Election Study GLES
- Datenarchiv für die Sozialwissenschaften: Datenservice zu nationalen und international-vergleichenden Umfragen zu soziologischen und politikwissenschaftlichen Fragestellungen, ALLBUS

### 3.3 Internationale Organisationen

- UN (Statistical Yearbook, Demographic Yearbook, Yearbook of National Accounts Statistics, New York), <http://unstats.un.org>
- OECD (Paris), <http://www.oecd.org>
- EUROSTAT, <http://epp.eurostat.ec.europa.eu>
- ILO (International Labour Organization, Genf), <http://www.ilo.org>
- WHO (World Health Organisation, Genf), <http://www.who.int/GHO/>
- IMF (International Monetary Fund, Washington), <http://www.imf.org>
- Multinationale Konzerne (Fachabteilungen)

### 3.4 Aufgaben und Quellen der Wirtschaftsstatistik im vereinten Deutschland

#### Ehemalige DDR

- Ministerrat der DDR, Staatliche Zentralverwaltung für Statistik
- oft Vollerhebung, z.B. Berufstätigenerhebung (BTE) ohne den X-Bereich (Stasi, Armee)

Gravierende Unterschiede in den Statistiksystemen (vor allem: Wirtschaftsstatistik)

Im Rahmen der Wiedervereinigung wurde eine Umstrukturierung bzw. ein Neuaufbau der Einrichtungen vorgenommen (z.B. Arbeitsämter, BA)

Zur Vereinheitlichungsdiskussion vgl. Allgemeines Statistisches Archiv, Bd. 76, 1992

## **Statistiken für die fünf neuen Bundesländer**

- Sozialreport '90
- Statistisches Bundesamt: Neue Publikationen (monatlich)
- Übersicht zum Stand der Einführung wichtiger ausgewählter Bundesstatistiken
- IAB-Werkstattbericht 'Neue Bundesländer'
- Presseinformationen der Bundesagentur für Arbeit
- Sozio-ökonomisches Panel (SOEP-Ost)
  - 1. Welle 1984 (West), 1. Welle 1990 (Ost)
- Infratest- 'Befragung Ost'

Bei einem Ereignis wie der deutschen Wiedervereinigung ist der sich hieraus ergebende strukturelle Bruch in einer ökonomisch/statistischen Bewertung besonders zu beachten. Konkret entsteht dieser hier durch die Ausweitung der Grundgesamtheit um rund 16 Millionen Menschen mit grundlegend anderen demografischen Voraussetzungen z.B. im Bereich Lebensumstände, Einkommen, Lebensstandard, Gesundheit usw. Die sich durch die Einbeziehung dieser Faktoren ergebenden Veränderungen müssen vor allem bei der Vergleichbarkeit der Daten berücksichtigt werden (z.B. wird in Gutachten meist gesondert die Situation vor (1989) und nach der Wiedervereinigung (1990) ausgewiesen → siehe z.B. tabellarischen Anhang im Gutachten der Sachverständigenrates Wirtschaft).

## **4 Das Adäquationsproblem und einige wissenschaftstheoretische Bemerkungen**

### **4.1 Wissenschaftstheoretische Grundlagen: Zur Struktur und Anwendung wissenschaftlicher Theorien**

Erklärung der Welt, 'Kritischer Rationalismus'

Der Kritische Rationalismus versteht unter einer Theorie allgemein ein System wissenschaftlicher Sätze über die Realität.

#### **Albert (1964):**

"Die zentralen Bestandteile realwissenschaftlicher Theorien haben den Charakter von nomologischen Hypothesen (Gesetzen), also empirisch gehaltvollen Aussagen über die Struktur der Realität, die infolgedessen anhand der Tatsachen nachgeprüft werden können."

#### **Logische Struktur einer Theorie:**

Aus Axiomen (Grundsätzen) werden Theoreme (abgeleitete Sätze) deduziert.

#### **Inhaltliche Struktur einer Theorie:**

1. Geltungsmodus: Nur Aussagen mit empirischem Geltungsanspruch (nicht nur denkbare Situationen)
2. Widerspruchsfreiheit
3. Operationalität (eindeutig definierte Begriffe, überprüfbar)
4. Empirischer Gehalt
5. Prüfbarkeit (Falsifizierbarkeit) und Bewährung
6. Allgemeinheit

Für ein **Explanandum**, das den zu erklärenden Tatbestand beschreibt, ist ein **Explanans** zu finden, das ein **allgemeines Gesetz** und **Anwendungsbedingungen** enthält (Albert 1964).

**Beispiel:**

Explanandum: 'Die Ausgaben für den privaten Konsum sind gestiegen'

Konsumhypothese: 'Wenn sich das verfügbare Einkommen der privaten Haushalte um einen bestimmten Betrag erhöht, dann steigen die Konsumausgaben im Mittel um einen bestimmten (anderen) Betrag'

→ Allgemeines Gesetz: 'Das verfügbare Einkommen ist um einen bestimmten Betrag gestiegen'

Explanans = Allgemeines Gesetz und Anwendungsbedingung

Aus Explanans kann dann das Explanandum logisch abgeleitet werden.: Wenn das verfügbare Einkommen um x steigt, erhöhen sich die Konsumausgaben um f(x).

**Karl Popper (1964):**

$$\left. \begin{array}{l} G \text{ (allgemeines Gesetz: Wenn A, dann C)} \\ A \text{ (singuläre Aussage: Nun A)} \end{array} \right\} \text{Explanans (Prämissen)}$$

Also C (Conclusion)

Explanandum

Prüfung: Bewährungsgrad einer Theorie durch permanente Falsifikationsversuche feststellen

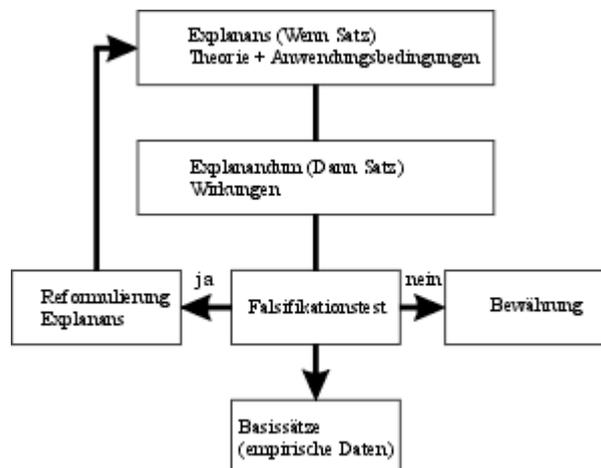


Abb. I.9: Der Falsifikationsprozess

Quelle: Hujer und Cremer (1978), S. 12

**Thomas Kuhn** betrachtet in seinem Buch 'Structure of Scientific Revolutions', 2nd ed., Chicago 1970, ausführlich den **Paradigmenwechsel**. Kuhn interpretiert ein Paradigma als eine Menge von Wechselbeziehungen, die aber noch unterentwickelt sind. Ein Paradigma ist dann ein Gedankengebäude, das Antworten zu bestimmten Fragen liefern kann, auf denen dann Ausweitungen der Theorie vorgenommen werden können. Diese 'normal science' beschäftigt sich also mit dem 'puzzle solving': Offen gebliebene Fragen einer revolutionären Theorie werden gelöst.

Neue Paradigmen tauchen dann auf, wenn Widersprüche in den bestehenden Paradigmen entdeckt werden, d.h. alte Paradigmen werden dann fallengelassen, wenn sie immer mehr Fragen nicht beantworten können. Der Zeitablauf ist also durch wechselseitige Phasen von normaler und revolutionärer Wissenschaft gekennzeichnet.

## 4.2 Das Adäquationsproblem: Allgemeine Problemstellung und statistische Operationalisierung

Der theoretische (idealtypische) Begriff ist mit einem empirisch feststellbaren Begriff zu verbinden.

### Adäquationsproblem:

Die Diskrepanz zwischen theoretischem und statistischem Begriff sollte so klein wie möglich werden (Grohmann (1986a), S. 18, Blind, **Frankfurter Schule**).

### Beispiel:

---

Frage: Es ist zu klären, ob und wie die Käufe eines Verbrauchsgutes von der Anzahl, Größe und dem Einkommen der Haushalte abhängen.

Theorie: Wirtschaftswissenschaften, Welche Haushaltsdefinition soll verwendet werden? (möglich: Haushalt = Wirtschaftseinheit, d.h. Gruppe von Personen, die einen gemeinsamen Verbrauchsplan aufstellen, die gemeinsam wirtschaftlich handeln) (Mikroökonomie).

Realität: Eine gemeinsame Kaufentscheidung wird eher selten getroffen.

Praxis: Laut Volkszählung umfaßt ein Haushalt alle diejenigen Personen, die in der gleichen Wohnung leben und den Lebensunterhalt überwiegend gemeinsam betreiben.

Welche Einkommensdefinition soll verwendet werden? Das Geldeinkommen (Lohn und Gehalt) gehört grundsätzlich zum Einkommen. Was ist aber mit dem 13. Monatsgehalt, einmaligen Zahlungen etc.? Wie verhält es sich mit einer mietfreien Werkswohnung, einem Firmenwagen oder Natureinkommen, 'fringe benefits' oder laufendem Brutto-/Nettoeinkommen?

---

## 5 Sachgerechte Interpretation: 'How (not) to lie with statistics'

Illustratives zu 'How to lie with statistics': Huff (1978), Krämer (1991), Schwarze (1990), S. 17-19.

### 5.1 Some pitfalls

- willkürlicher Bezug

#### Beispiel:

Von 2500 Studenten nehmen 50 an der Statistik-Klausur teil. Keiner von ihnen besteht die Klausur. Der Dozent behauptet, die Durchfallquote betrage zwei Prozent.

- fehlende Sachlogik

#### Beispiel:

Es wird beobachtet, dass der Anstieg von Storchbrütungen in einer Region mit einem Anstieg der Geburten einher geht. Basierend auf dieser Beobachtung wird ein statistisch 'bewiesener' Zusammenhang vermutet.

- Herausgreifen bestimmter 'passender' Werte

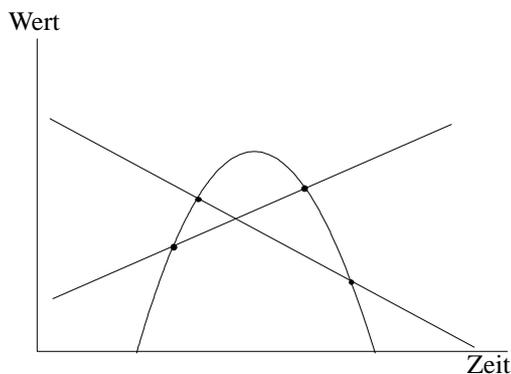


Abb. I.10: Herausgreifen bestimmter 'passender' Werte

- Maßstabsmanipulation:

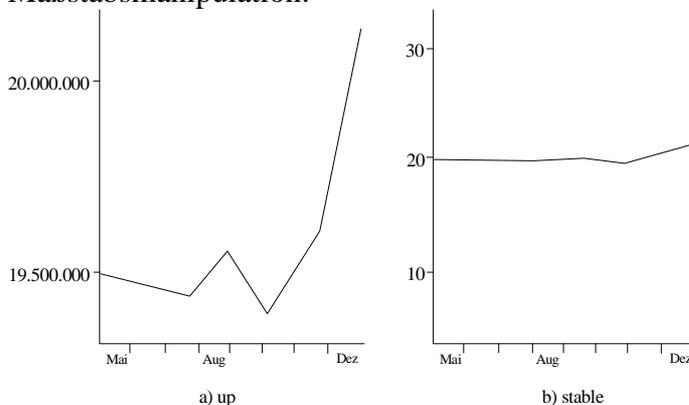


Abb. I.11: Maßstabsmanipulation

## 5.2 How not to lie with statistics

- Wahl eines problemgerechten operablen statistischen Begriffs (Adäquationsproblem)
- eindeutige Bezugs- und Berechnungsangaben
- sachgerechte, ehrliche Präsentation
- adäquater Vergleich
- Vorsicht mit Extrapolationen, Vorhersagen weit außerhalb des Datenstützbereichs unzulässig
- ... 'be honest'

## 6 Statistische Einheiten und statistische Massen

### 6.1 Statistische Einheiten

**Statistische Einheiten** = Merkmalsträger der Untersuchung, Einzelobjekt, Proband

Statistische Einheiten sind real, klar voneinander abgrenzbar, zählbar:

z.B. Haushalt, Person, Betrieb, Gemeinde, Fläche, Region, PKW, Beobachtungspunkt in einer Stadt (bei Verkehrszählung) etc.

Identifikations- oder Abgrenzungskriterien:

- sachlich
- räumlich
- zeitlich

**Beispiel:** \_\_\_\_\_

Sfb 3 Nebenerwerbstätigkeitsumfrage 1984 (Merz, Helberger und Schneider (1985)):

Statistische Einheit: Personen in Privathaushalten (nicht in Anstalten)

Identifikation:

- sachlich: Person über 14 Jahre mit definierter Nebenerwerbstätigkeit (Mehrfacherwerbstätigkeit oder Mehrfachstätigkeit?)
  - räumlich: Gebiet der Bundesrepublik Deutschland (einschließlich West-Berlin)
  - zeitlich: vergangene drei Monate aus einem 'Quartal' 1984
-

## 6.2 Statistische Massen

**Statistische Masse** = Gesamtheit aller statistischen Einheiten, die vom Untersuchungsziel her gleichartig sind (übereinstimmende Identifikationskriterien, sachliche, räumliche und zeitliche Abgrenzung)

### Beispiele:

---

- Zahl der Arbeitslosen in der Bundesrepublik im Monat Februar 2010
  - Rechnungen des Unternehmens McAlles im Monat Oktober 2010
- 

### Arten statistischer Massen:

#### - Bestandsmassen

Für einen **Zeitpunkt** definiert, z.B. Kassenbestand eines Warenhauses am 31.12.2010, Wohnbevölkerung in der Bundesrepublik am Stichtag (z.B. 25.5.1987, Stichtag der letzten Volkszählung)

#### - Ereignis- (Bewegungs-) Massen

Für einen **Zeitraum** definiert, z.B. Eheschließungen in der Bundesrepublik im Jahre 2010, Scheckeingänge der Bank X im Monat März 2010

Die Verknüpfung von Bestands- und Bewegungsmassen erfolgt durch **Fortschreibung**:

$$\begin{array}{ccccc} \text{Anfangsbestand} & + & \text{Zugang} & \text{./. Abgang} & = & \text{Endbestand} \\ \text{(Bestandsmasse)} & & \text{(Bewegungsmasse)} & & & \text{(Bestandsmasse)} \end{array}$$

### Beispiel:

---

Zugelassene Kraftfahrzeuge in Lüneburg am	1.1.2010
+ Neuzulassungen	1.1. - 31.12.2010
./. Abmeldungen	1.1. - 31.12.2010
= zugelassene Kfz in Lüneburg am	31.12.2010

---

## 7 Merkmale, Merkmalsausprägungen und Meßskalen

### 7.1 Merkmale und Merkmalsausprägungen

**Merkmal** = **Eigenschaft** einer statistischen Einheit (Merkmalsträger)

### Beispiele:

---

- Merkmale einer Person (statistische Einheit):  
Alter, Geschlecht, Einkommen, Berufsausbildung... (sozioökonomische Merkmale)
- Merkmale eines Haushalts (statistische Einheit):

Haushaltsgröße, Alter des 'Haushaltvorstandes', Anzahl der Kinder, Anzahl der Erwerbstätigen...

- Merkmal X: ALTER, SEX, AGE OF HEAD...

**Merkmalsausprägung = Mögliche Werte** (Kategorien) eines Merkmals

**Beispiele:**

Statistische Einheit = Studentin

Merkmal:	ALTER	STUDIENFACH
Merkmalsausprägung: (Merkmalswert, Beobachtungswert)	21 Jahre	BWL

## 7.2 Meßskalen und ihre Eigenschaften

Meßskalen der Merkmalsausprägungen haben unterschiedliches Meßniveau:

### Nominalskala

- keine natürliche Reihenfolge, Merkmalsausprägungen sind gleichberechtigt nebeneinander:  
z.B. Geschlecht, Hautfarbe, Religion, Staatsangehörigkeit (Codes)

### Ordinalskala

- Rangskala, natürliche Rangordnung, Abstände nicht quantifizierbar:  
z.B. Examensnoten, Güteklassen, Bundesligatabelle, \* oder \*\*\* [Sterne] Hotel

### Metrische Skala

- Kardinalskala, Abstände sind angebbbar (Maßsystem):
  - *Intervallskala:*
    - mit Abständen, aber ohne Bezugspunkt:  
z.B. Abstand zwischen Gefrier- und Siedepunkt des Wassers in 100 Teilen, Kalenderzeit
  - *Verhältnisskala:*
    - mit Abständen und mit natürlichem Bezugspunkt:  
z.B. Körpergröße (cm), Alter (Jahre), Einkommen
  - (- *Absolutskala:*
    - metrische Skala mit natürlichem Nullpunkt und natürlicher Einheit:  
z.B. Stückzahlen)

Zur schematischen Abgrenzung von Meßskalen siehe Tab. I.2.

Tab. I.2: Schematische Abgrenzung von Meßskalen

Merkmale	Skala	gleich oder verschieden	natürliche Reihenfolge	konstanter Wertabstand	natürlicher Nullpunkt	natürliche Einheit	Rechenoperationen
qualitative	Nominalskala	X					Häufigkeiten
intensitätsmäßige	Ordinalskala	X	X				Median
quantitative	Intervallskala	X	X	X			Addition und Subtraktion
quantitative	Verhältnisskala	X	X	X	X		Division und Multiplikation
quantitative	Absolutskala	X	X	X	X	X	Division und Multiplikation

### 7.3 Diskrete und stetige Merkmale

#### Diskretes Merkmal

abzählbare Ausprägungen:

z.B. Erwerbstätigkeit (0/1), Anzahl der Studentinnen und Studenten im Hörsaal

#### Stetiges Merkmal

überabzählbare Ausprägungen (kontinuierlich):

z.B. Länge, Gewicht

(Approximativ stetig: z.B. Geld)

### 7.4 Quantitative und qualitative Merkmale

#### Quantitative Merkmale

Abstände zwischen Merkmalsausprägungen sind durch reelle Zahlen meßbar:

z.B. Länge, Alter

#### Qualitative Merkmale

Kategoriale Abstufung:

z.B. Farbe, Noten

Mit dieser Unterscheidung ist es schwierig, ordinalskalierte Daten einzuordnen.

## 8 Statistische Untersuchungen: Erhebung, Aufbereitung und Analyse

### 8.1 Vorgehensweise bei statistischen Untersuchungen

#### 1. Schritt: Abbildung materieller Fragestellung in statistisches Konstrukt

Ausgangspunkt: Problemstellung aus Theorie (Empirische Überprüfung von Hypothesen) oder Praxis (Wert- oder Zielvorstellungen): z.B. Überprüfung der Theorie Dualer Arbeitsmärkte, Erreichung des Ziels 'Hoher Beschäftigungsstand'

Theorie, Praxis → materielle Fragestellung

- Entwicklung eines Begriffssystems aus der Fachwissenschaft (idealtypischer Begriff):
- z.B. Entwicklung eines Modells über den Arbeitsmarkt, Definition von Arbeitsmarktindikatoren für den Zielkomplex 'Hoher Beschäftigungsstand'
- Übersetzung von materieller Fragestellung in statistische Konstrukte, Messung (Adäquationsproblem)
- z.B. Messung der Arbeitslosenquote, Ermittlung der Zahl der offenen Stellen

#### 2. Schritt: Erhebung

Nach Festlegung der zu untersuchenden Objekte und Merkmale erfolgt die Beobachtung oder Befragung (Erhebung) (siehe VI.2).

#### 3. Schritt: Aufbereitung und Darstellung des Beobachtungsmaterials

Ziel: Verdichtung, Straffung und Strukturierung des Urmaterials (bei Befragungen, Umfragen: 'Editing' der Daten)

Gruppierung der Daten nach Merkmalsklassen, grafische/tabellarische Darstellung der klassifizierten Daten

Berechnung von beschreibenden Maßzahlen wie Mittelwerte, Streuung, Zusammenhangsmaße

#### 4. Schritt: Analyse durch Schluß von der Stichprobe auf die Grundgesamtheit

Ziel: Aussagen über unbekannte Gesamtmasse (Grundgesamtheit), z.B. Studenten der Universität Lüneburg, Bevölkerung Niedersachsens

Die Ergebnisse der bekannten Teilmasse (Stichprobe) werden auf die Grundgesamtheit mit einer bestimmten Wahrscheinlichkeit übertragen, z.B. durchschnittliche Körpergröße einer Stichprobe der Studenten der Statistik I Vorlesung → alle Studenten (Lüneburg)

Methoden:

- Schätzen der unbekanntenen Größen (Parameter)
- Testen von Hypothesen über diese Größen

**Beispiele:**

---

- a) Anteil der Kommilitoninnen und Kommilitonen an den Studenten
- b) Individuelle Wirkungen der Steuerreform '90: Ist der geschätzte Koeffizient  $b$  zu der Anzahl der Kinder signifikant von Null verschieden, um das Arbeitsangebot im Nebenerwerb zu erklären?

$$= f\left(\dots, \underset{*}{b} \cdot \text{Anzahl der Kinder}, \dots\right)$$

\*=signifikant von Null verschieden?

---

## 5. Schritt: Sachgerechte Interpretation der Ergebnisse

Interpretation der Ergebnisse im Sinne der untersuchten materiellen Frage

Beachtung der Einschränkungen aus:

- Definition
- verwendeten Methoden
- zeitliche, räumliche und sachliche Begrenzung
- Güte der Daten

Falsifikation: wissenschaftliche Theorie oder politische Zielvorstellung widerlegt oder nicht?

Aus den fünf Schritten seien zwei vertieft: Erhebung sowie Aufbereitung und Analyse.

## 8.2 Erhebung: Erhebungsarten und Erhebungstechnik

### Erhebung

Bei einer vorbestimmten Menge von **Merkmalsträgern** (Untersuchungseinheiten, Objekte, Probanden) werden eine Anzahl von **Merkmalen** erhoben und deren **Ausprägungen** erfaßt.

### Erhebungsarten

- Primärstatistische Erhebungen
  - ausschließlich zu statistischen Zwecken
- Sekundärstatistische Erhebungen
  - bereits vorhandene, zunächst für andere Zwecke gesammelte Daten (z.B. Lohnsteuerstatistik)
- Vollerhebung
  - alle Einheiten werden erfaßt (z.B. Volkszählung)
- Teilerhebung
  - ausgewählte Einheiten (z.B. Mikrozensus)

## Erhebungstechnik

- Schriftliche Befragung, Fragebogen (Questionnaire)
  - offene Fragen (ohne Antwortvorgabe), geschlossene Fragen (mit vollständigen Antwortvorgaben)
- Mündliche Befragung, Interviewer, CATI (Computer Aided Telephone Interview), CAPI
  - z.B. ISR, Ann Arbor Michigan, Panel Study of Income Dynamic (PSID), Infratest Sozialforschung, München
- Online-Erhebung bzw. -Umfrage
  - z.B. FFB-Online Erhebung zu Freien Berufen 2005/2006
- Beobachtung

## 8.3 Aufbereitung und Analyse

Aufbereitung und Analyse umfaßt alles vom Urmaterial (individueller Fragebogen, Zählblätter etc.) bis zum Ergebnis (Grafik, Tabelle)

### 1. Prüfen des Urmaterials auf Vollständigkeit, Widerspruchsfreiheit und Glaubwürdigkeit (Editing)

Widersprüche:

sozial:	80jähriger Schüler?
ökonomisch:	Großbetrieb mit Jahresumsatz von EUR 100,-
gesetzlich:	dreijährige Witwe
institutionell:	katholischer Pfarrer, verheiratet

### 2. Verschlüsseln (Kodierung) von qualitativen Merkmalsausprägungen:

Zuordnung von Variablenwerten wie z.B.:

SEX =	1: männlich, 2: weiblich (dummy variables)
FAMSTD =	1: ledig, 2: verheiratet, 3: geschieden, 4: verwitwet

Mehrstellige Codes: Arbeitsstätte, Beruf, Waren, Krankheit etc.  
z.B. Beruf: ISCO-Code (Systematik der Wirtschaftszweige)

### 3. Übertragen von Informationen auf Datenträger (z.B. CD-ROM) und maschinell unterstütztes Editing

### 4. Auswerten des Urmaterials nach Gruppen, Ausprägungen

- Deskription: Grafiken, Tabellen
- Inferenz: Regressionsanalyse, Multivariate Analyse
- Erwerb von fertigen Ergebnissen (Tabellen aus dem Statistischen Bundesamt)
- Eigene Software, Anwendungssoftware zur Analyse der Individualinformationen
- Datenbanken (dbase, ORACLE mit SQL,...)

## 9 Tabellarische und grafische Darstellung

### 9.1 Zur Präsentation von Informationen

# KLAR ... **GROß** ... *motivierend!*

'Weniger ist **MEHR!!!**', vollständige Beschriftung!

#### Beispiele:

---

- VENN-Diagramm aus einführendem Beispiel zur Mikroanalyse der Steuerreform
- Abbildungen mit Bildinformation (vgl. Abb. I.12a,b)
- charts (Software)
- Wirkung auch durch Text alleine:

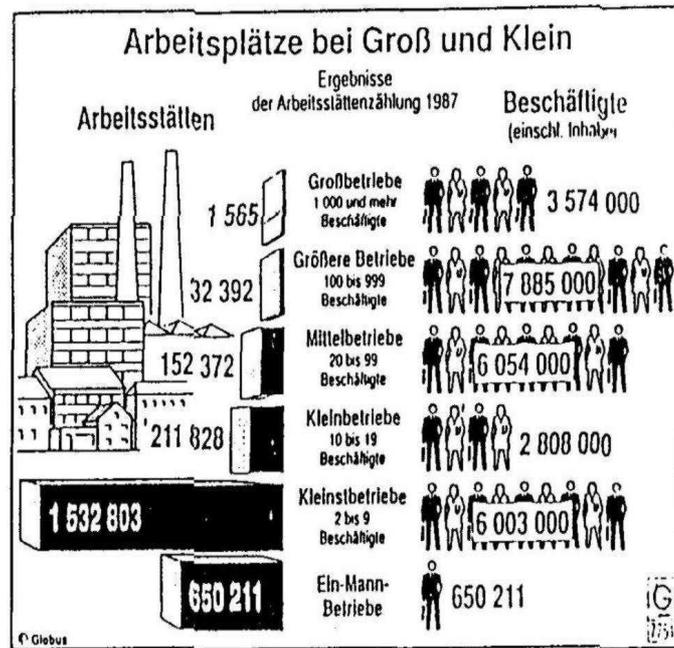
*Die Erde ist etwa 4.500.000.000 Jahre alt.*

*Vergleicht man diese Zeitspanne mit dem Leben eines 45jährigen Menschen, so traten die ersten Säugetiere vor acht Monaten in Erscheinung und Menschen gibt es erst seit wenigen Tagen.*

*Vor etwa einer Stunde erlernte der Mensch den Ackerbau und vor einer Minute begann die industrielle Revolution.*

*In diesen 60 Sekunden hat der Mensch die Rohstoffreserven unseres Planeten geplündert. Boden, Wasser und Luft verseucht und unzählige Pflanzen und Tiere ausgerottet.*

---



## Wirtschaftsdaten der EU-Länder

EU der 27 - Bevölkerung und Wirtschaftskraft

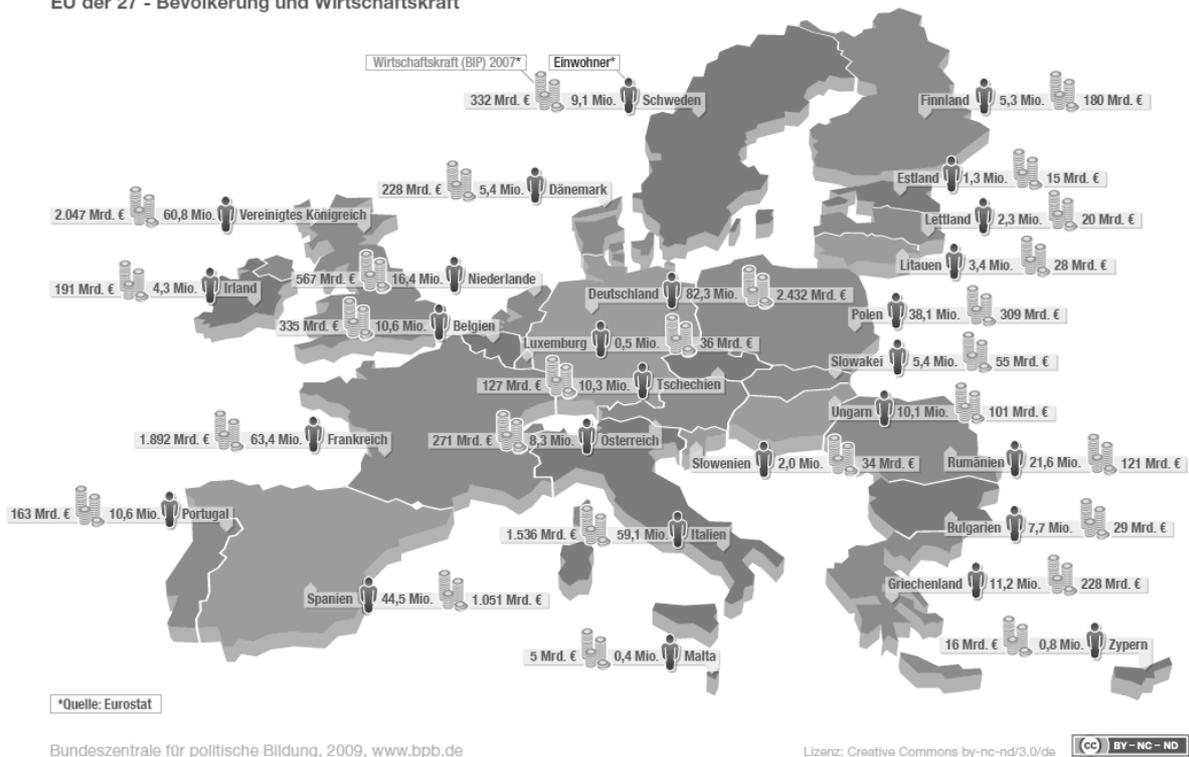
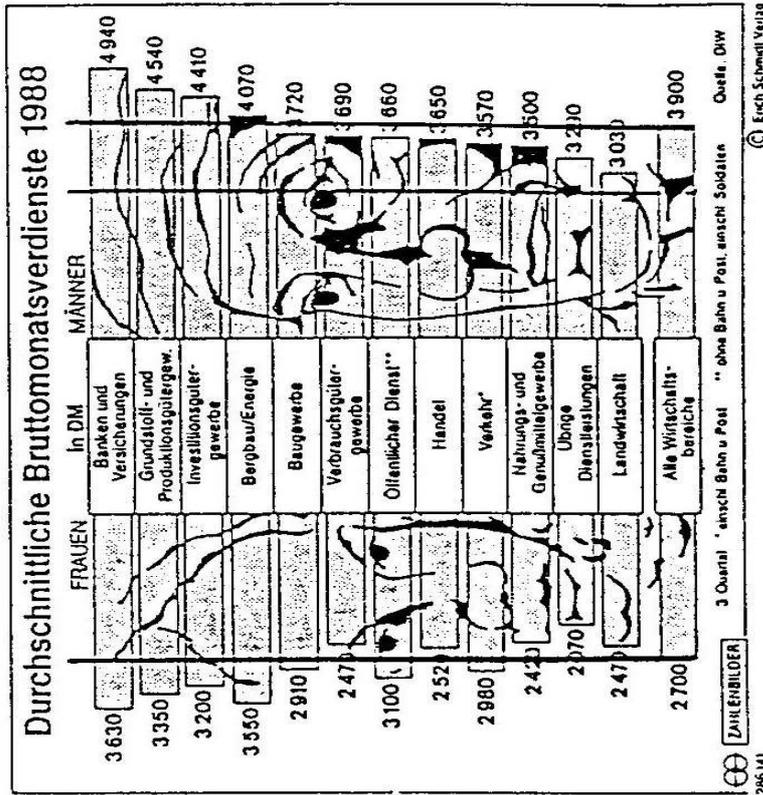
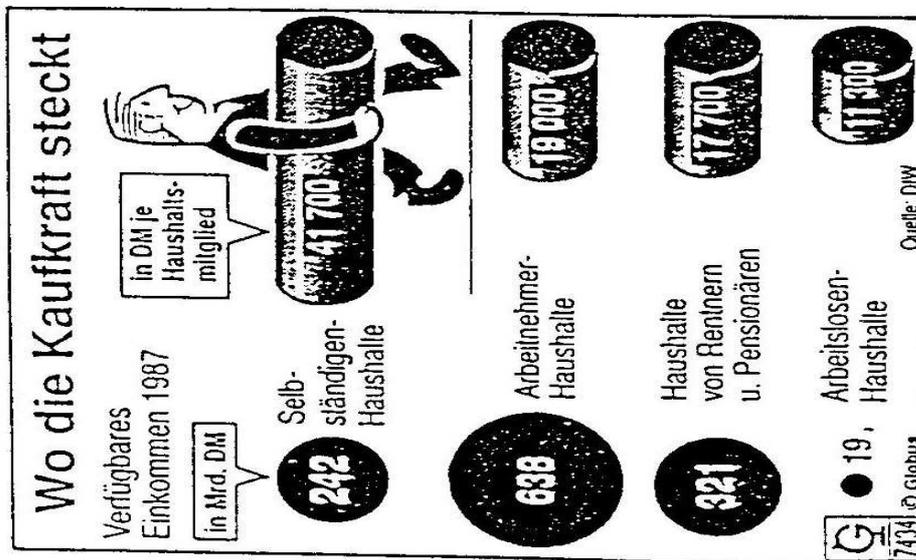


Abb. I.12a: Grafische Darstellung mit Bildsymbolen

Quelle: transcontact Verlagsgesellschaft, Bonn, für die KKB Bank AG, 1989, S. 31, Bundeszentrale für politische Bildung (2009)



Wo mit Geld gehandelt wird, da verdient man auch am meisten: Bei Banken und Versicherungen. Der öffentliche Dienst hält schönes Mittelmaß.



Auch wenn es in den Selbständigen-Familien ein Stück wohlhabender zugeht: Die viel zahlreicheren Arbeitnehmer-Haushalte geben als einzelne weniger, insgesamt aber viel mehr aus. Und die vielen Alten unter uns sind insgesamt weitaus kaufkräftiger als die Gruppe der „reichen“ Selbständigen.

Abb. I.12b: Grafische Darstellung mit Bildsymbolen

Quelle: transcontact Verlagsgesellschaft, Bonn, für die KKB Bank AG, 1989, S. 42-43

## 9.2 Tabellenaufbau und grafische Darstellung

Systematische und übersichtliche Zusammenstellung von Daten mit

- ausreichend informierender Überschrift
- möglichst Zwischensummen
- kein leeres Tabellenfeld (Information einfügen, wie z.B. nicht vorhanden, not available oder keine ausreichende Besetzungszahl)

**Normblatt DIN 55301:**

Überschrift (Titel und wichtige Angaben)

Vorspalte Kopf    zum zur    Kopf Vorspalte	<b>Tabellenkopf</b>				
<b>Vorspalte</b>					
		Fach			

(Erläuterungen)

**Grafische Darstellung:**

- Stab- oder Säulendiagramm, Balkendiagramm (für nominal und ordinal skalierte Daten)
- Flächendiagramm (flächenproportionale Darstellung)
- Kreisdiagramm ('Pie charts', Tortendiagramm)
- Piktogramm (mit Bildsymbolen)
- Kartogramm (innerhalb einer Landkarte)
- Kurvendiagramm
- Histogramm (Häufigkeiten eines klassifizierten Merkmals)
- Polygonzug (Verbindungsline der Mittelpunkte der Oberkanten des Histogramms)

**Beispiele:**

- Schwarze (1990), S. 50ff.
- Grafiken aus der Vorlesung
- 2 D- und 3 D-Grafiken aus Computerprogrammen  
(Chart, Boeing-Graph, Harvard Graphics, EXCEL etc., vgl. Abb. I.13)

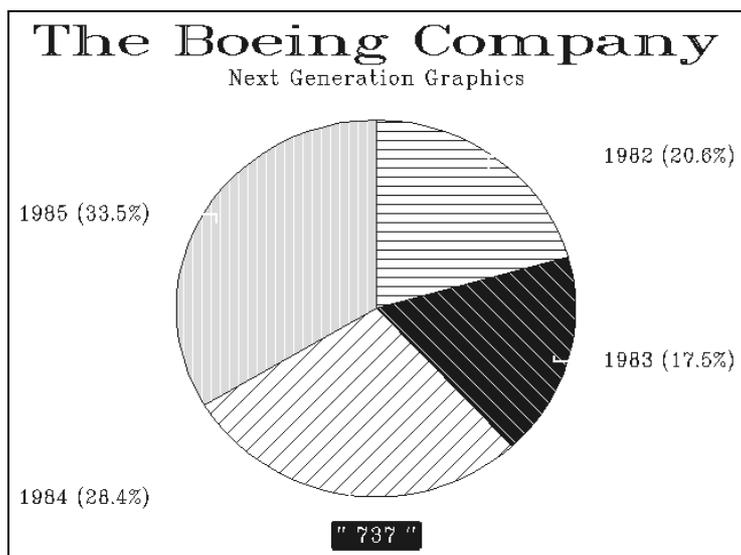
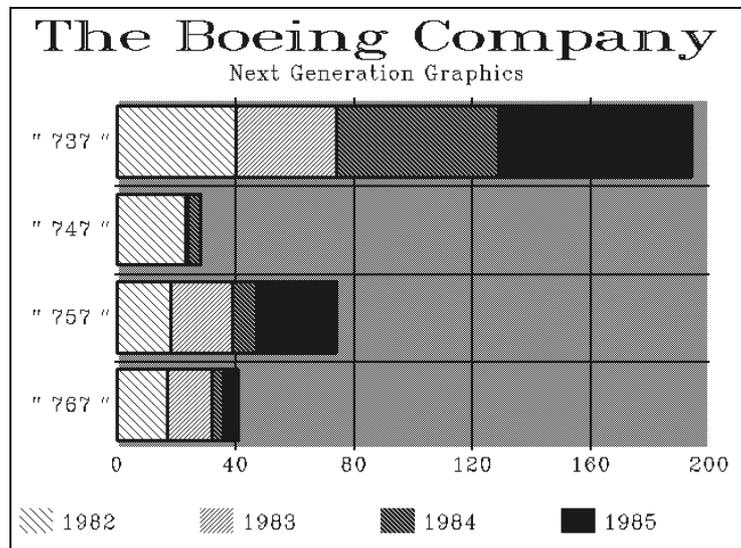
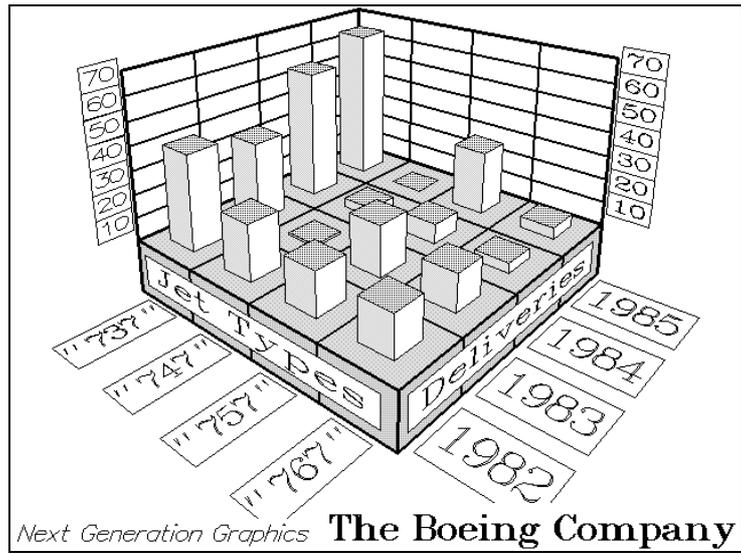


Abb. I.13: Boeing-Graph: 3 D/2 D-Darstellung

## 10 Datenschutz und Datensicherheit

**Datenschutz** ist die Aufgabe, Daten vor Mißbrauch u.a. bei der Datenverarbeitung zu bewahren und damit der Beeinträchtigung schutzwürdiger Belange der Betroffenen entgegenzuwirken.

Diskussion besonders im Umfeld des Volkszählungsgesetzes ('gläserner' Mensch, Orwell's 1984, 'big brother')

Berufskodex für Statistiker (Internationales Statistisches Institut ISI, 1986, S.238, zitiert nach Rinne (1994), S. 32):

*Statistical data are unconcerned with individual identities. They are collected to answer questions such as "how many?" or "what proportional?", not "who?". The identities and records of co-operating (or non-cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged.*

*Statisticians should take appropriate measures to prevent their data from being published or otherwise released in a form that would allow any subject's identity to be disclosed or inferred.*

Grundgesetz:

- Persönlichkeitsrecht nach Grundgesetz (GG) Art. 2, Abs. 1  
(Individuelles Persönlichkeitsrecht auf Achtung seiner Würde und Eigenwertes abgeleitet aus GG Art. 1, Abs. 2)
- Recht auf Informationsfreiheit nach GG Art. 5, Abs. 1, S. 1 ('jeder hat das Recht, ...sich aus allgemein zugänglichen Quellen ungehindert zu unterrichten.')

Grundlegende gesetzliche Regelungen:

- Bundesstatistikgesetz (BStatG) 1987 (Schutz von Daten aus statistischen Erhebungen)
- Bundesdatenschutzgesetz (BDSG) vom 20.12.1990
- Landesdatenschutzgesetze

Datenschutzbeauftragte des Bundes, der Länder, von Institutionen

### Datensicherheit

Organisatorische und technische Aufgabe zur Gewährleistung der Sicherheit von Datenbeständen und Datenverarbeitungsabläufen (Datenzugriff nur für Berechtigte, unverfälschte Verarbeitung der Daten etc.)

Weitere Informationen zu Datenschutz und Datensicherheit: Rinne (1994), Kap. 2.3

und im Internet unter:

[www.gesetze-im-internet.de/bdsg\\_1990/index.html](http://www.gesetze-im-internet.de/bdsg_1990/index.html)

***Keyconcepts***

*Begriff und Aufgaben der Statistik*

*Träger der Wirtschaftsstatistiken*

*Adäquationsproblem*

*Sachgerechte Interpretation*

*Statistische Massen und Einheiten*

*Merkmale, Merkmalsausprägungen*

*Meßskalen*

*Vorgehensweise bei statistischen Untersuchungen*

*Tabellarische, grafische Darstellung*

## II Wirtschafts- und Sozialstatistik



*Bedeutung und Quantifizierung wichtiger Bereiche der Wirtschafts- und Sozialstatistik*

### 1 Zur Aufgabe und Bedeutung der Wirtschafts- und Sozialstatistik

Statistiken über das wirtschaftliche und soziale Geschehen als Grundlage unternehmerischer Entscheidung, Entscheidungen im Haushalt und wirtschaftspolitischer Maßnahmen.

#### Primärstatistiken

Für Zwecke der Statistik allgemein werden die Ausprägungen der Erhebungsmerkmale direkt (mündlich/schriftlich) bei den Merkmalsträgern erfaßt.

#### Beispiele:

---

Volkszählung, Arbeitsstättenzählung, Einkommens- und Verbrauchsstichproben (EVS), Betriebserhebungen, Haushaltsbücher, Zeitbudgeterhebung

---

#### Sekundärstatistiken

Die direkte Datenerhebung bei den Merkmalsträgern entfällt; die Merkmalsausprägungen werden aus bereits vorliegenden Unterlagen gewonnen.

#### Beispiele:

---

Bevölkerungsbewegung (Geburten, Sterbefälle, Heiraten aus Unterlagen der Standesämter), Kraftfahrzeugstatistiken (Kfz-Zulassungsstellen), Volkswirtschaftliche Gesamtrechnungen (Unterlagen des Statistischen Bundesamtes), Statistiken über veröffentlichte Unternehmensbilanzen etc.

---

#### Problemorientierte, aufbereitete Statistiken

- Sozialberichterstattung: Soziale Indikatoren (vgl. Abschnitt II.8)
- Soziologischer Almanach (Ballerstedt/Glatzer)
  - SPES: Sozialpolitisches Entscheidungs- und Indikatorensystem für die Bundesrepublik Sozialpolitische Forschergruppe Frankfurt/Mannheim (1972-1979), Johann Wolfgang Goethe-Universität Frankfurt
- Datenreport
  - Der Datenreport ist ein Gemeinschaftsprojekt des Statistischen Bundesamtes (Destatis), des Wissenschaftszentrums Berlin für Sozialforschung (WZB) und der Gesellschaft sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS).

Ein globaler statistischer Überblick über die Bundesrepublik Deutschland findet sich in Tab. II.1.

Die Leistungen des Sozialbudgets im Überblick enthält Tab. II.2.

Im folgenden werden wichtige Statistiken der Wirtschafts- und Sozialstatistik behandelt:

Bevölkerung, Produktion, Dienstleistungen, Erwerbstätigkeit, Einkommen und Verbrauch, Preise, Umwelt, Volkswirtschaftliche Gesamtrechnung mit Bruttoinlandsprodukt und Ökosozialprodukt, Soziale Indikatoren

Tab. II.1: Die Bundesrepublik Deutschland in Zahlen

<b>Bundesrepublik Deutschland (in Mrd. Euro)</b>	<b>1995</b>	<b>2000</b>	<b>2009</b>
Bruttoinlandsprodukt	1848,5	2062,5	2407,2
Konsum	1429,0	1606,07	1887,8
Private Konsumausgaben	1067,2	1214,2	1414,7
Konsumausgaben des Staates	361,8	391,9	473,1
Bruttowertschöpfung			
Land- und Forstwirtschaft, Fischerei	21,3	23,5	18,1
Produzierendes Gewerbe (ohne Baugewerbe)	424,5	465,3	473,8
Baugewerbe			
Baugewerbe	112,3	96,2	98,2
Handel, Gastgewerbe und Verkehr	300,3	337,3	378,2
Finanzierung, Vermietung u. Unternehmensdienstleister	441,5	510,9	666,5
Öffentliche und private Dienstleister	371,3	423,0	516,0
Volkseinkommen	1397,2	1524,4	1815,0
Arbeitnehmerentgelt	997,0	1100,0	1224,0
Unternehmens- und Vermögenseinkommen	400,2	424,4	591,1
Bruttosozialprodukt (Bruttonationaleinkommen)	1834,8	2043,2	2450,4
Außenbeitrag (Ausfuhr minus Einfuhr)	8,7	7,3	109,9

Quelle: Statistisches Jahrbuch (2010), S. 623

Tab. II.2: Leistungen des Sozialbudgets

Institutionen	2000 Mrd. EUR	2000 %	2006 Mrd. EUR	2006 %	Leistungs- bereiche
Sozialbudget insgesamt	680,8	100,0	1306,5	100,0	
Allgemeine Systeme	445,6	65,4	856,2	65,5	
Rentenversicherung	217,4	31,9	410,9	31,5	A,G
Krankenversicherung	132,0	19,4	253,9	19,4	G
Pflegeversicherung	16,7	2,4	31,9	2,4	
Unfallversicherung	10,8	1,6	20,3	1,6	G,A
Arbeitsförderung	64,8	9,5	135,9	10,4	B
Kindergeld	0,1	0,0	0,2	0,02	F
Erziehungsgeld	3,7	0,5	3,1	0,2	F
Sondersysteme	5,2	0,8	6,8	0,5	
Altershilfe für Landwirte	3,3	0,5	3,1	0,2	A
Versorgungswerke	1,9	0,3	3,6	0,3	A
Leistungssyst. d. öff. Dienstes	49,4	7,2	49,8	3,8	
Pensionen	33,4	4,9	35,7	2,7	A
Familienzuschläge	7,0	1,0	3,0	0,2	F
Beihilfen	8,9	1,3	11,2	0,9	G
Arbeitgeberleistungen	54,2	7,9	56,5	4,3	
Entgeltfortzahlung	26,3	3,9	25,3	1,9	G
Betriebliche Altersversorgung	14,2	2,0	19,1	1,5	A
Zusatzversorgung	7,9	1,2	9,5	0,7	A,G
Sonstige	5,8	0,8	2,7	0,2	G,W,F
Arbeitgeberleistungen					
Entschädigungen	6,6	1,0	4,3	0,3	
Soziale Entschädigung	5,2	0,8	3,4	0,3	
Lastenausgleich	0,1	0,0	0,1	0,01	P,A
Wiedergutmachung	1,2	0,2	0,7	0,1	A
sonstige	0,1	0,0	0,1	0,0	A,P
Entschädigungen					
Soziale Hilfen und Dienste	48,9	7,2	45,2	3,5	P,L
Sozialhilfe	25,7	3,8	21,9	1,7	
Jugendhilfe	16,8	2,5	19,0	1,5	G,A
Ausbildungsförderung	0,9	0,1	1,8	0,1	F,L
Wohngeld	4,3	0,6	1,7	0,1	B
Förderung der Vermögensbildung	1,2	0,2	0,8	0,1	W
Indirekte Leistungen	70,9	10,4	72,4	5,5	G
Steuerliche Maßnahmen (ohne FLA)	39,3	5,8	35,7	2,7	S
Familienleistungsausgleich (FLA)	31,6	4,6	36,7	2,8	F,S, W

A: Alter und Hinterbliebene, B: Beschäftigung, F: Ehe und Familie, G: Gesundheit, L: Allgemeine Lebenshilfen,  
P: Folgen politischer Ereignisse, S: Sparen, W: Wohnen

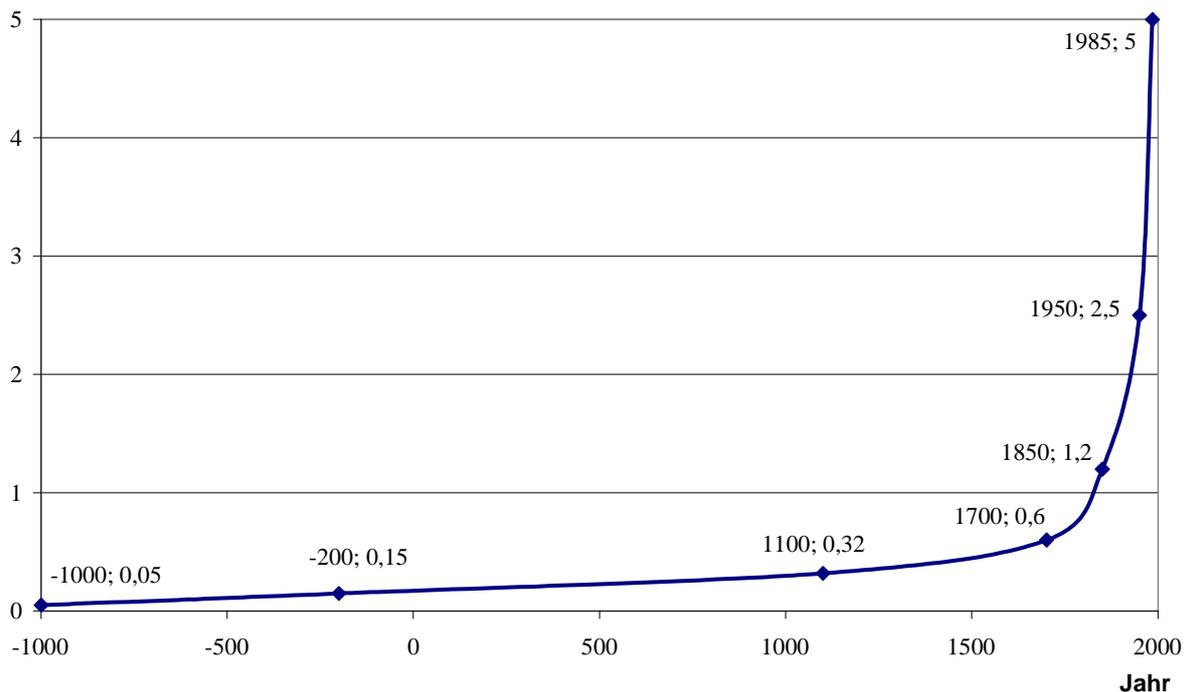
Quellen: Datenreport (2002), S. 194; Datenreport (2008), S. 269.

## 2 Bevölkerung

Die Bevölkerung steht im Zentrum der Gesellschaftswissenschaften und ist die Basis für viele Bereiche der Politik (Arbeitsmarkt-, Familien-, Wohnungsbau-, Regional- und Wirtschaftspolitik und Sozialpolitik generell). Die Bevölkerungsentwicklung hat nachhaltige Konsequenzen für praktisch alle Lebensbereiche (Umwelt: Stadt, Land, Wald, Ernährung, Ressourcenverbrauch, Gefährdung der Tier- und Pflanzenwelt, Umweltverschmutzung etc.).

Zum Wachstum der Weltbevölkerung vgl. die Abb II.1-3.

### Bevölkerung in Mrd.



### Verdopplungszeit:



Abb. II.1: Entwicklung der Weltbevölkerung in Mrd. Menschen

Quelle: Myers, N. (1985) und eigene Berechnungen#

Städte	2005 (Millions)	2015 (projected Millions)	Population Change 2005 - 2015 (projected %)
Tokyo	35,3	36,2	2,55
Mexico City	19,2	20,6	7,29
New York - Newark	18,5	19,7	6,49
Mumbai (Bombay)	18,3	22,6	23,50
Sao Paulo	18,3	20,0	9,29
Delhi	15,3	20,9	36,60
Calcutta	14,3	16,8	17,48
Buenos Aires	13,3	14,6	9,77
Jakarta	13,2	17,5	32,58
Shanghai	12,7	12,7	0,00
Dhaka	12,6	17,9	42,06
Los Angeles	12,1	12,9	6,61
Karachi	11,8	16,2	37,29
Rio de Janeiro	11,5	12,4	7,83
Osaka-Kobe	11,3	11,4	0,88
Cairo	11,1	13,1	18,02
Lagos	11,1	17,0	53,15
Beijing	10,8	11,1	2,78
Metro Manila	10,7	12,6	17,76
Moscow	10,7	10,9	1,87

Abb. II.2: Bevölkerung von Städten mit 10 Millionen und mehr Einwohnern für 2005 und prognostiziert bis 2015

Quelle: United Nations (2005)

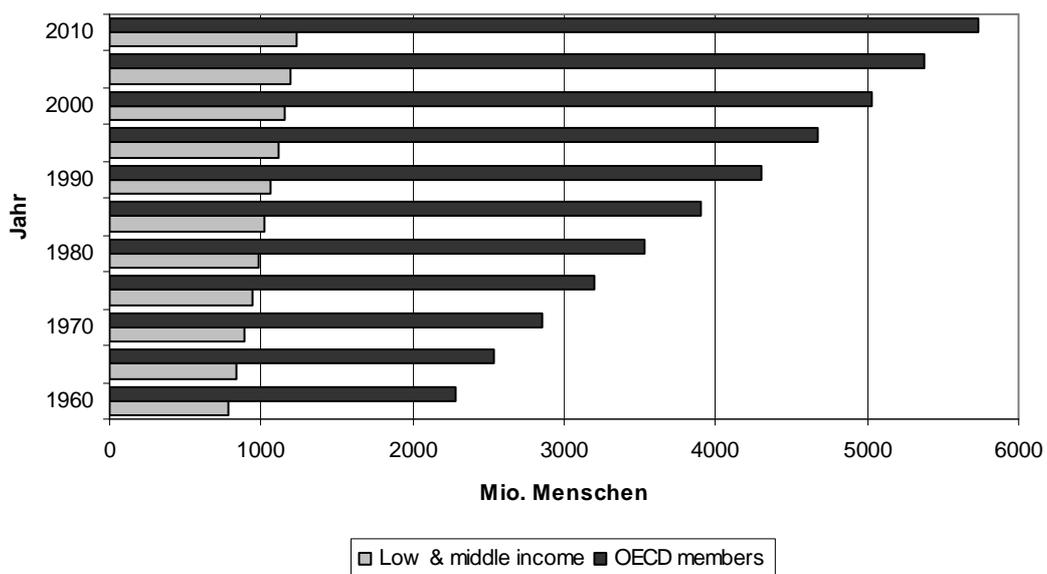


Abb. II.3: Unterschiedliches Wachstum in industrialisierten und „Low and middle Income“ Ländern

Quelle: Weltbank, New York (2011)

## UN-Konferenz

Internationale Konferenz über Bevölkerung und Entwicklung der Vereinten Nationen - IPCD 1994 - in Kairo (5. - 13.9.1994)

Wieviel Menschen *erträgt* diese Welt?

"Pro Sekunde werden drei Menschen geboren, jeden Tag ein neues Braunschweig, jedes Jahrzehnt ein neues Indien ...." (Die Zeit (2.9.1994): Die drei Gebote von Kairo: Mehr Verhütung und Gesundheit, weniger Verschwendung)

## 2.1 Aufgaben und Abgrenzung

Bevölkerungsstatistik für die

- ♦ Beschreibung,
  - ♦ Erklärung und
  - ♦ Prognose
- demografischer Zustände und Prozesse.

### Aufgaben der Bevölkerungsstatistik

Informationen bereitstellen über

#### Bevölkerungsbestand

- Größe
- Zusammensetzung der Bevölkerung

#### Bevölkerungsbewegung

- Veränderung des Bestandes durch Geburten, Sterbefälle, Wanderungen

#### Entwicklung von Bevölkerungsmodellen

- Prognosen, Erklärung
- Wichtig für öffentliche Verwaltung, politische Entscheidungsträger, Firmenpolitik...

## 2.2 Statistik des Bevölkerungsbestandes

### Bevölkerungsbestand

Bestandsdaten zu einem bestimmten Zeitpunkt nach

- Anzahl
- personen- und haushaltsbezogene Merkmale  
wie Alter, Geschlecht, Familienstand, Haushaltsgröße, Konfession, Nationalität

### International: ortsanwesende Bevölkerung

Stichtag, z.B. auch Kur- und Urlaubsgäste

**BRD: Wohnbevölkerung**

## - bis 1983

- Personen mit einer Wohnung: in Wohngemeinde gezählt  
 Personen mit mehreren Wohnungen: Zuordnung über die Wohnung, von der aus der Arbeit oder Ausbildung nachgegangen wurde  
 Nicht in Arbeit /Ausbildung: Zuordnung über überwiegenden Aufenthalt

## - seit 1983

- Hauptwohnungsprinzip (nach neuem Melderecht):  
 Zuordnung zu der Gemeinde, in der sich die **vorwiegend von der Familie genutzte Wohnung** (Hauptwohnung) befindet (**Schwerpunkt der Lebensbeziehungen des Einwohners**)

**Erhebungsmethode:**

Primärerhebungen (Volkszählung, Mikrozensus)

Ergebnisse: Fachserie 1, Reihen 1 und 3 des Statistischen Bundesamtes

Zur Bevölkerungsentwicklung in der BRD vergleiche Tab. II.3.

Tab. II.3a: Bevölkerung in der Bundesrepublik Deutschland 1970, 1987

Land	Jahr	insgesamt	Männlich	Weiblich	Deutsche	Ausländer
		Anzahl				
Schleswig-Holstein	1970	2.494.104	47,6	52,4	98,5	1,5
	1987	2.554.241	48,1	51,9	97,0	3,0
Hamburg	1970	1.793.823	46,1	53,9	96,7	3,3
	1987	1.592.770	46,8	53,2	90,7	9,3
Niedersachsen	1970	7.082.158	47,8	52,2	97,9	2,1
	1987	7.162.103	48,2	51,8	96,3	3,7
Bremen	1970	722.732	47,0	53,0	97,9	2,1
	1987	660.084	47,3	52,7	93,2	6,8
Nordrhein-Westfalen	1970	16.914.118	47,8	52,2	95,9	4,1
	1987	16.711.845	47,9	52,1	92,5	7,5
Hessen	1970	5.381.705	48,1	51,9	94,9	5,1
	1987	5.507.777	48,2	51,8	91,4	8,6
Rheinland-Pfalz	1970	3.645.437	47,7	52,3	97,8	2,2
	1987	3.630.823	48,1	51,9	95,8	4,2
Baden-Württemberg	1970	8.895.048	48,2	51,8	92,8	7,2
	1987	9.286.387	48,3	51,7	90,9	9,1
Bayern	1970	10.479.386	47,4	52,6	96,3	3,7
	1987	10.902.643	48,0	52,0	94,4	5,6
Saarland	1970	1.119.742	47,6	52,4	97,5	2,5
	1987	1.055.660	47,9	52,1	95,9	4,1
Berlin (West)	1970	2.122.346	43,4	56,6	96,5	3,5
	1987	2.012.709	46,5	53,5	88,9	11,1
Bundesgebiet	1970	60.650.599	47,6	52,4	96,0	4,0
	1987	61.077.042	48,0	52,0	93,2	6,8

Quelle: Statistisches Bundesamt (1989)

Tab. II.3b: Bevölkerung in der Bundesrepublik Deutschland 1995, 2009

Land	1995			2009		
	insgesamt	Männlich	Weiblich	insgesamt	Männlich	Weiblich
	Anzahl	%		Anzahl	%	
Baden-Württemberg	10.319.367	48,9	51,1	10.744.921	49,2	50,8
Bayern	11.993.484	48,8	51,2	12.510.331	49,0	51,0
Berlin	3.471.418	48,2	51,8	3.442.675	49,0	51,0
Brandenburg	2.542.042	49,1	50,9	2.511.525	49,5	50,5
Bremen	679.757	48,2	51,8	661.716	48,7	51,3
Hamburg	1.707.901	48,2	51,8	1.774.224	48,8	51,2
Hessen	6.009.913	49,0	51,0	6.061.951	49,0	51,0
Mecklenburg- Vorpommern	1.823.084	49,2	50,8	1.651.216	49,5	50,5
Niedersachsen	7.780.422	48,8	51,2	7.928.815	49,1	50,9
Nordrhein-Westfalen	17.893.045	48,5	51,5	17.872.763	48,8	51,2
Rheinland-Pfalz	3.977.919	48,9	51,1	4.012.675	49,1	50,9
Saarland	1.084.370	48,5	51,5	1.022.585	48,7	51,3
Sachsen	4.566.603	48,1	51,9	4.168.732	48,9	51,1
Sachsen-Anhalt	2.738.928	48,4	51,6	2.356.219	49,0	51,0
Schleswig-Holstein	2.725.461	48,8	51,2	2.832.027	49,0	51,0
Thüringen	2.503.785	48,5	51,5	2.249.882	49,4	50,6
insgesamt	81.817.499	48,6	51,4	81.802.257	49,0	51,0

Quelle: Statistisches Bundesamt (2011)

## **Volkszählung (VZ)**

Eine Volkszählung liefert Grundinformationen über die Bereiche Bevölkerung und Erwerbsleben, wobei eine tiefe Gliederung vorgenommen werden kann.

Volkszählungen sind Totalerhebungen. Die letzte Volkszählung in der Bundesrepublik fand am 09. Mai 2011 statt, davor fanden Volkszählungen in den Jahren 1950, 1961, 1970, 1987 statt.

1970 wurde erstmals eine Erhebung bezüglich des Abschlusses an allgemeinbildenden Schulen (mittlere Reife, Abitur...) vorgenommen.

Die Volkszählung ist eine Vollerhebung mit Auskunftspflicht. 1987 bestand das Zählwerk aus

- Volks- und Berufszählung (VZ)
- Gebäude- und Wohnungszählung (GWZ)
- Arbeitsstättenzählung (AZ)

Die Volkszählung bezieht die gesamte im Bundesgebiet wohnende Bevölkerung ein (Inlandskonzept).

Erhebungseinheiten sind Personen mit Wohnungs- und Haushaltsangaben (vgl. Abb.II.4a,b).

Die zunächst für 1983 geplante Volkszählung scheiterte am Urteil des Bundesverfassungsgerichts vom 15.12.1983, in dem einige Bestimmungen über die

- Weitergabe personenbezogener Daten für verfassungswidrig erklärt und
- verwaltungsmäßige Vorkehrungen für einen noch besseren Datenschutz gefordert werden.

Das Volkszählungsgesetz vom 8.11.1985 trägt diesem Urteil Rechnung.

Weitere Informationen zur Diskussion um die VZ 1983/1987 finden sich in Rinne (1994), S.61ff.

Die Volkszählung ist aus folgenden Gründen notwendig:

- Bestandsaufnahme der Bevölkerung in großen Zeitabständen (gegliedert wie sonst nicht anders erfahrbare)
- Basisfunktion für die gesamte übrige Statistik:  
bildet den Ausgangspunkt für Fortschreibung und Auswahl des Mikrozensus und für viele andere Stichproben
- die Volkszählung erlaubt als umfassende und tiefgehende Rahmencensenzählung die Bedeutung anderer Erhebungen in einem Gesamtzusammenhang einzuordnen. Sie bildet für diese einen Hochrechnungsrahmen.

# VOLKSZÄHLUNG 1987

## Wohnungsbogen mit Gebäudeangaben

*Bitte so markieren*

1

**Rechtsgrundlage:**  
 Siehe Haushaltsmantelbogen oder Erläuterungsblatt, die Bestandteile der Erhebungsvordrucke sind. Stichtag: 25. Mai 1987

Die Wohnung steht leer

Die Wohnung ist von Angehörigen ausländischer Streitkräfte, diplomatischer, berufskonsularischer Vertretungen oder der Ständigen Vertretung der DDR privatrechtlich gemietet.

Bitte Gemeinde angeben

---

**Wohnungsangaben**

**FÜR ALLE HAUSHALTE**

1. Bewohnen Sie die Wohnung/Räume als

a) Eigentümer(in), Mitigent. oder Kaufanwärt.

b) Hauptmieter(in) (einschließlich Altenteil)

c) Untermieter(in)   
*Untermieter(innen) brauchen im Wohnungsbogen keine weiteren Fragen zu beantworten.*

2. Ist die Wohnung eine Freizeitwohnung?  nein  ja

3. Wann ist der Haushalt in diese Wohnung angezogen?  
*Bitte das Einzugsjahr für das am längsten hier wohnende Haushaltsmitglied angeben*

vor 1970

1970-1974

1975-1979

1980-1984

1985

1986

1987

**FÜR EIGENTÜMER, HAUPTMIETER**

4. Sind innerhalb der Wohnung

Küche

Kochnische

Bad/Dusche

WC

Mehrere Antworten möglich

5. a) Wird die Wohnung überwiegend beheizt mit

Fern-/Blockheizung

Zentralheizung

Etagenheizung

Einzel- oder Mehrraumöfen (auch Elektropeicher)

b) Welcher Brennstoff, welche Wärmequelle wird verwendet?

Gas

Heizöl

Strom

Kohle, Holz usw.

Fernwärme

Sonnenenergie, Wärmepumpe

Mehrere Antworten möglich

**FÜR EIGENTÜMER, HAUPTMIETER**

6. Wie viele Räume der Wohnung haben mindestens 6 Quadratmeter?  
*(ohne Küche, Bad, WC, Flur)*

1  2  3  4  5  6  7  8  9

Falls davon Räume untervermietet oder gewerblich genutzt sind:

a) Anzahl der untervermieteten Räume 1  2  3

b) Anzahl der gewerblich genutzten Räume 1  2  3

7. Wie groß ist die Fläche der gesamten Wohnung?  
*(einschließlich Küche, Bad, WC, Flur, Mansarden und ähnliches sowie untervermieteter oder gewerblich genutzter Räume in der Wohnung)*

volle Quadratmeter

8. Ist die Wohnung mit Mitteln des sozialen Wohnungsbaus gefördert?  nein  ja

9. Wie hoch ist die Monatsmiete?  
*Zur Miete rechnen auch die monatlich aufzuwendenden Beträge für Wasser, Kanalisation, Straßenreinigung, Müllabfuhr, Treppenhäusbeleuchtung, Schornsteinreinigung. Nicht zur Miete rechnen Umlagen für Zentralheizung, Warmwasserversorgung, Garagenmiete, Untermietenzuschlag, Zuschlag für Moblierung.*

volle DM

10. a) Handelt es sich um eine Dienst-, Werkwohnung, Berufs- oder Geschäftsmietwohnung?  nein  ja

b) Ist die Wohnung verbilligt, kostenlos überlassen oder ist die Miete wegen finanzieller Vorleistungen (z. B. Mieterdarlehen) ermäßigt?  nein  ja

**FÜR LEERSTEHENDE WOHNUNGEN**

11. Seit wie vielen Monaten steht die Wohnung leer?

bis 3  4-6  7-12  13

**Gebäudeangaben: Bitte nur ausfüllen, wenn Sie Eigentümer(in) oder Verwalter(in) dieses Gebäudes sind.**

1. Gebäudeart

Wohngebäude   
*(Gebäude mindestens zur Hälfte für Wohnzwecke genutzt)*

sonstiges Gebäude mit Wohnraum

bewohnte Unterkunft (z. B. Baracke)

Außerdem bei Wohnheimen:  
 Wird dieses Gebäude vollständig oder teilweise als Wohnheim genutzt?  vollständig  teilweise

2. Wer ist Eigentümer(in), Kaufanwärt. bzw. Erbbauberechtigte(r) des Gebäudes?

Einzelperson oder Ehepaar, Erbengemeinschaft oder ähnliche Personengemeinschaft

Gemeinschaft von Wohnungseigentümern (nur bei Eigentumswohnungen)

gemeinnütziges Wohnungsunternehmen, Wohnungsbaugenossensch. oder Organ der staatl. Wohnungspolitik

freies Wohnungsunternehmen

sonstige(r) Eigentümer(in)

3. Baujahr des Gebäudes  
*(Jahr der Bezugsterminstellung)*

Bei Erweiterungs- und Umbauten ist das ursprüngliche Baujahr anzugeben bei Wiederaufbau nach Trübschaden das Jahr des Wiederaufbaus

bis 1900  1901-18  1919-48  1949-57  1958-62  63-64  65-66  67-68  69-70  71-72  73-74  75-76  77-78  79-80  81-82  83-84  85-86  87

4. Sind Wohnungen im Gebäude mit Mitteln des sozialen Wohnungsbaus gefördert?  
*(nur Erster Forderungsweg)*

keine

alle

nur ein Teil

Fläche (7)      Miete (9)      Gebäude

0

**NUR VOM STATISTISCHEN LANDESAMT AUSZUFÜLLEN**

\* siehe Erläuterungen im Haushaltsmantelbogen

Abb. II.4a: Volkszählung 1987: Wohnungsbogen mit Gebäudeangaben  
 Quelle: Statistisches Bundesamt (1986)

## VOLKSZÄHLUNG 1987 Personenbogen

Bitte so markieren

2 1

**Rechtsgrundlage:**  
Siehe Haushaltsmantelbogen oder Erläuterungsblatt, die Bestandteile der Erhebungsvordrucke sind.  
Stichtag: 25. Mai 1987

Falls Sie einen Abschluß an einer allgemeinbildenden bzw. berufsbildenden Schule/Hochschule haben:

**9** Welchen höchsten allgemeinen Schulabschluß haben Sie?

· · · · ·	Volksschule, Hauptschule
· · · · ·	Realschule/gleichwertiger Abschluß (z. B. Mittlere Reife)
· · · · ·	Hochschulreife (Abitur), Fachhochschulreife

**10** a) Welchen höchsten Abschluß an einer berufsbildenden Schule oder Hochschule haben Sie?

· · · · ·	Berufsfachschule (ohne Berufsschule)
· · · · ·	Fachschule
· · · · ·	Fachhochschule (Ing.-Schule, höhere Fachschule)
· · · · ·	Hochschule (einschließlich Lehrerausbildung)

b) Welche Hauptfachrichtung hat dieser Abschluß?

**11** Falls Sie eine praktische Berufsausbildung (z. B. Lehre) abgeschlossen haben:

a) Auf welchen Lehrberuf bezog sich diese Ausbildung?

b) Wie lange dauerte diese Ausbildung? (Jahre)

**12** Bitte Name und Anschrift Ihrer Arbeitsstätte oder Schule/Hochschule angeben.

Name: \_\_\_\_\_

Straße/Hausnummer: \_\_\_\_\_

PLZ: \_\_\_\_\_ Gemeinde: \_\_\_\_\_

**13** Welches Verkehrsmittel benutzen Sie hauptsächlich (langste Strecke) auf dem Hinweg zur Arbeit oder Schule/Hochschule?

· · · · ·	kein Verkehrsmittel (zu Fuß)
· · · · ·	Fahrrad
· · · · ·	Pkw
· · · · ·	U-Bahn, S-Bahn, Straßenbahn
· · · · ·	Eisenbahn
· · · · ·	Bus, sonst. öffentl. Verkehrsmittel
· · · · ·	sonstiges (Motorrad, Moped, Mofa)

**14** Wieviel Zeit benötigen Sie normalerweise für den Hinweg zur Arbeit oder Schule/Hochschule?

· · · · ·	entfällt, da auf gleichem Grundstück
· · · · ·	unter 15 Minuten
· · · · ·	15 bis unter 30 Minuten
· · · · ·	30 bis unter 45 Minuten
· · · · ·	45 bis unter 60 Minuten
· · · · ·	60 Minuten und mehr

**15** Sind Sie zur Zeit tätig als

· · · · ·	Facharbeiter(in)
· · · · ·	sonstige(r) Arbeiter(in)
· · · · ·	Angestellte(r)
· · · · ·	Auszubildende(r)
· · · · ·	Beamter/Beamtin, Richter(in), Soldat, Zivildienstl. (auch Beamtenanw.)
· · · · ·	Selbständige(r) mit bezahlten Beschäftigten
· · · · ·	ohne bezahlte Beschäftigte
· · · · ·	mithelfende(r) Familienangehörige(r)

**16** Zu welchem Wirtschaftszweig (Branche, Behörde) gehört der Betrieb (Firma, Dienststelle), in dem Sie tätig sind?

**17** Welche Tätigkeit, welchen Beruf üben Sie aus?

**18** Falls Sie eine Nebenerwerbstätigkeit ausüben, handelt es sich um eine

· · · · ·	landwirtschaftliche
· · · · ·	nichtlandwirtschaftliche

**FÜR ALLE PERSONEN**

Bitte Gemeinde angeben: \_\_\_\_\_

**1** Geburtsangaben

a) Geburtsjahr: \_\_\_\_\_

b) Geburtsmonat: 1. Januar bis 24. Mär. · · · · ·  
25. Mär. bis 31. Dez. · · · · ·

**2** Geschlecht: männlich · · · · ·  
weiblich · · · · ·

**3** Familienstand: ledig · · · · ·  
verheiratet · · · · ·  
verwitwet · · · · ·  
geschieden · · · · ·

**4** Rechtliche Zugehörigkeit zu einer Religionsgesellschaft

· · · · ·	Römisch-katholische Kirche
· · · · ·	Evangelische Kirche
· · · · ·	Evangelische Freikirche
· · · · ·	Jüdische Religionsgesellschaft
· · · · ·	Islamische Religionsgemeinschaft
· · · · ·	andere Religionsgesellschaften
· · · · ·	keiner Religionsgesellschaft rechtlich zugehörig

**5** Welche Staatsangehörigkeit haben Sie?

· · · · ·	deutsch
· · · · ·	griechisch
· · · · ·	italienisch
· · · · ·	übrige EG-Staaten
· · · · ·	jugoslawisch
· · · · ·	türkisch
· · · · ·	sonstige/keine

**6** Wird von Ihnen noch eine weitere Wohnung (Unterkunft/Zimmer) in der Bundesrepublik Deutschland einschließlich Berlin (West) bewohnt?

nein · · · · ·  
ja · · · · ·

Falls ja:

a) Für Verheiratete, die nicht dauernd getrennt leben ist die hiesige Wohnung die vorwiegend benutzte Wohnung der Familie? nein · · · · ·  
ja · · · · ·

b) Für alle übrigen Personen ist die hiesige Wohnung die vorwiegend benutzte Wohnung? nein · · · · ·  
ja · · · · ·

c) Außerdem für Erwerbstätige, Schüler/Studenten: Gehen Sie vorwiegend von der hiesigen Wohnung aus zur Arbeit oder Schule/Hochschule? nein · · · · ·  
ja · · · · ·

**7** Sind Sie erwerbstätig?

· · · · ·	Vollzeit (über 36 Std. <sup>2)</sup> in der Woche)
· · · · ·	Teilzeit (bis zu 36 Std. <sup>2)</sup> in der Woche)
· · · · ·	arbeitslos, arbeitssuchend
· · · · ·	nicht erwerbstätig
· · · · ·	Hausfrau, Hausmann
· · · · ·	Schüler(in), Student(in)

Mehrere Antworten möglich

**8** Leben Sie überwiegend von

· · · · ·	Erwerbs-, Berufstätigkeit
· · · · ·	Arbeitslosengeld, -hilfe
· · · · ·	Rente, Pension
· · · · ·	eigenem Vermögen, Vermietung, Verpachtung, Altenteil
· · · · ·	Zuwendungen, Unterhalt durch Eltern, Ehegatten usw.
· · · · ·	sonstigen Unterstützungen (z. B. Sozialhilfe, BAföG)

**NUR VOM STATISTISCHEN LANDESAMT AUSZUFÜLLEN**

Geburtsjahr (1 a) 2, 3 u. 4 Stelle      Hauptfachrichtung (10 b)      Elterner Beruf (11 a)      Dauer (11 b)

Arbeitsstätte, Schule/Hochschule (Pendler) Land (12) Gemeinde (12)      Straße (12)      Hausnummer (12)

Wirtschaftszweig (16)      Ausgewählte Tätigkeit (17)

\* siehe Erläuterungen im Haushaltsmantelbogen oder im Erläuterungsblatt

Abb. II.4b: Volkszählung 1987: Personenbogen  
Quelle: Statistisches Bundesamt (1986)



noch: Persönliche Angaben	noch: Persönliche Angaben
<p><b>6 Welche Staatsangehörigkeit/-en haben Sie?</b> Mehrfachnennungen sind möglich.</p> <p>Deutsche Staatsangehörigkeit ..... <input type="checkbox"/></p> <p>Staatsangehörigkeit eines anderen EU-Staates .... <input type="checkbox"/></p> <p>Staatsangehörigkeit eines Nicht-EU-Staates ..... <input type="checkbox"/></p> <p>Staatenlos ..... <input type="checkbox"/></p> <p>Ungeklärt ..... <input type="checkbox"/></p>	<p><b>9 Welchen Familienstand haben Sie?</b></p> <p>Ledig ..... <input type="checkbox"/></p> <p>Verheiratet ..... <input type="checkbox"/></p> <p>Geschieden ..... <input type="checkbox"/></p> <p>Verwitwet ..... <input type="checkbox"/></p> <p>Eingetragene Lebenspartnerschaft (gleichgeschlechtlich) ..... <input type="checkbox"/></p> <p>Eingetragene Lebenspartnerschaft (gleichgeschlechtlich) aufgehoben ..... <input type="checkbox"/></p> <p>Eingetragener Lebenspartner/ eingetragene Lebenspartnerin (gleichgeschlechtlich) verstorben ..... <input type="checkbox"/></p>
<p><b>7 Welcher Religionsgesellschaft gehören Sie an?</b></p> <p>Römisch-katholische Kirche ..... <input type="checkbox"/></p> <p>Evangelische Kirche ..... <input type="checkbox"/></p> <p>Evangelische Freikirchen ..... <input type="checkbox"/></p> <p>Orthodoxe Kirchen ..... <input type="checkbox"/></p> <p>Jüdische Gemeinden ..... <input type="checkbox"/></p> <p>Sonstige öffentlich-rechtliche Religionsgesellschaft ..... <input type="checkbox"/></p> <p>Keiner öffentlich-rechtlichen Religionsgesellschaft ..... <input type="checkbox"/></p>	<p><b>10 Wohnen Sie in Ihrer Wohnung mit einem Partner/einer Partnerin in einer Lebensgemeinschaft zusammen, die weder Ehe noch eingetragene gleichgeschlechtliche Lebenspartnerschaft ist?</b></p> <p>Ja ..... <input type="checkbox"/></p> <p>Nein ..... <input type="checkbox"/></p>
<p><b>8 Zu welcher der folgenden Religionen, Glaubensrichtungen oder Weltanschauungen bekennen Sie sich?</b></p> <p style="background-color: #cccccc; padding: 2px;">Die Beantwortung der Frage ist freiwillig.</p> <p>Christentum ..... <input type="checkbox"/></p> <p>Judentum ..... <input type="checkbox"/></p> <p>Islam</p> <p>    Sunnitischer ..... <input type="checkbox"/></p> <p>    Schitischer ..... <input type="checkbox"/></p> <p>    Alevitischer ..... <input type="checkbox"/></p> <p>Buddhismus ..... <input type="checkbox"/></p> <p>Hinduismus ..... <input type="checkbox"/></p> <p>Sonstige Religion, Glaubensrichtung oder Weltanschauung ..... <input type="checkbox"/></p> <p>Keiner Religion, Glaubensrichtung oder Weltanschauung ..... <input type="checkbox"/></p>	<p><b>11 Wie viele Personen leben insgesamt in Ihrer Wohnung?</b></p> <p>Anzahl der Personen (Sie einbezogen) ..... <input type="text"/></p> <p><b>12 Bewohnen Sie eine weitere Wohnung in Deutschland?</b></p> <p>Ja ..... <input type="checkbox"/></p> <p>Nein ..... <input type="checkbox"/></p> <p><b>13 Hauptwohnsitz</b></p> <p><small>Bitte beantworten Sie abhängig von Ihrem Familienstand nur eine der beiden Fragen.</small></p> <p>Für Verheiratete bzw. in eingetragener Lebenspartnerschaft (gleichgeschlechtlich) Lebende, die nicht dauernd getrennt leben:</p> <p><b>Ist die hiesige Wohnung die vorwiegend benutzte Wohnung der Familie?</b></p> <p>Ja ..... <input type="checkbox"/></p> <p>Nein ..... <input type="checkbox"/></p> <p>Für alle übrigen Personen:</p> <p><b>Ist die hiesige Wohnung die vorwiegend benutzte Wohnung?</b></p> <p>Ja ..... <input type="checkbox"/></p> <p>Nein ..... <input type="checkbox"/></p>

Abb. II.4d: Volkszählung 2011: Personenbogen (Seite2/10)  
Quelle: Statistisches Bundesamt (2011), Fragebögen erhältlich unter www.zensus2011.de

Weitere Informationen zu diesem Thema: [www.destatis.de/zensus/index.htm](http://www.destatis.de/zensus/index.htm)

Weitere Hinweise zur Volkszählung: Rinne (1994), Kap. 3.2, Statistisches Bundesamt (1986).

### **Mikrozensus (MZ)**

Der Mikrozensus ist die amtliche Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt in Deutschland. Die Arbeitskräftestichprobe der Europäischen Union (EU-Arbeitskräftestichprobe) ist in den Mikrozensus integriert. Bereits seit 1957 - in den neuen Ländern (einschließlich Berlin-Ost) seit 1991 - liefert der Mikrozensus statistische Informationen in tiefer fachlicher und regionaler Gliederung über die Bevölkerungsstruktur, die wirtschaftliche und soziale Lage der Bevölkerung, der Familien, Lebensgemeinschaften und Haushalte, die Erwerbstätigkeit, Arbeitsuche, Aus- und Weiterbildung, Wohnverhältnisse und Gesundheit.

Kennzeichnend für die Erhebungsform des Mikrozensus bis 2004 war das Konzept der festen Berichtswoche, das heißt die meisten Fragen bezogen sich auf die Gegebenheiten in einer einzelnen Berichtswoche im Jahr. In der Regel handelte es sich um die letzte feiertagsfreie Woche im April. Die Mikrozensusergebnisse bis 2004 lieferten somit eine Momentaufnahme der Verhältnisse im Frühjahr und waren - je nach Merkmal mehr oder weniger stark - durch saisonale Schwankungen beeinflusst.

Das MZG 2005 ordnet eine unterjährige, kontinuierliche Erhebung an. Bei dieser Erhebungsform soll das gesamte Befragungsvolumen gleichmäßig auf alle Kalenderwochen des Jahres verteilt werden, wobei die letzte Woche vor der Befragung die Berichtswoche darstellt (so genannte gleitende Berichtswoche). Damit sollten den Nutzerinnen und Nutzern des Mikrozensus neben jährlichen auch vierteljährliche Durchschnittsergebnisse zur Verfügung gestellt werden. Die vierteljährliche Veröffentlichung von Quartalsergebnissen wird allerdings wegen methodischer Schwierigkeiten gegenwärtig noch geprüft.

Der Auswahlsatz liegt für alle Merkmale einheitlich bei 1% der Bevölkerung

Erhebungsmerkmale:

Demografische Merkmale, Fragen nach Erwerbsbeteiligung, überwiegende Unterhaltsquelle, Wirtschaftszweig, Stellung im Beruf, wöchentliche Arbeitszeit und Nettoeinkommen (Nebenerwerb); siehe auch Sfb 3 Nebenerwerbstätigkeitsumfrage '84 (Merz, Helberger, Schneider (1985)). Inhaltliche Neuerungen im Erhebungsprogramm des Mikrozensus ab 2005 bestehen im Wesentlichen in der Aufnahme des neuen Themenkomplexes "Migration und Integration".

Erhebungszeitraum:

unterjährige, kontinuierliche Erhebung.

Die letzte Woche vor der Befragung stellt die Berichtswoche dar (so genannte gleitende Berichtswoche).

Es gibt Zusatzerhebungen zum Mikrozensus (wie z.B. zur Situation berufstätiger Mütter, Ausbildung und berufliche Fortbildung).

Der MZ ist eine Stichprobe mit geschichteter Zufallsauswahl und geschulten Interviewern

Der MZ wird zur laufenden Fortschreibung der Volkszählung verwendet (→ Fehler z.B. in der Volkszählung 1987, eine Millionen Wohnungen fehlend).

Eine Regionalisierung ist wegen des kleinen Auswahlsatzes nur begrenzt möglich.

Kurzinformationen für die Befragten des Mikrozensus 1991 finden sich in Abb. II.5. Zur internationalen Verbreitung vgl. Abb. II.6.

Weitere Informationen: Esser, Grohmann, Müller und Schäfer (1989).



## Mikrozensus

### - Kurzinformation für die Befragten -

**➤ Was ist der Mikrozensus?**

Der Mikrozensus ist eine amtliche Befragung bei ein Prozent der Bevölkerung, bei der die Mitglieder der ausgewählten Haushalte grundsätzlich durch Erhebungsbeauftragte interviewt werden. Seit 1987 ermittelt die amtliche Statistik grundlegende Daten über die Struktur der Bevölkerung, die Entwicklung des Arbeitsmarktes, die Art der Erwerbstätigkeit sowie über Formen des Zusammenlebens in allen Ländern der Bundesrepublik Deutschland. Ihre Durchführung ist gesetzlich vorgeschrieben und geregelt.

**➤ Wozu dient der Mikrozensus?**

Um nur einige Beispiele zu nennen: Wie groß ist die Zahl allein stehender Frauen und Männer, allein erziehender Mütter und Väter, Kinderreicher Familien, älterer Menschen, die in Einpersonenhaushalten oder Heimen leben? Wie viele Menschen in den verschiedenen Regionen Deutschlands erwerbstätig sind, in welchen Berufen, welchen Branchen sie arbeiten? Das wissen wir nicht ohne die Ergebnisse des Mikrozensus.

Die Ergebnisse werden von den Statistischen Landesämtern und vom Statistischen Bundesamt in Wiesbaden u. a. im Internet veröffentlicht. Sie stehen allen interessierten Bürgerinnen und Bürgern zur Verfügung, nicht nur der Regierung und Verwaltung, der Wirtschaft, Wissenschaft und Presse.

Weitergehende Informationen enthält die Broschüre „Information für die Befragten“, die Ihnen die/der Erhebungsbeauftragte gern aushändigt.

**➤ Warum werden gerade Sie befragt?**

Nach einem mathematisch-statistischen Zufallsverfahren wurden in Hamburg etwa 1.100 Auswahlbezirke und in Schleswig-Holstein etwa 1.700 Erhebungsbeauftragten befragt, die über 9.000 Haushalte in Hamburg und 14.000 Haushalte in Schleswig-Holstein in diesen Bezirken maximal 4 aufeinander folgenden Jahren hinterreinander. Auch Ihr Haushalt gehört dazu. Da Stichprobenergebnisse nur dann zuverlässig sind, wenn die Auswahlordnung eingehalten wird, kann Ihr Haushalt nicht gegen einen anderen ausgetauscht werden.

**Wir bitten Sie für den Mikrozensus um Ihre Mitarbeit!**

---

**➤ Sind Sie zur Auskunft verpflichtet?**

Ja, Sie sind zur Auskunft verpflichtet!

Gerade bei Stichproben ist die Vollständigkeit der Auskünfte besonders wichtig. Deshalb schreibt auch das Mikrozensusgesetz die Auskunftspflicht für Volljährige, sowie für Minderjährige, die einen eigenen Haushalt führen, vor. Darüber hinaus sind Sie auch zur Auskunft für minderjährige oder solche Mitglieder Ihres Haushaltes, die auf Grund einer Behinderung nicht selbst antworten können, verpflichtet. Es kann aber auch eine andere Vertrauensperson beauftragt werden. Widerspruch und Anfechtungsklage gegen die Auforderung zur Auskunftserteilung haben keine aufschiebende Wirkung.

Für einen Teil des Frageprogramms ist die Auskunftserteilung freiwillig, darauf wird im Einzelnen hingewiesen.

In unserem Auftrag wird Ihnen ein/e Erhebungsbeauftragte/r in den nächsten Tagen einen Besuchstermin für das Interview vorschlagen.

**➤ Welche Möglichkeiten der Auskunftserteilung bestehen? Können Sie auch schriftlich Auskunft erteilen?**

Es bestehen drei Möglichkeiten Ihrer Auskunftspflicht nachzukommen:

- Das persönliche Interview
- Das telefonische Interview
- Der Haushalt füllt den Erhebungsbogen selbst aus („Selbstausschüttung“)

Grundsätzlich wird das Interview mit Unterstützung eines Laptops durch die/den Erhebungsbeauftragten durchgeführt und hat sich bewährt. Die besonders geschulten Erhebungsbeauftragten sind mit dem Fragen vertraut.

Falls Sie aus irgendwelchen Gründen die Auskunft nicht in Form des Interviews geben wollen, können Sie auch als „Selbstausschütter“ schriftlich Auskunft erteilen und den ausgefüllten Fragebogen innerhalb einer Woche dem Statistischen Amt ausreißend frankiert zusenden. Vermerken Sie in diesem Fall bitte Ihren Namen und Ihre Anschrift auf dem Umschlag. Fehlen diese Angaben, kann der Fragebogen nicht bearbeitet werden und gilt dann als nicht abgegeben. Bitte beantworten Sie die Fragen vollständig und wahrheitsgemäß.

---

**➤ Wie werden Ihre Angaben geheim gehalten?**

Die bei Ihnen erhobenen Einzelangaben werden nach dem § 18 des Bundesstatistikgesetzes grundsätzlich geheim gehalten. Sie dürfen nur für die gesetzlich bestimmten Zwecke verwendet werden. Alle Mitarbeiterinnen und Mitarbeiter des Statistischen Amtes sowie die Erhebungsbeauftragten sind gesetzlich zur Geheimhaltung verpflichtet. Die ehrenamtlich tätigen Erhebungsbeauftragten sind vom Statistischen Amt mit besonderer Sorgfalt ausgewählt und geschult worden. Interessenskonflikte auf Grund Ihrer beruflichen oder dienstlichen Tätigkeit sind ausgeschlossen. Sie sind schriftlich verpflichtet worden, über die Wahrung des Statistikerheimnisses, hinaus sämtliche Erkenntnisse über Auskunftspläne im Zusammenhang mit Ihrer Interviewtätigkeit geheim zu halten. Sie können sich durch einen Ausweis des Statistischen Amtes legitimieren. Die Erhebungsbeauftragten sind also Vertrauenspersonen, die Ihnen mit Rat und Tat zur Verfügung stehen.

**➤ Zusätzliche Informationen zur EU-Arbeitskräfteerhebungsprobe**

Die Mitgliedsstaaten der Europäischen Union (EU) führen eine gemeinsame Arbeitskräfteerhebungsprobe seit 1988 regelmäßig durch. Sie ist wie der Mikrozensus eine amtliche Haushaltsbefragung und dient der Ermittlung wichtiger, international vergleichbarer Ergebnisse. Mit Ihren Daten liefert die EU-Arbeitskräfteerhebungsprobe Grundlagen für arbeitsmarkt- und regionalpolitische Initiativen der EU (z. B. Verteilung der Mittel aus dem EU-Sozialfonds zur Unterstützung strukturschwacher Gebiete).

Beide Erhebungen, Mikrozensus und EU-Arbeitskräfteerhebungsprobe werden gemeinsam durchgeführt. Dadurch reduzieren sich die zeitliche Belastung der Befragten sowie die Erhebungskosten in erheblichem Maße.

**➤ Welche Fragen werden gestellt?**

Die Fragen richten sich an alle Mitglieder Ihres Haushalts. Gefragt wird z. B. nach den Angaben zur Person, der Erwerbstätigkeit und dem Beruf, der Arbeitssuche, der Bildung, der Altersversorgung sowie nach dem Lebensunterhalt.

**➤ Auf welcher Rechtsgrundlage beruht der Mikrozensus?**

Rechtsgrundlagen sind das Mikrozensusgesetz 2005 (MZG 2005) vom 24. Juni 2004 (BGBl. I S. 1350) in Verbindung mit der Verordnung (EG) Nr. 577/98 des Rates vom 9. März 1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft (ABl. EG Nr. L 77 S. 3) sowie diverse Ausführungsverordnungen der Kommission und das Bundesstatistikgesetz (BStatG) vom 22.1.1987 (BGBl. I S. 462, 565) in der jeweils gültigen Fassung.

Die gesetzlichen Regelungen zum Mikrozensus wurden bereits 1988 vom Bundesverfassungsgericht überprüft. In seinen Beschlüssen vom 1. März 1988 - 1 BvR 93/88 und 15. April 1988 - 1 BvR 222/88 hat es dabei u. a. festgestellt, die gesetzlich festgelegte Auskunftsspflicht stehe im Einklang mit dem im Volkszählungsurteil entwickelten verfassungsrechtlichen Grundsätzen. Das Erhebungsprogramm greife nicht in den Kernbereich privater Lebensgestaltung ein und führe auch nicht zu einer mit der Würde des Menschen unvereinbaren Registrierung oder Katalogisierung der Persönlichkeit.

**➤ Ist der Datenschutz gewährleistet?**

Ja. Der Gesetzgeber hat genaue Regelungen erlassen, um den Datenschutz zu gewährleisten. Die Datenschutzbeauftragten der Länder Hamburg und Schleswig-Holstein haben das Verfahren des Mikrozensus begutachtet und ihre Zustimmung gegeben.

Abb. II.5: Mikrozensus 2011 - Kurzinformationen für die Befragten

Quelle: Statistisches Bundesamt

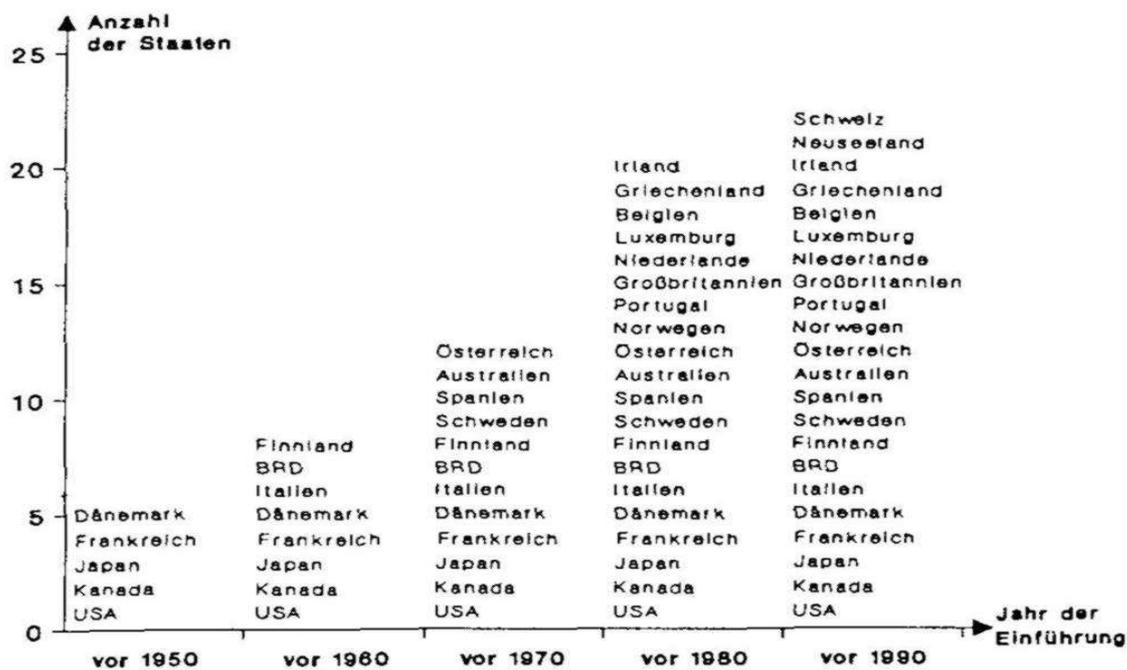


Abb. II.6: Zur internationalen Verbreitung von Mikrozensus: Einführung mikrozensusähnlicher Erhebungen in den OECD-Mitgliedsstaaten  
Quelle: Esser, Grohmann, Müller und Schäfer (1989), S. 76

## Bevölkerungsbestandsmerkmale

Struktur der Bevölkerung über Merkmale wie Alter, Geschlecht, Familienstand, Haushaltsgröße etc.:

### Bestand

Tab. II.4: Bevölkerung in der BRD und in Niedersachsen

	BRD	Niedersachsen	%
1970	61,001 Mio.	7,122 Mio.	11,7
1987	61,238 Mio.	7,164 Mio.	11,7
1991	80,275 Mio.	7,476 Mio.	9,3
2000	82,260 Mio.	7,926 Mio.	9,6
2009	81,802 Mio.	7,929 Mio.	9,7

Quelle: Statistisches Bundesamt (2011) (vgl. Abb. II.7)

### Altersaufbau

Alterspyramide (vgl. Abb. II.7,8)

Form geprägt durch Geburten, Sterbefälle, Wanderungen

Auswirkungen historischer Ereignisse werden gut sichtbar

### Geschlechterverhältnis

$$GV = \frac{\text{Frauen}(F)}{\text{Männer}(M)} \cdot 1000$$

Gibt an, wieviele Frauen rechnerisch auf 1000 Männer entfallen.  
z.B. Wohnbevölkerung der BRD 31.12.2009

$$GV = \frac{41,70 \text{ Frauen}}{40,10 \text{ Männer}} \cdot 1000 = 1040$$

- Altersspezifisches GV: In vielen Ländern kann beobachtet werden, dass
- weniger Mädchen als Jungen geboren werden (Neugeborenen GV < 1000)
  - durch höhere Sterblichkeit bei Männern in allen Altersklassen gilt:
    - GV in mittleren Altersgruppen nähert sich 1000
    - GV in höheren Altersgruppen > 1000

### Altersaufbau der Bevölkerung Deutschlands am 31.12.2000

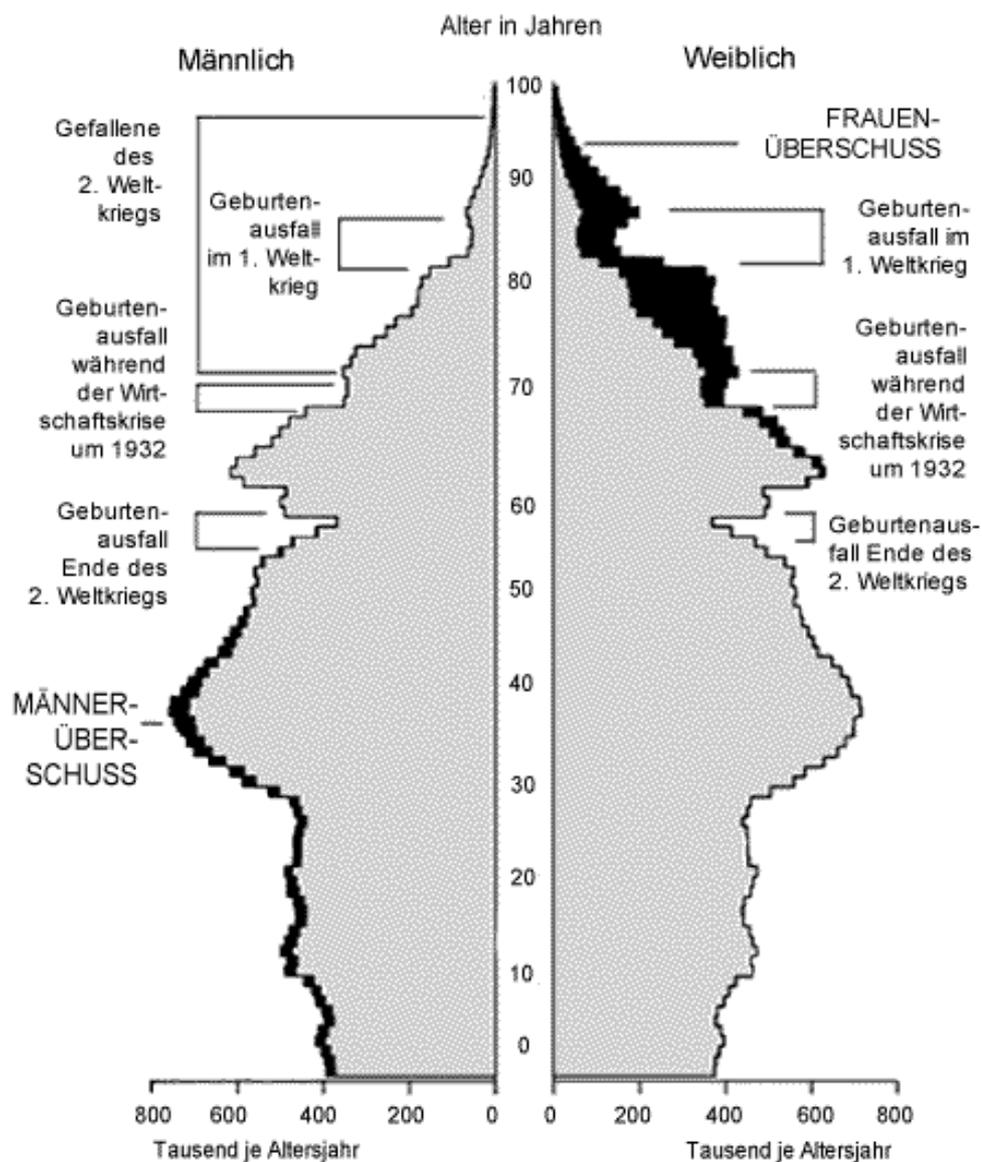


Abb. II.7: Kommentierter Altersaufbau der Bevölkerung der Bundesrepublik Deutschland am 31.12.2000  
Quelle: Statistisches Bundesamt (2002)

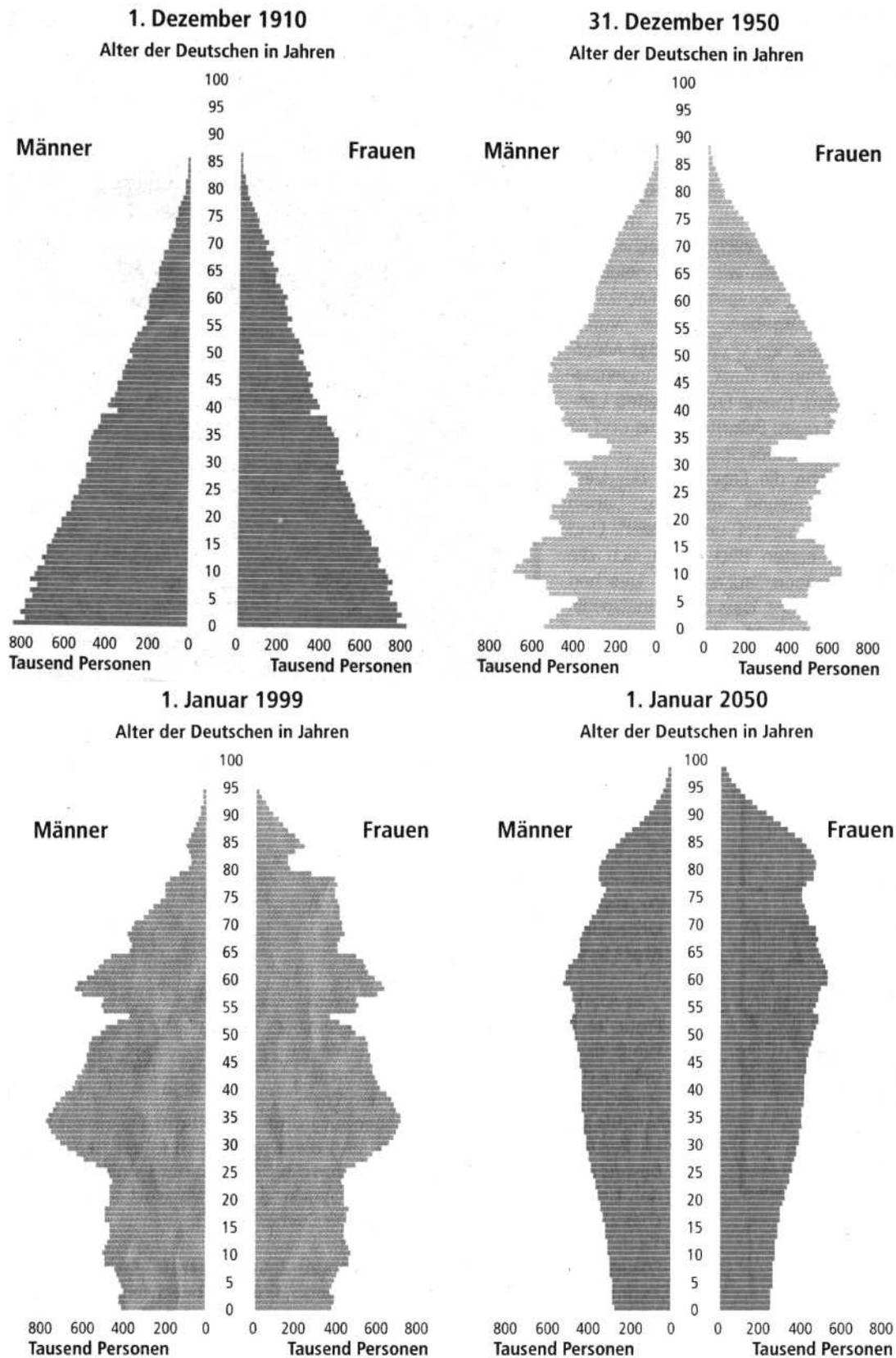


Abb. II.8: Altersaufbau in Deutschland, 1910 bis 2008: Fortschreibung, 2009 bis 2060  
Vorausberechnung

Quelle: Der siebte Tag (30.11.2002), S. 2-3; Statistisches Bundesamt (2009)

## Erwerbstätigkeit

Verbindung zwischen Bevölkerungs- und Wirtschaftsstatistik

z.B.:

$$\text{Allgemeine Erwerbsquote} = \frac{\text{Erwerbspersonen}}{\text{Wohnbevölkerung}} \cdot 100[\%]$$

z.B. 2010 (in Mio.):

Tab. II.5: Bevölkerung nach Erwerbstätigkeit (Männer und Frauen)

	Männer	Frauen
Erwerbstätige	20,637	17,633
+ Erwerbslose	1,695	1,294
= Erwerbspersonen	22,333	18,882

Quelle: Statistisches Bundesamt, Wiesbaden 2011

Mit einer Wohnbevölkerung von 53,902 Mio. und 38,270 Mio. Erwerbspersonen ist die

$$\text{allgemeine EQ}_{2010} = \frac{38,270}{53,902} \cdot 100 = 70,88 \%$$

Spezifische Erwerbsquoten (z.B. alters-, geschlechts-, familienstandsspezifische Quoten):

$$\text{spezifische EQ}_{2010} (\text{Männer}) = \frac{20,637}{53,902 \cdot 0,50} \cdot 100 = 76,6 \%$$

$$\text{spezifische EQ}_{2010} (\text{Frauen}) = \frac{17,633}{53,902 \cdot 0,50} \cdot 100 = 65,4 \%$$

Anmerkung: Der Bevölkerungsanteil der Männer/Frauen ist 2010 ca. 50% / 50%.

## Familienstand

Bevölkerung aufgeteilt nach: Ledige, Verheiratete, Verwitwete und Geschiedene:

Tab. II.6: Bevölkerung nach Familienstand

	BRD 2010					insgesamt
	ledig	verheiratet zusammen lebend	verheiratet getrennt lebend	verwitwet	geschieden	
Männer	44,3	45,4	2,0	2,6	5,6	40223
Frauen	36,0	44,0	1,8	11,0	7,3	41534

Quelle: Statistisches Bundesamt, Wiesbaden (2011)

## Haushaltstyp

Privathaushalte: z.B. x-Personenhaushalt (in BRD) (vgl. auch Tab. II.7)

Tab. II.7: Bevölkerung nach Privathaushalten

Jahr	Insgesamt in Mio.	Personen (Haushalte in %)			
		1	2	3	4 und mehr
1970	21,991	25,1	27,1	19,6	28,2
1987	27,006	33,4	28,4	17,7	20,4
1990	28,175	35,0	30,2	16,7	18,1
2000	38,124	36,1	33,4	14,7	15,9
2010	40,301	40,2	34,2	12,6	13,0

1970, 1987: Ergebnis der VZ. Bis 1990: Früheres Bundesgebiet.

Quelle: Statistisches Bundesamt (1989), VZ 1970, 1987, Statistisches Bundesamt (2011)

Zur Struktur privater Haushalte: Rapin (1990)

## Bevölkerung und Fläche

Bevölkerungsdichte: durchschnittliche Einwohnerzahl pro Flächeneinheit

$$\text{allg. Bevölkerungsdichte} = \frac{\text{Bevölkerung}}{\text{Fläche in Quadratkilometern}}$$

Die regionale Bevölkerungsdichte ist wichtig für Verkehrs- und Wohnungsbauplanung.

Das Spektrum der Bevölkerungsdichte in den Flächenländern im Jahr 2000 reichte von 528 Einwohner pro qkm in Nordrhein-Westfalen bis 77 Einwohner pro qkm in Mecklenburg-Vorpommern.

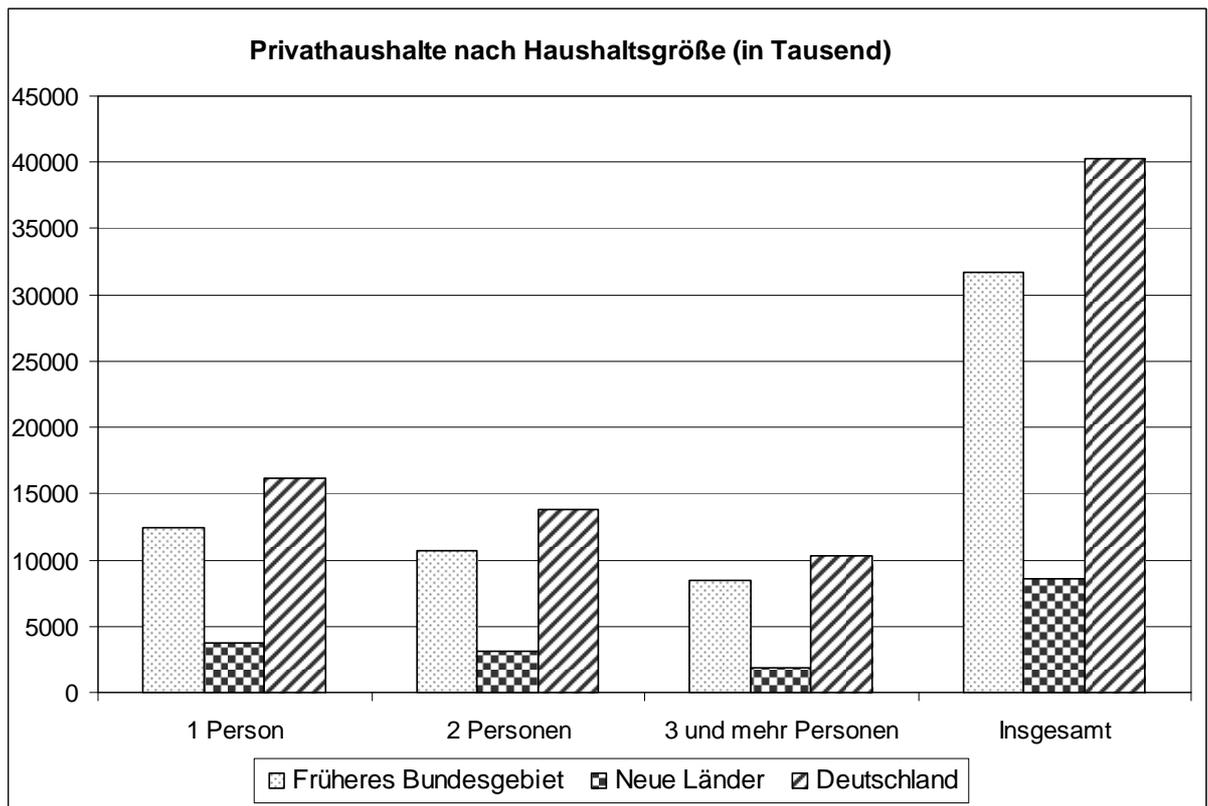


Abb. II.9: Privathaushalte nach Haushaltsgröße 2010  
Quelle: Statistisches Bundesamt (2011)

## 2.3 Statistik der Bevölkerungsbewegung

Größen, die den Bevölkerungsbestand verändern:

- Geburten, Sterbefälle, Wanderungen
- Eheschließungen, Ehescheidungen

dienen zur:

- Fortschreibung des Bevölkerungsbestandes
- Entscheidungen über zukünftigen Bedarf, z.B. an Ausbildungseinrichtungen, Wohnraum etc.

### Erhebungsmethoden

- Geburten, Sterbefälle, Eheschließungen: Standesämter
- Ehescheidungen: Gerichte
- Zu- und Fortzüge: Meldebehörden (Einwohnermeldeamt)

### Bevölkerungsbewegungsmerkmale:

Statistiken oft als Relation von Bewegungsmasse dividiert durch die sie hervorbringende Bestandsmasse:

### Allgemeine Geburten- und Sterbeziffer

Gesamtzahl der Geborenen in einem Jahr, bezogen auf den mittleren Bevölkerungsbestand (Mittel zwischen Anfangs- und Endbestand des Jahres):

$$\text{Allg. Geburtenziffer} = \frac{\text{Zahl der Lebendgeborenen im Jahr } i \ (G)}{\text{Mittlerer Bevölkerungsbestand im Jahr } i \ (B)} \cdot 1000$$

Ergibt Geburten je 1000 Einwohner

Nach dem gleichen Prinzip werden die allg. Sterbeziffer, allg. Heiratsziffer und allg. Ehescheidungs ziffer errechnet.

Interpretation:

Die Aussagefähigkeit ist beschränkt, da der Nenner unterschiedliche Strukturen hat.

Z.B. niedrige allg. Geburtenziffer nicht nur

- bei geringer Geburtenzahl von Frauen im gebärfähigen Alter, sondern auch
- bei hohem Bevölkerungsanteil alter Menschen

Verbesserte Aussagefähigkeit durch:

### Besondere Geburten- und Sterbeziffer

Fertilitätsrate, auch allg. Fruchtbarkeitsrate:

Zahl der Geborenen/Zahl der Frauen im gebärfähigen Alter (amtliche Statistik: 15 bis unter 45jährige)

$$\text{allg. Fertilitätsrate} = \frac{\text{Lebendgeborene im Jahr } i \cdot 1000}{\text{Anzahl der Frauen zwl 15 u. 45 Jahren im Jahr}}$$

Die Fertilitäts-(Fruchtbarkeits-)raten haben in Deutschland deutlich abgenommen (vgl. Tab. II.8)

Da auch innerhalb solcher Gruppen die Alterszusammensetzung noch eine erhebliche Rolle spielt, existieren zusätzliche Ziffern für Teilaltersgruppen, z.B.:

$$\begin{aligned} \text{altersspezifische Fruchtbarkeitsrate} &= \frac{\text{Anzahl der im Kalenderjahr von Frauen} \\ &\quad \text{im Alter } i \text{ lebendgeborenen Kinder} \cdot 1000}{\text{Anzahl der Frauen im Alter } i} \\ &= \frac{G_i}{F_i} \cdot 1000 = g_i \end{aligned}$$

Analoges gilt für die Sterbe-, Eheschließungs- und Ehescheidungs ziffern

Zur Entwicklung von Eheschließungen, Lebendgeborenen und Gestorbenen vgl. Abb. II.10.

Tab. II.8: Geburtenhäufigkeit in Deutschland (Lebendgeborene auf 1000 Frauen im Alter von 15 bis unter 49 Jahren)

	1960	1970	1980	1990	2000	2009
Ab 1991	2366	2016	1445	1454	1379	1358
Gesamtdeutschland						

Quelle: Sachverständigenrat Wirtschaft, Zeitreihen für Deutschland (2011)

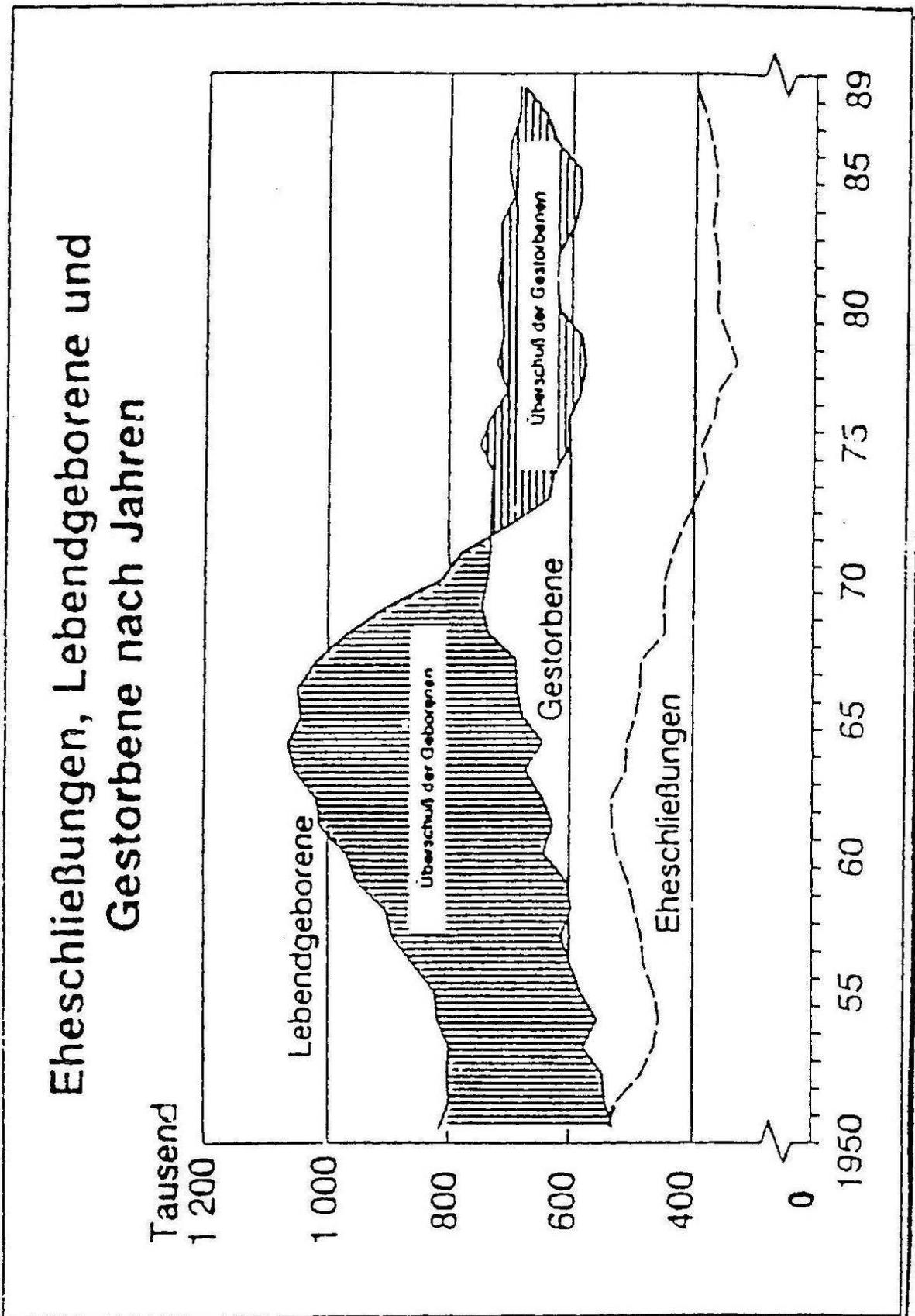


Abb. II.10: Eheschließungen, Lebendgeborene und Gestorbene in Deutschland

## Standardisierte Geburten- und Sterbeziffer

Für den Vergleich zwischen Ländern oder Perioden mit unterschiedlichem Altersaufbau verwendet man standardisierte Ziffern.

Die altersspezifische Fruchtbarkeitsrate  $g_i$  wird gewichtet mit dem Anteil der Frauen dieses Jahrgangs an der Gesamtzahl der Frauen ( $F_i/F$ ).

$$\sum_{i=1}^n g_i \frac{F_i}{F} = \sum_i \frac{G_i}{F_i} \cdot \frac{F_i}{F} \cdot 1000 = \sum_i \frac{G_i}{F} \cdot 1000 = \frac{\sum_i G_i}{F} \cdot 1000 = \frac{G}{F} \cdot 1000 \quad \left( \frac{\text{Geburten}}{\text{Frauen}} \right)$$

Die Effekte des Altersaufbaus lassen sich nun eliminieren, indem die altersspezifischen Fruchtbarkeitsraten auf eine **Standardbevölkerung** bezogen werden (übliche Annahme der Standardbevölkerung: in jeder Altersgruppe befänden sich 1000 Frauen).

→ standardisiertes Maß: Gesamtindex der Fruchtbarkeit oder

$$\text{totale Fruchtbarkeitsrate: } \sum_i g_i = \sum_i \frac{G_i}{F_i} 1000$$

Analoges gilt für Sterberaten etc.

## Reproduktionsraten

**Bruttoreproduktionsrate (BRR):** durchschnittliche Anzahl der Mädchengeburten im Laufe des Lebens von je 1000 Frauen eines Geburtenjahrgangs

Annahme: Diese 1000 Frauen unterliegen im Jahr  $i$  der altersspezifischen Fruchtbarkeitsrate  $g_i$  ohne Sterblichkeit.

**Nettoreproduktionsrate (NRR):** BRR - Sterblichkeit

Mit dieser Annahme wird aus einem Querschnitt (1000 Frauen zu einem best. Zeitpunkt) eine Längsschnitzaussage (1000 Frauen über einen Zeitverlauf) getroffen.

Tab. II.9: Ausgewählte Nettoreproduktionsraten der BRD

	Jahr	NRR
BRD:	1965	1,18
	1970	0,95
	1975	0,68
	1980	0,68
	1985	0,60
	1990	0,70
	1995	0,60
	2000	0,70
	2009	0,70

Ab 1990 gesamtes Bundesgebiet.

Quelle: Wolffs, M. (2002), S. 38, Tabelle 5, Bundesinstitut für Bevölkerungsforschung (2011)

Interpretation: Eine NRR=1 bedeutet die Konstanz des Bevölkerungsbestandes. Seit etwa 1970 reicht die Zahl der Geburten nicht mehr aus, den Bevölkerungsbestand zu erhalten.

## Sterbetafel

Mit den bisherigen Maßzahlen wird ein Vorgang nur summarisch charakterisiert: Mit der Tafelrechnung wird ein Ereignisablauf festgehalten.

Frage: Wie viele Personen einer Altersgruppe, die heute  $y$  Jahre alt sind, werden im Durchschnitt ein bestimmtes Alter erreichen bzw. vorher sterben?

Absterbeordnung, Lebenserwartung, die wichtig ist für:

- Bevölkerungsvorausschätzungen
- Rentenprobleme
- Versicherungsprobleme

### Sterbetafel

#### Generationstafel

zur Prognose: Annahme beobachtet  
Längsschnitt oder longitudinale  
Tafel; realer Geburtsjahrgang wird  
im Abgangsverhalten beschrieben  
(‘fertig’ in ca. 100 Jahren)  
(problematisch, da die Sterblichkei-  
ten sich verändern)

#### Querschnittstafel

Periodentafel; fiktive Kohorte, Sterb-  
lichkeit eines Querschnitts wird über-  
tragen; Periodentafeln sind vorherr-  
schend

Interpretation: für eine heute geborene  
Generation wird in 20 Jahren die  
gleiche Sterbewahrscheinlichkeit  
gelten, wie sie heute für die 20jährigen  
beobachtet wird.

Ausgangspunkt: Einjährige Sterbe-  
wahrscheinlichkeiten

Abb. II.11: Generations- und Querschnittstafel

### Sterbewahrscheinlichkeit (SW):

Ähnlich der Sterbeziffer (Gestorbene/mittlerer Bevölkerungsstand), aber:

SW = Quotient aus Gestorbenen und Lebenden desselben Geburtsjahrganges!

Altersspezifische Sterbeziffer	Einjährige Sterbewahrscheinlichkeit
$k_i = \frac{G}{B}$	$q_i = \frac{G^*}{B^*}$
G= Alle Personen, die im Verlauf des Kalenderjahres sterben und das $i$ -te, aber nicht das $(i+1)$ -te Lebensjahr vollendet haben.	G*= Gestorbene desselben Geburtsjahrgangs, die das Alter $i$ , aber nicht mehr das Alter $i+1$ erreicht haben.
B= durchschnittlicher Bestand an Lebenden, die während des Kalenderjahres das $i$ -te, aber nicht das $(i+1)$ -te Lebensjahr vollendet haben.	B*= durchschnittlicher Bestand an Lebenden eines Geburtsjahrgangs

Unter der Annahme, dass sich die Sterbefälle über das Jahr annähernd gleich verteilen, ergibt sich die folgende Beziehung zwischen der Sterbeziffer und der Sterbewahrscheinlichkeit:

$$q_i = \frac{2 \cdot k_i}{2 + k_i}$$

Sterbetafel (Absterbeordnung) (vgl. Abb. II.11)

Absterbeordnung aus den Sterbewahrscheinlichkeiten errechnet: fiktiver Anfangsbestand von 100.000

Anfangsbestand: 100.000  $l_0$  aus Spalte 2

Bestand an Überlebenden:  $l_i = l_{i-1} \cdot p_{i-1} = l_i \cdot (1 - q_{i-1})$

mit:  $l_i$  = Überlebende einer Altersgruppe  
 $q_i$  = Sterbewahrscheinlichkeit einer Altersgruppe (Spalte 4) -  
 (Komplementär zu  $p_i$ )  
 $p_i$  = Überlebenswahrscheinlichkeit einer Altersgruppe (Sp.5)

Tab. II.10: Auszug aus der allgemeinen Sterbetafel für Frauen 2007/09

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Vollendetes Alter	Überlebende im Alter i	Gestorbene im Alter i bis unter i+1	Sterbewahrscheinlichkeit vom Alter i bis i+1	Überlebenswahrscheinlichkeit vom Alter i bis i+1	Von den Überlebenden im Alter i bis zum Alter i+1 durchlebte Jahre	Von den Überlebenden im Alter i insgesamt noch zu lebende Jahre	Durchschnittl fernere Lebenserwartung in Jahren
i	$l_i$	$d_i$	$q_i$	$p_i = 1 - q_i$	$L_i$	$T_i$	$e_i = T/l$
0	100.000	320	0,00320	0,99680	99.727	8.252.546	82,53
1	99.680	28	0,00028	0,99972	99.666	8.152.819	81,79
2	99.652	16	0,00016	0,99984	99.644	8.053.153	80,81
3	99.636	14	0,00014	0,99986	99.628	7.953.509	79,83
4	99.621	12	0,00012	0,99988	99.615	7.853.881	78,84
5	99.609	9	0,00009	0,99991	99.605	7.754.265	77,85
6	99.600	8	0,00008	0,99992	99.596	7.654.660	76,85
7	99.592	8	0,00008	0,99992	99.588	7.555.064	75,86
8	99.584	6	0,00006	0,99994	99.581	7.455.476	74,87
9	99.578	6	0,00006	0,99994	99.575	7.355.895	73,87
10	99.572	8	0,00008	0,99992	99.568	7.256.320	72,88
11	99.564	8	0,00008	0,99992	99.560	7.156.753	71,88
12	99.556	8	0,00008	0,99992	99.552	7.057.193	70,89
13	99.548	10	0,00010	0,99990	99.543	6.957.641	69,89
14	99.538	12	0,00012	0,99988	99.532	6.858.098	68,90
15	99.526	13	0,00014	0,99986	99.519	6.758.566	67,91
16	99.512	16	0,00016	0,99984	99.504	6.659.047	66,92
17	99.496	17	0,00018	0,99982	99.488	6.559.542	65,93
18	99.479	22	0,00022	0,99978	99.468	6.460.055	64,94
19	99.457	21	0,00022	0,99978	99.447	6.360.587	63,95
20	99.436	21	0,00021	0,99979	99.425	6.261.140	62,97
21	99.415	21	0,00022	0,99978	99.404	6.161.715	61,98
22	99.393	23	0,00023	0,99977	99.382	6.062.311	60,99
23	99.371	23	0,00023	0,99977	99.359	5.962.929	60,01

Quelle: Statistisches Bundesamt, Wiesbaden 2010

Spalten (6)-(8) dienen zur Errechnung der durchschnittlichen Lebenserwartung nach den Sterblichkeitsverhältnissen von 2007/09.

Spalte (6):  $L_i =$  Die von den Überlebenden im Alter i bis zum Alter i+1 bereits durchlebten Jahre  
z.B. gemäß Spalte (2) erreichen  
 $l_{18} = 99.479$  Frauen das Alter 18 und  
 $l_{19} = 99.457$  Frauen das Alter 19

d.h. 22 der 100.000 Neugeborenen sterben im Alter von 18 Jahren. Sie verlebten nur einen Teil ihres Lebensjahres.

Annahme: Gleichmäßige Verteilung der Sterbefälle über das Jahr.

$$\rightarrow \text{vereinfachend: } L_{18} = \frac{l_{18} + l_{19}}{2} \quad \text{allg.: } L_i = \frac{l_i + l_{i+1}}{2}$$

Vereinfachung für  $L_0$ : Da höhere Geburtensterbefälle, ist Gleichverteilungsannahme nicht sinnvoll. Empirisch:  $L_0 = 0,83 \cdot (l_0 + l_1)$

Spalte (7):  $T_i =$  Die von den  $l_i$  Überlebenden insgesamt noch zu durchlebenden Jahre (= Summe aus den bereits durchlebten Jahre der jeweils höheren Altersstufen)

$$T_i = l_i + L_{i+1} + L_{i+2} + \dots = \sum_{j \geq i} L_j$$

Spalte (8):  $e_i =$  Mittlere fernere Lebenserwartung gibt für jedes Alter  $i$  die Zahl der Jahre an, welche die Überlebenden  $l_i$  von da an noch im Durchschnitt zu durchleben haben.

$$e_i = \frac{T_i}{l_i}$$

## 2.4 Bevölkerungsprognosen: Demografische Vorausschätzungen und computergestützte mikroanalytische Simulationsverfahren

Ziel der Bevölkerungsstatistik ist seit Alters her die Prognose der künftigen Bevölkerungsentwicklung.

Wichtig für:

- Investitionen im Sozial- und Bildungsbereich
- Energiepolitik
- Rentenversicherung
- etc.

### Modellrechnungen:

Eine empirische Ausgangssituation wird als Basis verwendet. Mit Hilfe bestimmter Annahmen/Hypothesen über die Sterblichkeit, Geburtenhäufigkeit und Wanderungen wird dann eine Modellrechnung durchgeführt:

- **Trendextrapolationen (vgl. Skriptum Kapitel VII)**

- **Modell des demografischen Übergangs**

Phase 1: hohe Geburten- und Sterblichkeitsziffern

Phase 2: sinkende Sterbeziffern bei hohen Geburtenziffern

Phase 3: Geburtenhäufigkeit sinkt beim Übergang vom Agrar- zum Industriestaat

Phase 4: Sterblichkeit stagniert, sinkende Geburtenziffern

Phase 5: Stagnierende Bevölkerung bei niedrigen Geburtenziffern und Sterbeziffern

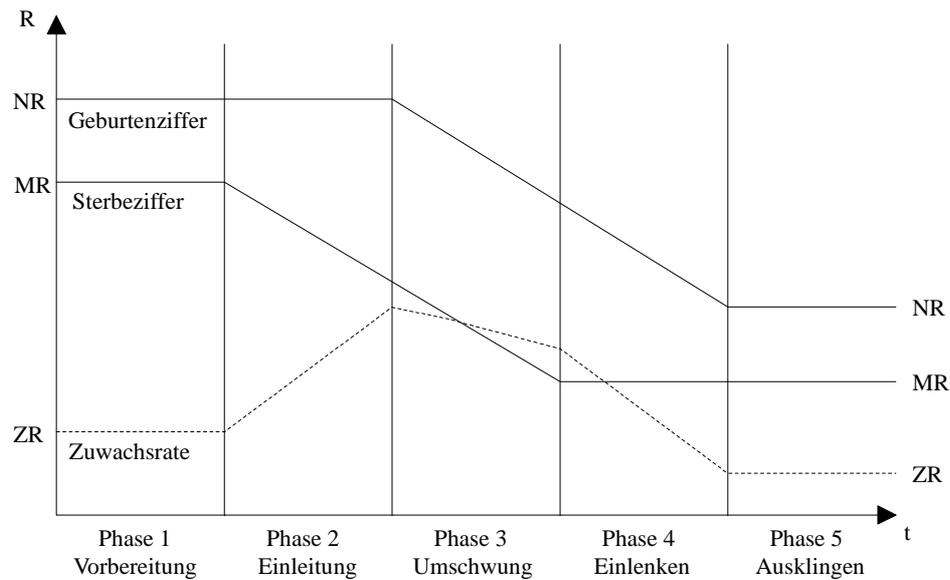


Abb. II.12: Schema des 'demografischen Übergangs'

Quelle: Grohmann (1986b), S. 75

### - Logistisches Modell

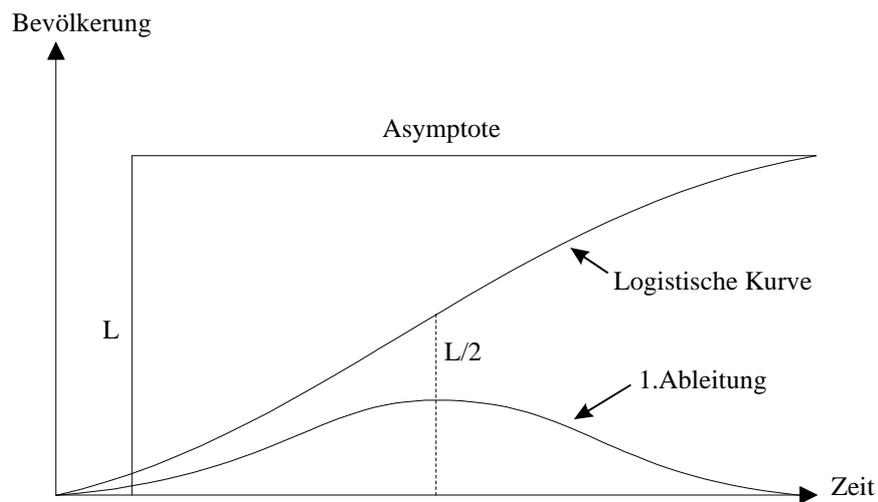


Abb. II.13: Logistisches Modell für eine Bevölkerungsprojektion:

Quelle: Esenwein-Rothe (1978), S. 65

### - Gruppensimulation

Bevölkerungsbestand wird gegliedert nach Alter und Geschlecht - gruppenweise - fortgeschrieben (Transition, Übergangsraten)

Jede Altersgruppe rückt dann Jahr für Jahr weiter (bereinigt um Sterbefälle, Wanderungen) (Grohmann 1986a).

### - Mikrosimulation

Die Merkmale jeder einzelnen Person werden im Haushalts- und Familienzusammenhang mittels empirischer Schätzansätze (Mikroökonomie) fortgeschrieben.

Dadurch ergeben sich vielfältige Auswertungsmöglichkeiten, wie Familienbildung etc.

Mikrosimulation (Überblick): Orcutt, Merz und Quinke (1986), Merz (1994c)

Zur statischen und dynamischen Mikrosimulation vgl. Abb. II.15.

'Mikroanalytische Grundlagen der Gesellschaftspolitik' der Universitäten Frankfurt und Mannheim

Dynamisches Sfb 3 - Mikrosimulationsmodell:

Steger (1980), Galler und Wagner (1986)

Statisches Sfb 3 - Mikrosimulationsmodell:

Merz (1994a)

Zum Ablauf des gesamten dynamischen Sfb 3-Mikrosimulationsmodells vgl. Abb. II.14

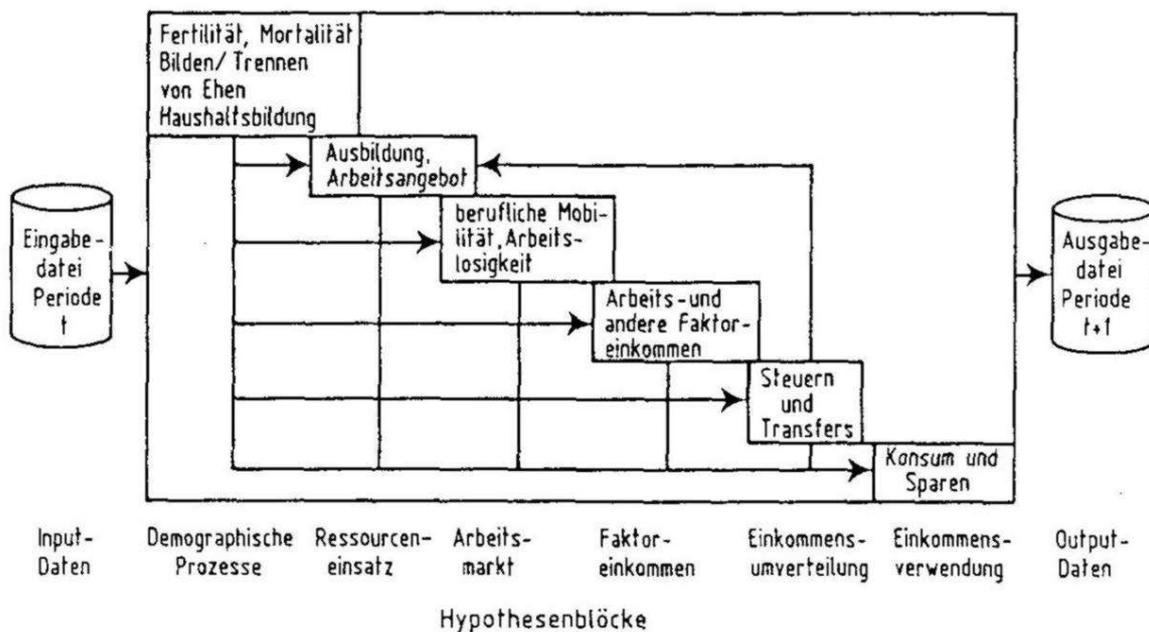


Abb. II.14: Die Grundstruktur des dynamischen Mikromodells

Quelle: Galler, Ott (1994), S. 406

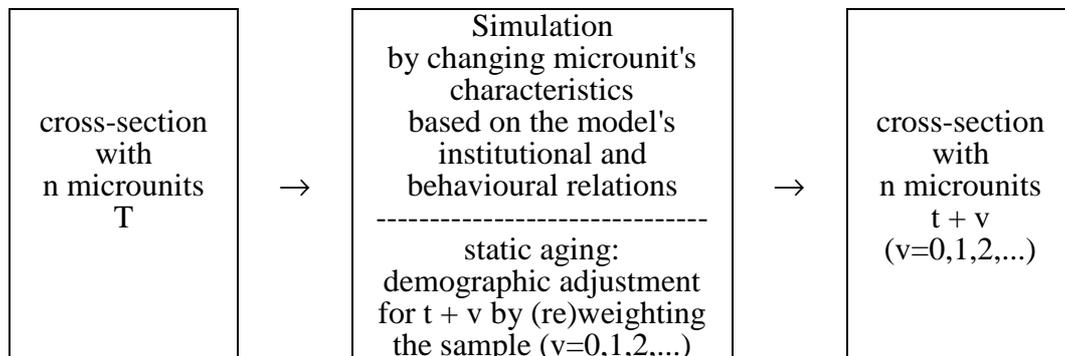
Die Einsatzgebiete der Sfb 3 - Mikrosimulationsmodelle sind:

- Rentenreform
- Steuerreformanalyse
- Leserverhalten
- Markt- und nichtmarktmäßige Aktivitäten privater Haushalte

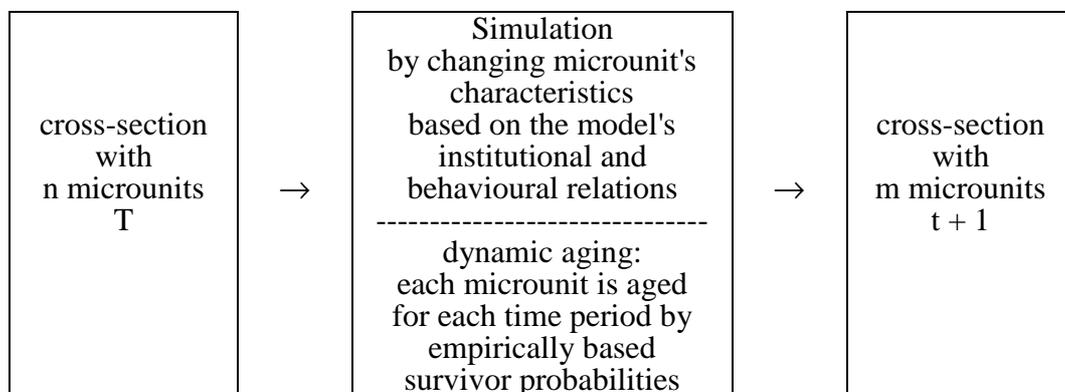
Mikrosimulation, ein aktuelles Beispiel: Verteilungswirkungen der Steuerreform 2000/2005 und Alternativen (Karlsruher Entwurf) (Merz, Stolze und Zwick (2002))

Weitere Informationen zur Bevölkerungsstatistik: Rinne (2001), Datenreport 2002, Teil I.1, Mueller (1993)

### STATIC MICROSIMULATION



### DYNAMIC CROSS-SECTION MICROSIMULATION



### DYNAMIC LIFE-CYCLE MICROSIMULATION

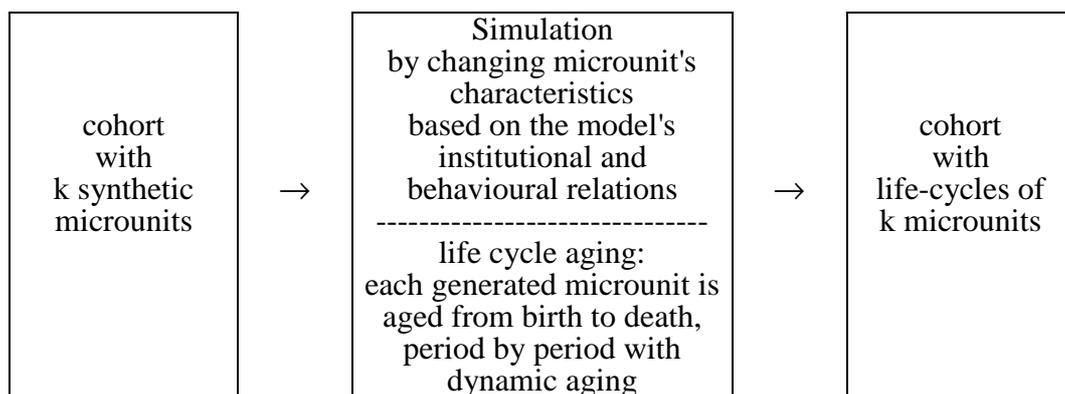


Abb. II.15: Statische und dynamische Querschnitts- und Lebenszyklus-Mikrosimulation

Quelle: Merz, J. (1994c)

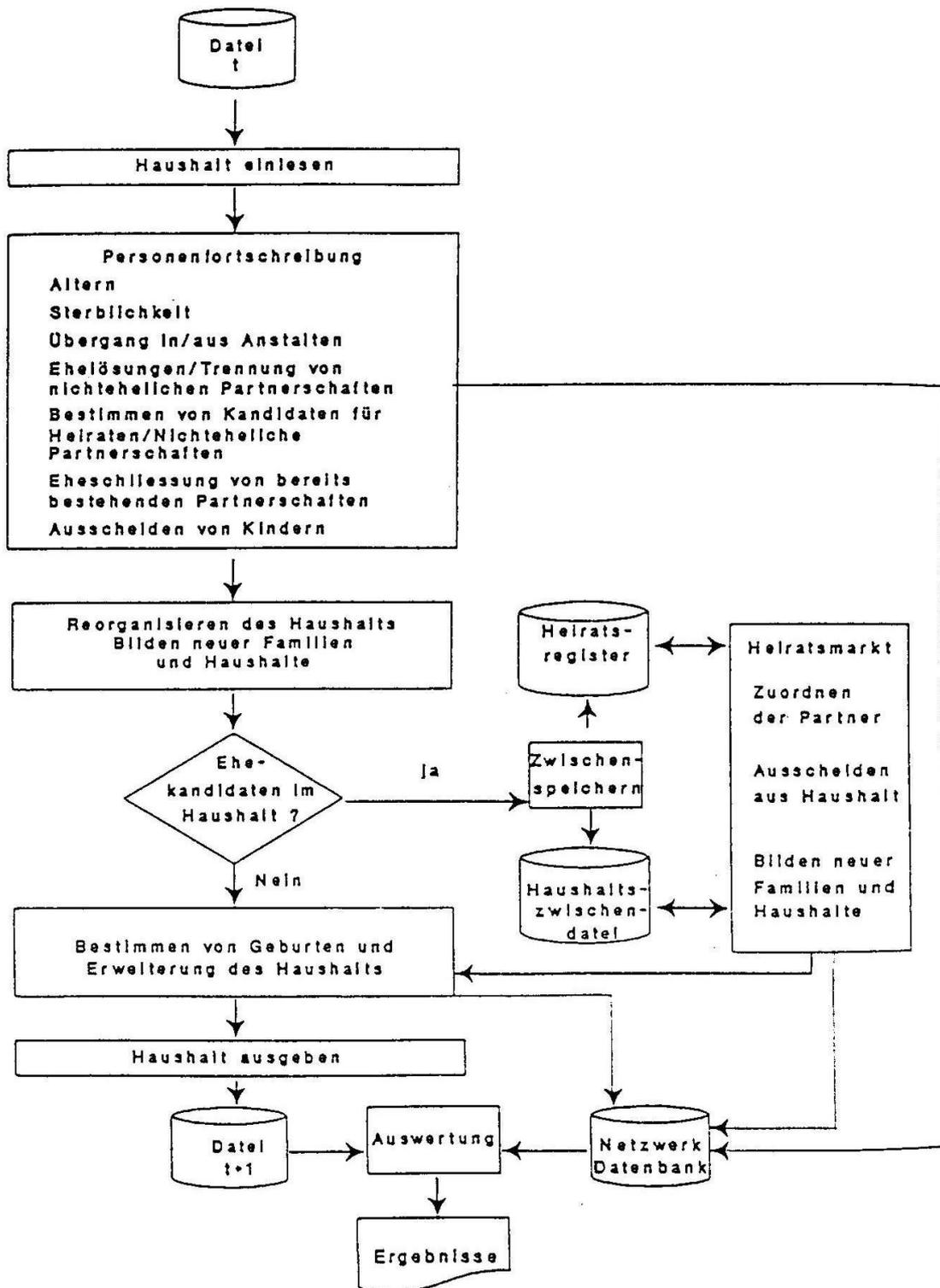


Abb. II.16: Demografie im 'Dynamischen Sfb 3 - Mikrosimulationsmodell'

Quelle: Galler, Ott (1994), S. 408

## 3 Produktion

### 3.1 Aufgabe, Abgrenzung und Messung

Produktionsstatistik = Statistik des warenproduzierenden Gewerbes

#### Aufgabe

- kurzfristige, regelmäßige Erfassung der industriellen Produktion
- Produktionsstatistik für
  - Indizes (Konjunkturindikatoren)
  - Volkswirtschaftliche Gesamtrechnung (VGR)
  - Wirtschaftsbeobachtung

#### Abgrenzung

Amtliche Statistik der Bundesrepublik Deutschland:

Produktionsstätten: Betriebe, Unternehmen, Arbeitsstätten → Systematik der Wirtschaftszweige

Das warenproduzierende Gewerbe (Energiewirtschaft und Bergbau, verarbeitendes Gewerbe (ohne Bergbau) und Baugewerbe) ist ein Wirtschaftszweig von zehn.

Tab. II.11: Wirtschaftszweige in der amtlichen Statistik

A	01.-03.	Land- und Forstwirtschaft, Fischerei	K	64.-66.	Erbringung von Finanz- und Versicherungsdienstleistungen
B	05.-09.	Bergbau	L	68.	Grundstücks- und Wohnungswesen
C	10.-33.	Verarbeitendes Gewerbe	M	69.-75.	Freiberufliche, wissenschaftliche und technische Dienstleistungen
D	35.	Energieversorgung	N	77.-82.	Sonstige wirtschaftliche Dienstleistungen
E	36.-39.	Wasserversorgung Abwasser- und Abfallentsorgung	O	84.	Öffentliche Verwaltung, Verteidigung; Sozialversicherung
F	41.-43.	Baugewerbe	P	85.	Erziehung und Unterricht
G	45.-47.	Handel; Instandhaltung und Reparatur von Fahrzeugen	Q	86.-88.	Gesundheits- und Sozialwesen
H	49.-53.	Verkehr und Lagerei	R	90.-93.	Kunst, Unterhaltung und Erholung
I	55.-56.	Gastgewerbe	S	94.-96.	Erbringung von sonstigen Dienstleistungen
J	58.-63.	Information, Kommunikation	T	97.-98.	Private Haushalte
			U	99.	Exterritoriale Organisationen und Körperschaften

Quelle: Statistisches Bundesamt, Wiesbaden (2008)

#### Definition und Messung der Produktion

Produktionsfaktoren  $\xrightarrow[\text{Produktion}]{\text{Erzeugung}}$  Sachgüter

(Industrie)

einschließlich der damit zusammenhängenden Dienstleistungen (Montage, Reparaturen)

Produktionsmessung: Erzeugnismenge, Preis

Ziel: Feststellen der echten Eigenleistung (Nettoleistung) der produzierenden Einheit →

**Nettoproduktionswert (NPW)** = Bruttoproduktionswert  $\cdot$  Vorleistungen  
 Vorleistungen: Materialverbrauch (Roh-, Hilfs- und Betriebsstoffe, Energie)

Der Nettoproduktionswert enthält aber immer noch die sog. sonstigen (anderen, personellen) Vorleistungen (vgl. Abb. II.17);

Sonstige Vorleistungen: Instandhaltungs-, Reparatur-, Werbe- und Vertriebskosten, Versicherungskosten, Versicherungsprämien, Bankspesen, Büromaterial, Porti, Lizenz- und Patentgebühren, Mieten und Pachten für Maschinen, EDV-Anlagen, Lagerräume Fahrzeuge etc. (vgl. v.d. Lippe (1990), S. 235ff.)

**Census Value Added** = NPW  $\cdot$  Kosten für industrielle oder handwerkliche Dienstleistungen Dritter

$$\text{Nettoquote} = \frac{\text{Nettoproduktionswert}}{\text{Bruttoproduktionswert}} (\text{NQ})$$

Je niedriger die Nettoquote ist, desto höher ist der Anteil der Vorleistungen am Bruttoproduktionswert (und desto höher ist die Endverbrauchsproduktion des Wirtschaftszweiges)

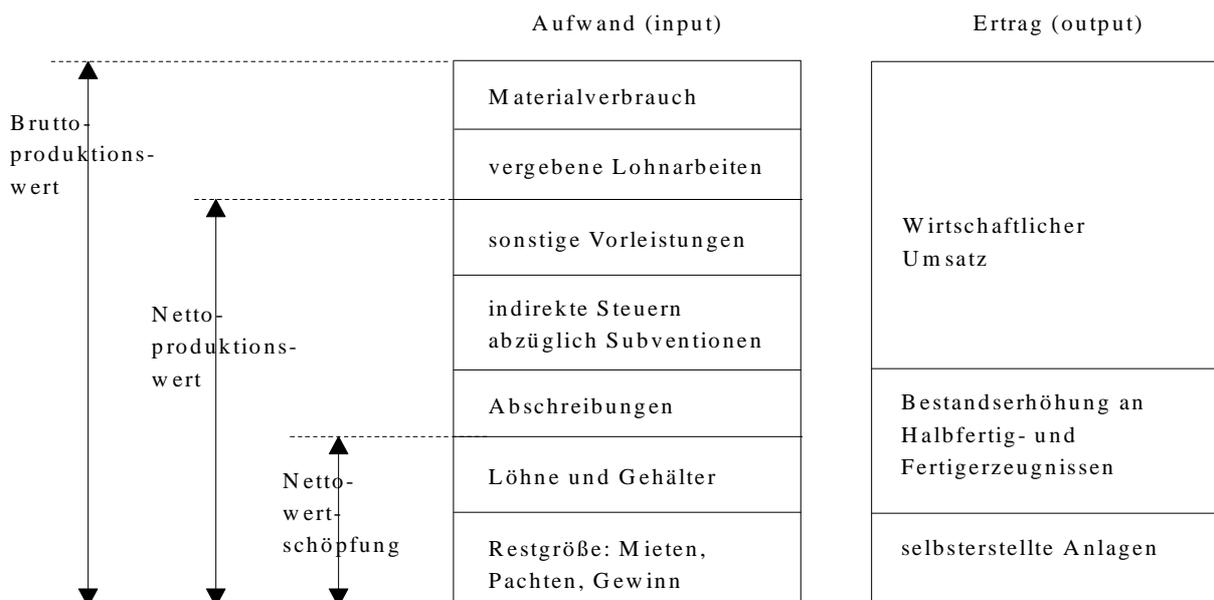


Abb. II.17: Produktionskonto einer Unternehmung

Quelle: Hujer (1990), S. 16

### 3.2 Das Erhebungssystem der Produktionsstatistik

Gesetz über die Statistik im Produzierenden Gewerbe (Bergbau, Verarbeitendes Gewerbe, Baugewerbe)

#### Erhebungsbereich

alle Wirtschaftseinheiten im Produzierenden Gewerbe (nach Neuordnung auch das Produzierende Handwerk)

## Erhebungseinheiten

- Unternehmen:
  - kleinste rechtlich selbständige Einheit in der Wirtschaftsstatistik (größtes Einteilungsmerkmal)
- Betriebe:
  - örtlich getrennte Niederlassungen des Unternehmens
  - Zuordnung von Unternehmen und Betrieben zu Wirtschaftszweigen nach Schwerpunkten der Produktion (Hauptbeteiligungskonzept). Zuordnungsprobleme→
- Fachliche Unternehmens- bzw. Betriebsteile
- Beteiligtenkonzept, so dass
  - der Produktionsausstoß möglichst homogen ist und
  - die erfaßten Produktionsfaktoren in enger Beziehung zum Produktionsausstoß stehen
- Arbeitsstätten:
  - örtliche Einheiten, in denen unter Einschluß des Leiters mindestens eine Person haupt- oder nebenberuflich ständig tätig ist (Arbeitsstättenzählung 1987 im Rahmen der Volkszählung)

**Berichtskreis:** berichtspflichtig bei

### 1. Unternehmenserhebungen

- alle Unternehmen mit mindestens zehn Beschäftigten, deren Schwerpunkt im Produzierenden Gewerbe liegt

### 2. Betriebserhebungen

- alle Betriebe des Produzierenden Gewerbes der unter 1 genannten Gewerbe
- Betriebe mit fachlichen Unternehmensteilen im Produzierenden Gewerbe mit mehr als zehn Beschäftigten, wobei der wirtschaftliche Schwerpunkt des Gesamtunternehmens außerhalb des Produzierenden Gewerbes liegt

## Wichtige Erhebungen:

### Monatserhebungen

liefern Indikatoren zur Konjunkturentwicklung (z.B. des Auftragseingangs)

auf Betriebsebene: Produktionsteilbericht

auf Unternehmensebene: Statistik über den Auftragsbestand

### Jahreserhebungen

liefern Indikatoren zur Konjunktur- und Strukturentwicklung

auf Betriebsebene:

- Investitionserhebung
- Statistik über den Auftragsbestand
- Erhebung für Kleinbetriebe

auf Unternehmensebene:

- Investitionserhebung
- Kostenstrukturerhebung (Stichprobe)

### Zensus im Produzierenden Gewerbe (alle vier bis sechs Jahre)

liefern Indikatoren zum Strukturwandel, regionale Sozialproduktionsberechnung  
Vordringliches Ziel: Ermittlung von Brutto- und Nettoproduktionswerten

Einen Überblick über das Berichtssystem im Produzierenden Gewerbe gibt Abb. II.18.

### Veröffentlichungen:

- Fachserie 4 (Produzierendes Gewerbe)
- 'Wirtschaft und Statistik'
- Sonderhefte

Erhebung	Energie- u. Wasser- versorgung	Bergbau u. Verarbeiten- des Gewerbe	Bauhaupt u. Ausbau- gewerbe
<b>Monatliche und vierteljährliche Erhebungen</b>			
<b>Monatsberichte</b>			
Beschäftigte, geleistete Arbeiterstunden, Lohn- u. Gehaltssumme			
Umsatz und Auftragseingang	X	X	X
Monatlicher Produktionseilbericht <sup>2</sup>	-	X	X <sup>1</sup>
Vierteljährliche Produktionserhebung <sup>2</sup>	-	X	-
Monatsstatistik über Bezug, Verbrauch, Erzeugung und Abgabe von Elektrizität <sup>4</sup>	-	X	X <sup>3</sup>
Vierteljahresstistik über Bestand und Verbrauch an Brennstoffen <sup>4</sup>	-	X	-
Fachstatistiken in bestimmten Produktionsbereichen	- X <sup>5</sup>	X	-
<b>Jährliche Erhebungen</b>			
<b>Jahreserhebungen für Unternehmen über</b>			
Beschäftigte, Lohn- u. Gehaltssummen und Umsatz	X <sup>6</sup>	- <sup>7</sup>	X
Investitionserhebung	X	X	X
Kostenstrukturerhebung	X	X	X
Erhebung für Kleinbetriebe <sup>8</sup>	-	X	-
Totalerhebung im bauhauptgewerbe und 'Zusatzerhebung' im Ausbaugewerbe	-	-	X
<b>Mehrfährliche Erhebungen</b>			
Zensus <sup>9</sup>	-	X	X
Statistik über den Material- und Wareneingang <sup>10</sup>	-	X	-

<sup>1</sup> Auftragseingang nur im Bauhauptgewerbe. Dort außerdem vierteljährliche Erhebung des Auftragsbestands.

<sup>2</sup> Produktion nach Güterarten (Menge und Wert).

<sup>3</sup> Produktionserhebung im Fertigteilbau.

<sup>4</sup> Im Rahmen des Monatsberichts erhoben.

<sup>5</sup> Dadurch erübrigen sich in der Energieversorgung die vier zuvor genannten.

- <sup>6</sup> Hier auch 'geleistete Arbeiterstunden'.  
<sup>7</sup> Zahlen werden durch die Zusammenfassung von Monatsergebnissen gewonnen.  
<sup>8</sup> Betriebe unterhalb der Abschneidegrenze von im allgemeinen 20 Beschäftigten, jedoch ohne handwerkliche Betriebe.  
<sup>9</sup> Für die Jahre 1967, 1979, 1985 und danach alle sechs Jahre.  
<sup>10</sup> Für die Jahre 1967, 1978 und danach alle vier Jahre (im Bauhauptgewerbe nur für 1978 und 1982).

Abb. II.18: Übersicht über das Berichtssystem im Produzierenden Gewerbe  
 Quelle: Kunz (1987), S. 177

### 3.3 Indizes der industriellen Produktion und Produktionsmessung

#### Indizes der Nettoproduktion

Verwendung als Konjunkturindikator (Auftragseingang → Produktion → Beschäftigung)

Gesamtindex:

Wirtschaftszweigindizes mit jeweiligem Anteil am gesamten Nettoproduktionswert gewichtet  
 Neuberechnung der Gewichte: 3-8 Jahre

Indizes der Nettoproduktion für das Verarbeitende Gewerbe von 2006-2011 finden sich in Abb. II.19.

#### Indizes der Produktion Verarbeitende Gewerbe

Produktionsindex (2005 = 100) ausgewählte Wirtschaftszweige	Jahre					
	2006	2007	2008	2009	2010	2011
Kohlenbergbau	101	98	94	87	87	84
Gewinnung von Erdöl und Erdgas	111	107	76	66	64	60
Herstellung von Nahrungs- und Futtermitteln	95	99	104	100	99	99
Getränkeherstellung	84	88	94	81	75	79
Herstellung von Textilien	97	101	100	71	75	83
Herstellung von Bekleidung	100	88	83	59	61	60
Herstellung von Leder, Lederwaren und Schuhen	95	94	110	94	105	112
Herstellung von Papier, Pappe und Waren daraus	103	112	117	99	103	109
Kokerei und Mineralölverarbeitung	98	101	100	89	87	91
Herstellung von Gummi- und Kunststoffwaren	99	110	111	85	92	105
Metallerzeugung und -bearbeitung	102	115	115	77	84	101
Herstellung von Metallerzeugnissen	97	108	115	83	83	102
Herstellung von elektrischen Ausrüstungen	101	111	116	86	86	104
Maschinenbau	87	102	112	84	73	90
Herstellung von Kraftwagen und -teilen	94	104	109	66	77	99
Herstellung von Möbeln	97	105	106	87	79	84
Energieversorgung	125	104	107	109	113	109
Verarbeitendes Gewerbe	99	103	104	84	85	92

Abb. II.19: Index der Nettoproduktion für das Verarbeitende Gewerbe  
 Quelle: Statistisches Bundesamt, Wiesbaden (2011)

Es gibt Statistiken mit

- 356 Mengenreihen des Produktionsausstoßes
- 664 preisbereinigte Wertereihen
- 40 preisbereinigte Umsatzreihen
- 9 Arbeiterstundenreihen

Einschränkungen der Aussagekraft, da diese nur als Hilfsreihen für den idealtypischen Begriff 'Konjunktursituation' verwendet werden.

### **Index der Bruttonproduktion für Investitions- und Verbrauchsgüter**

nach Warengruppen gegliedert; wichtig für Input/Output-Analysen

Zuordnungsprobleme:

- Verwendungszweck Haushalte → Verbrauchsgut
- Unternehmen/Staat → Investitionsgut

### **Indizes des Auftragseingangs und des Auftragsbestands**

Ist ein Konjunktur-Frühindikator; wird nach Wirtschaftszweigen berechnet.

#### **Auftragseingang**

- = Wert der im abgelaufenen Kalendermonat eingegangenen und angenommenen Bestellungen
- (soll weitgehende Übereinstimmung zwischen Auftragseingängen und späteren Umsätzen ermöglichen)
- Gemessen: bei Anbieter (Bestellungen bei inländischen Unternehmen)

Monatlich, zu jeweiligen Preisen und zu konstanten Preisen eines Basisjahres

#### **Auftragsbestandsindex (seit 1970)**

#### **Auftragsbestand:**

- = Summe aller vorliegenden, akzeptierten, aber noch nicht ausgeführten Bestellungen bewertet gegen Ende des Berichtsmonats

Auftragseingang und -bestand können gänzlich unterschiedliche Bedeutung haben (z.B. bei Gütern mit langen Produktionsprozessen). Deshalb ist die Kausalkette 'Auftragseingang (-bestand) → Produktion → Beschäftigung' kritisch zu hinterfragen.

### 3.4 Produktivitätsmessung

$$\text{Produktivität} = \frac{\text{Produktionsergebnis (Output)}}{\text{Faktoreinsatz (Input)}}$$

z.B. Arbeitsproduktivität (Faktor Arbeit)

Statistisches Bundesamt berechnet vier Indizes der Arbeitsproduktivität (vgl. Tab. II.12)

Output: industrielle Nettoproduktion

Input: Arbeitseinsatz durch

- Zahl der Arbeiter
- Zahl der Beschäftigten
- Zahl der Arbeiterstunden
- Zahl der Beschäftigtenstunden

Probleme: mehrere Faktoren bestimmen gemeinsam den Output (unterschiedliche Kapital-Arbeit-Ausstattung)  
unterschiedliche Qualität, Ausbildungsstand des Produktionsfaktors 'Arbeit'

Produktivitätssteigerungen sind differenziert zu bewerten (z.B. bei Rationalisierungsinvestitionen mit → Freisetzung von Arbeitskräften → Einkommen ↓ → Konsumgüternachfrage ↓ ...)

Tab. II.12: Index der Arbeitsproduktivität für den Bergbau und das Verarbeitende Gewerbe (2005 = 100)

	2005	2006	2007	2008	2009
Produktionsergebnis je Beschäftigten	100	109	112	110	94
je Arbeitsstunde	100	108	110	109	100

Quelle: Statistisches Bundesamt, Wiesbaden (2011)

Weitere zusammenfassende Hinweise zur Produktionsstatistik: Rinne (1994), S. 227-260, Kunz (1987), S. 173-197.

## 4 Dienstleistung

### 4.1 Allgemeine Bedeutung

Der Dienstleistungsbereich (tertiärer Sektor) produziert im Gegensatz zum primären (Land- und Forstwirtschaft) und sekundären (Produzierendes Gewerbe) keine materiellen Güter.

Leistungen des Handels, Gastgewerbes, der Banken und Versicherungen, Verkehrsgewerbes und der Freien Berufe

Fourastié: Der Tertiäre Sektor ist die große Hoffnung des 20. Jahrhunderts für die Wirtschaftsstruktur- und Beschäftigungsentwicklung

Tab. II.13: Beschäftigte im tertiären Sektor

Jahr	Beschäftigte im tertiären Sektor
1900	jeder vierte
1950	jeder dritte
1990	mehr als jeder zweite Erwerbstätige
2001	Mehr als zwei Drittel (67 %)
2005	ca. 72 %
2010	ca. 74%

Quelle: Datenreport 1992, S. 329, Datenreport 2002, S. 303, Datenreport 2006, S. 291, Statistische Ämter des Bundes und der Länder (2011)

### 4.2 Handel, Gastgewerbe, Kreditinstitute

#### Handel (Vertrieb von Waren)

Mittelständische Struktur: mehr als die Hälfte aller Handelsunternehmen beschäftigte 1987 mehr als zwei Personen.

Statistiken: Handels- und Gaststättenzählungen, Arbeitsstättenzählung 1987

Handelsbereich 1987: ca. 50.000 Unternehmen, 3,9 Mio. Beschäftigte (jeder siebte Erwerbstätige)

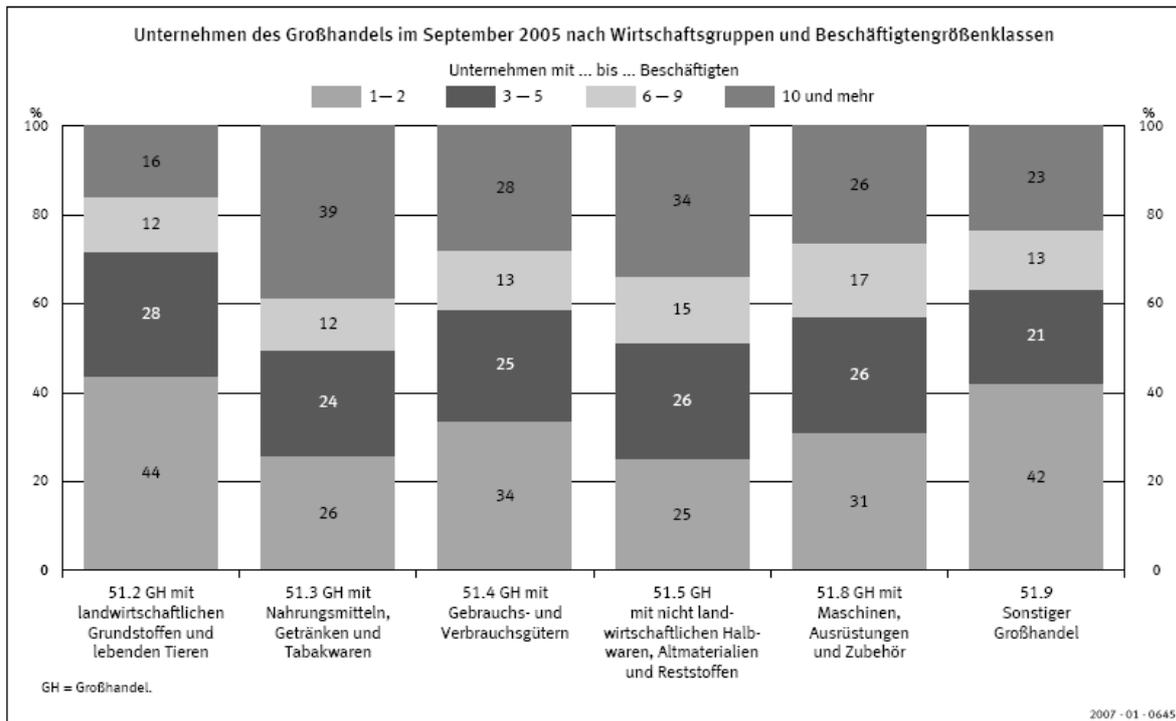


Abb. II.20: Unternehmen des Großhandels im September 2005  
Quelle: Datenreport 2008

### Gastgewerbe

Größte Gruppe: Gaststättengewerbe (vgl. Abb. II.21)

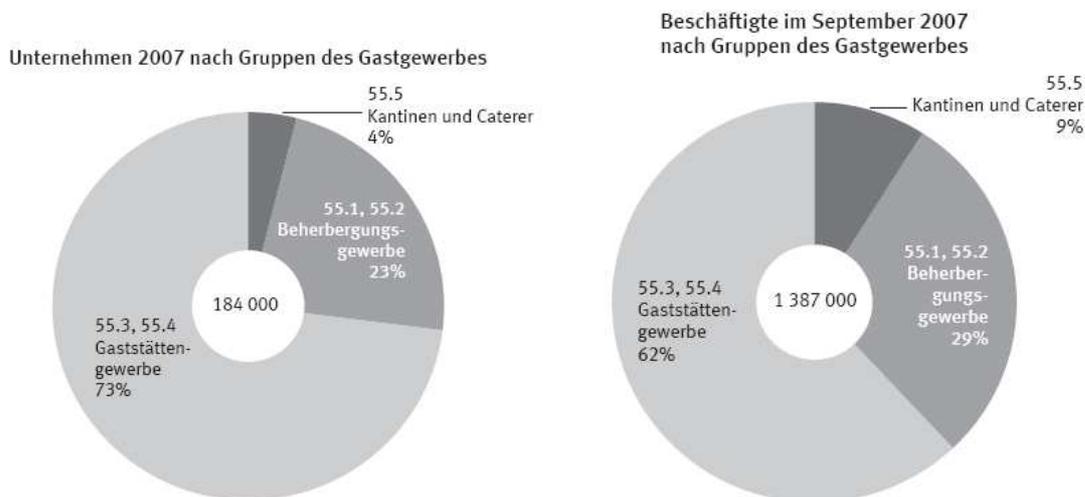


Abb. II.21: Unternehmen und Beschäftigte des Gastgewerbes 2007  
Quelle: Datenreport 2008

## Kreditinstitute

Durch die Zunahme von Bankvorgängen per Online-Banking ist zwischen den Jahren 2003 und 2009 die Anzahl der Bankstellen in Deutschland von fast 50.000 auf etwa 41.600 gesunken.

Die Anzahl der Beschäftigten in der Bankenbranche hat in den letzten Jahren stark abgenommen. Waren im Jahr 2000 noch rund 775.000 Personen in dem Sektor tätig, reduzierte sich die Anzahl der Erwerbstätigen bis zum Jahr 2009 auf knapp über 660.000.

Zu den Schwerpunkten der geschäftlichen Tätigkeit vgl. Abb. II.22

### Aktiva und Passiva

Bilanzergebnisse der Banken in Mill. EUR	2005	2006	2007	2008	2009
<b>Aktiva</b>					
<b>Bilanzsumme</b>	6 903 169	7 187 714	7 625 737	7 956 390	7 509 829
<b>Barreserven</b>	63 262	65 895	82 819	120 375	96 193
<b>Wechselbestand</b>	2 966	2 502	2 285	1 635	1 049
<b>Kredite an in- und ausländische Banken</b>	1 956 772	2 089 342	2 328 112	2 429 520	2 124 062
<b>Kredite an Nichtbanken</b>	3 020 110	3 050 739	3 140 365	3 227 428	3 161 182
<b>Schuldverschreibungen und andere festverzinsliche Wertpapiere</b>	1 209 293	1 279 116	1 365 192	1 438 264	1 446 137
<b>Sonstige Aktiva</b>	650 766	700 120	706 964	739 168	681 206
<b>Passiva</b>					
<b>Bilanzsumme</b>	6 903 169	7 187 714	7 625 737	7 956 390	7 509 829
<b>Einlagen von inländischen und ausländischen Banken</b>	1 943 192	2 031 262	2 211 146	2 278 595	2 001 236
<b>Einlagen von inländischen und ausländischen Nichtbanken</b>	2 554 049	2 663 896	2 834 760	3 017 674	2 994 150
<b>dar.: Sicht- und Termineinlagen</b>	1 862 734	1 979 908	2 159 833	2 344 559	2 295 713
<b>Spareinlagen</b>	611 877	594 858	563 823	544 121	604 111
<b>Verbriefte Verbindlichkeiten insgesamt</b>	1 616 781	1 638 400	1 659 812	1 640 653	1 530 347
<b>Eigenkapital insgesamt (gezeichnetes Kapital, Rücklagen, abzüglich ausgewiesener Verlust)</b>	260 168	287 845	306 426	331 557	342 794
<b>Sonstige Passiva</b>	528 979	566 311	613 593	687 911	641 365

Quelle: Deutsche Bundesbank.

Abb. II.22: Aktiva und Passiva der Kreditinstitute  
Quelle: Bundesbank (2011)

## 4.3 Verkehr, Freie Berufe

### Verkehr

Verknüpft Produzenten und Märkte, Beschäftigte und Arbeitsstellen.

Statistische Informationen bspw. zur Verkehrsinfrastruktur (Straßennetz, Schienennetz, Telefon- und Telefaxanschlüsse etc.), Fahrzeugbestände, Verkehrsleistungen (Personen- und Güterverkehr), Energieverbrauch oder Verkehrsunfälle.

Fahrzeugbestand: Deutschland: 4. größter Kfz-Bestand, aber nur ca. 5 % des Weltweiten Bestands (2009), (zur Aufteilung in Deutschland im Jahr 2011 vgl. Abb. II.23)

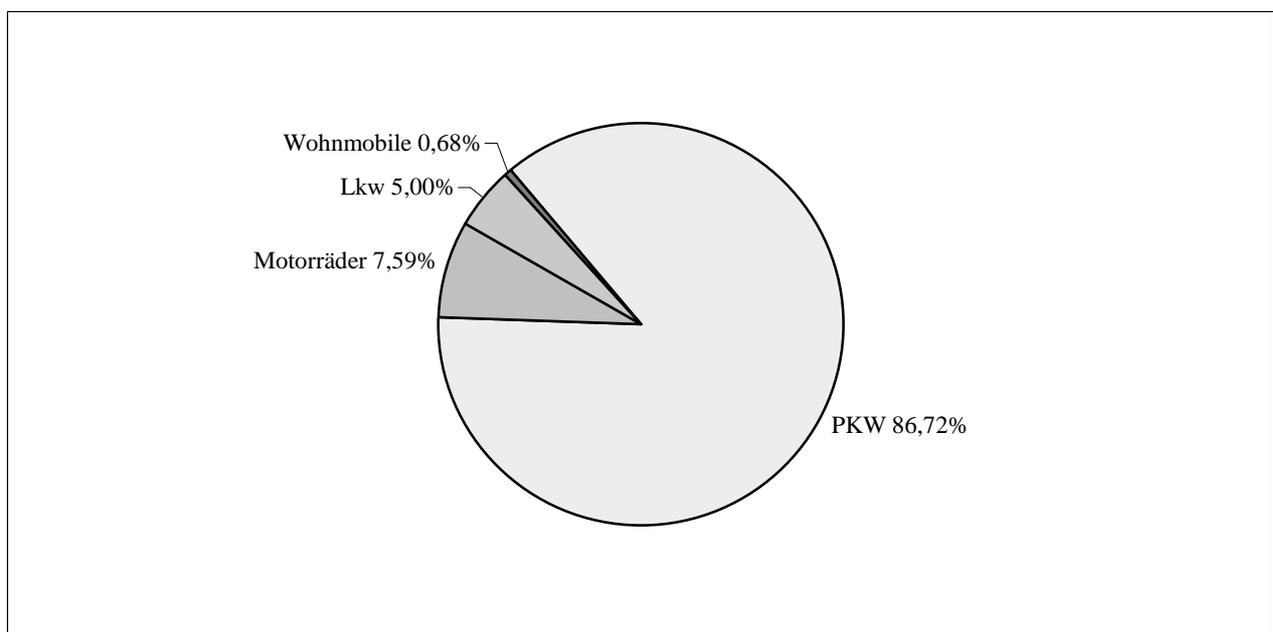


Abb. II.23: Fahrzeugbestand in Deutschland 01.01.2011

Quelle: Statistisches Bundesamt, Wiesbaden (2011)

### Freie Berufe

Alle Personen, die selbständig in einem wissenschaftlichen, künstlerischen, ärztlichen, beratenden (Katalogberufe nach EStG § 18) oder ähnlichen Beruf (nichtselbständige Gewerbetreibende) arbeiten.

z.B.: Ärzte, Rechtsanwälte, Wirtschaftsprüfer, Steuerberater, Architekten bis Künstler

Nur unsystematische und lückenhaft Statistiken gegeben.

Informationen:

- Forschungsinstitut Freie Berufe (FFB) der Universität Lüneburg
- Institut für Freie Berufe an der Universität Erlangen-Nürnberg
- Kammern, Verbände, Organisationen

Während die Selbständigenanteile an den Erwerbstätigen seit 1950 stark zurückgehen (von ca. 16 % auf 8 %), nehmen die Freiberufleranteile stark zu (vgl. Abb. II.24a,b).

Freiberuflich Tätige 2000: ca. 873.000

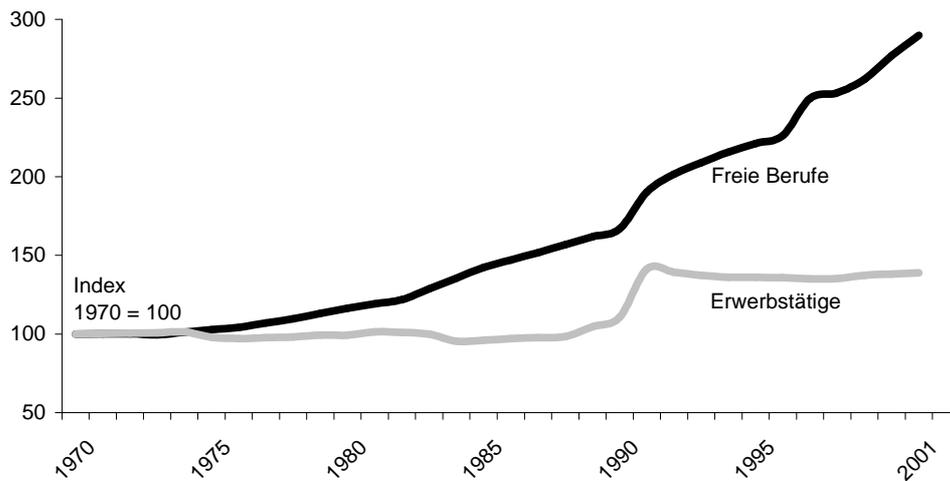
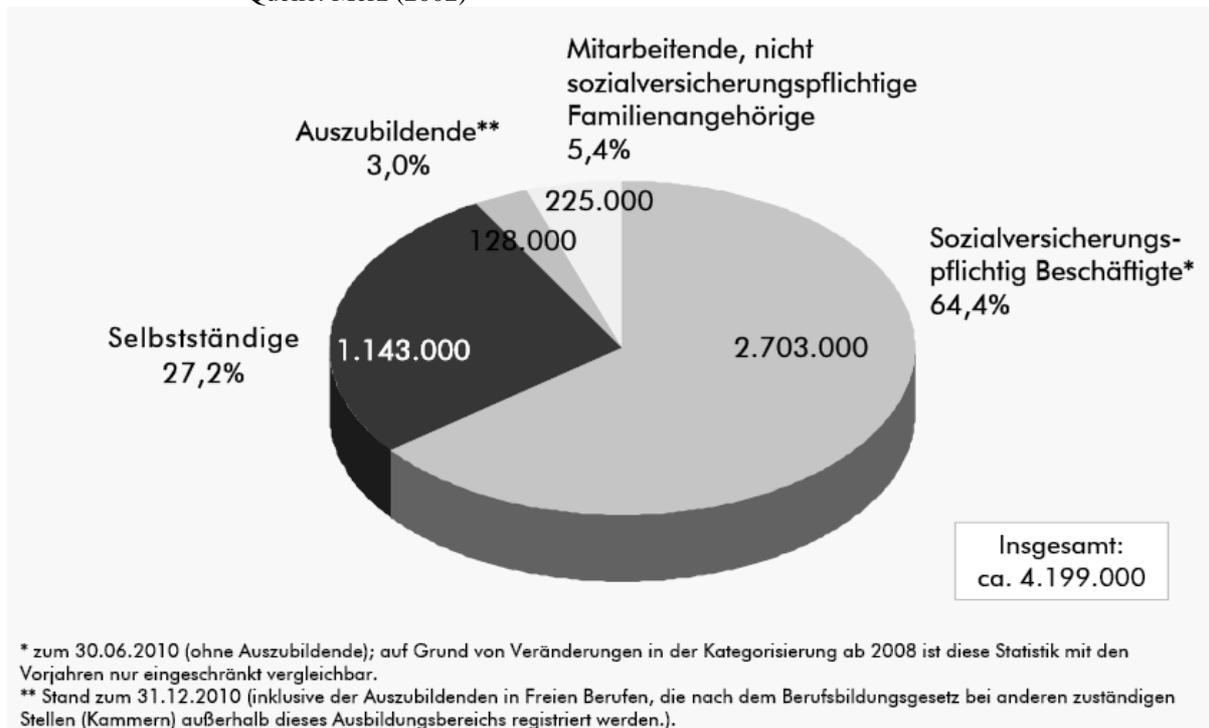


Abb II.24a: Entwicklung der Freien Berufe und der Selbstständigen 1970-2001  
Quelle: Merz (2002)



\* zum 30.06.2010 (ohne Auszubildende); auf Grund von Veränderungen in der Kategorisierung ab 2008 ist diese Statistik mit den Vorjahren nur eingeschränkt vergleichbar.

\*\* Stand zum 31.12.2010 (inklusive der Auszubildenden in Freien Berufen, die nach dem Berufsbildungsgesetz bei anderen zuständigen Stellen (Kammern) außerhalb dieses Ausbildungsbereichs registriert werden.).

Abb. II.24b: Erwerbstätige in Freien Berufen in Deutschland (Stand: 01.01.2011; z.T. vorläufige Ergebnisse)

Quelle: Institut für Freie Berufe Nürnberg (2011)

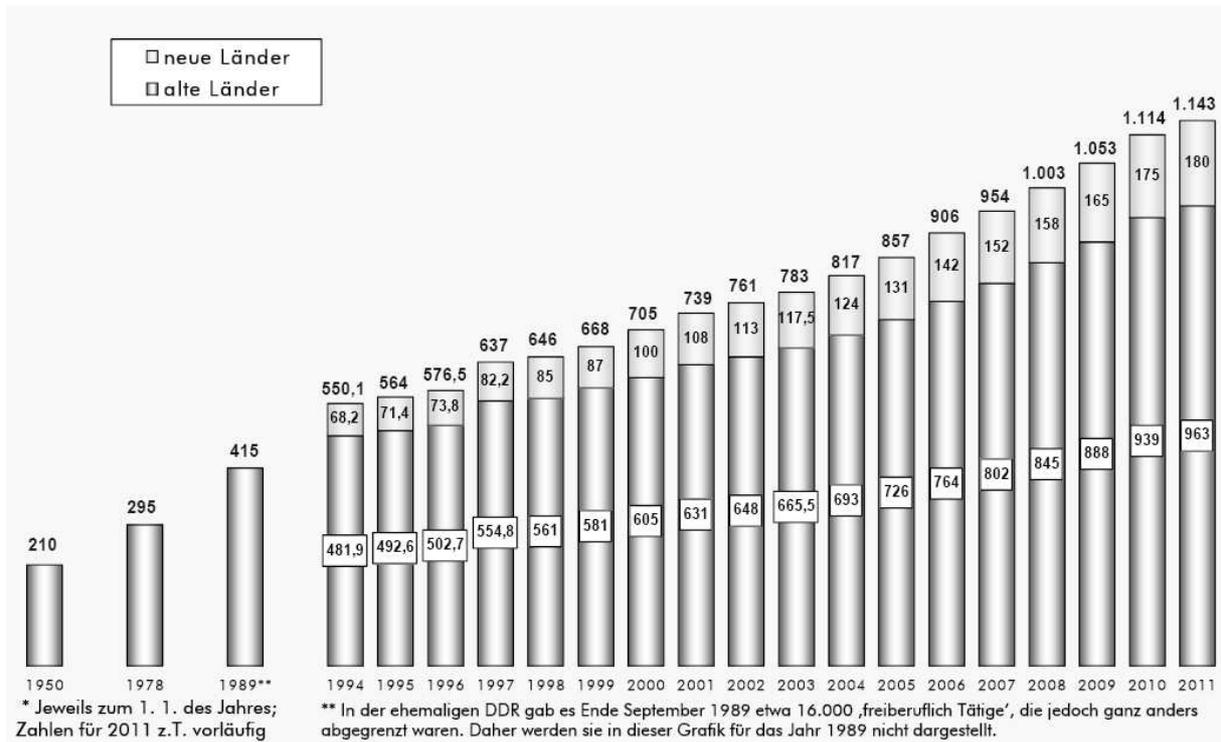


Abb. II.24c: Entwicklung der Freien Berufe und den Selbstständigen 1950-2011\* (in Tsd.)  
Quelle: Institut für Freie Berufe Nürnberg (2011)

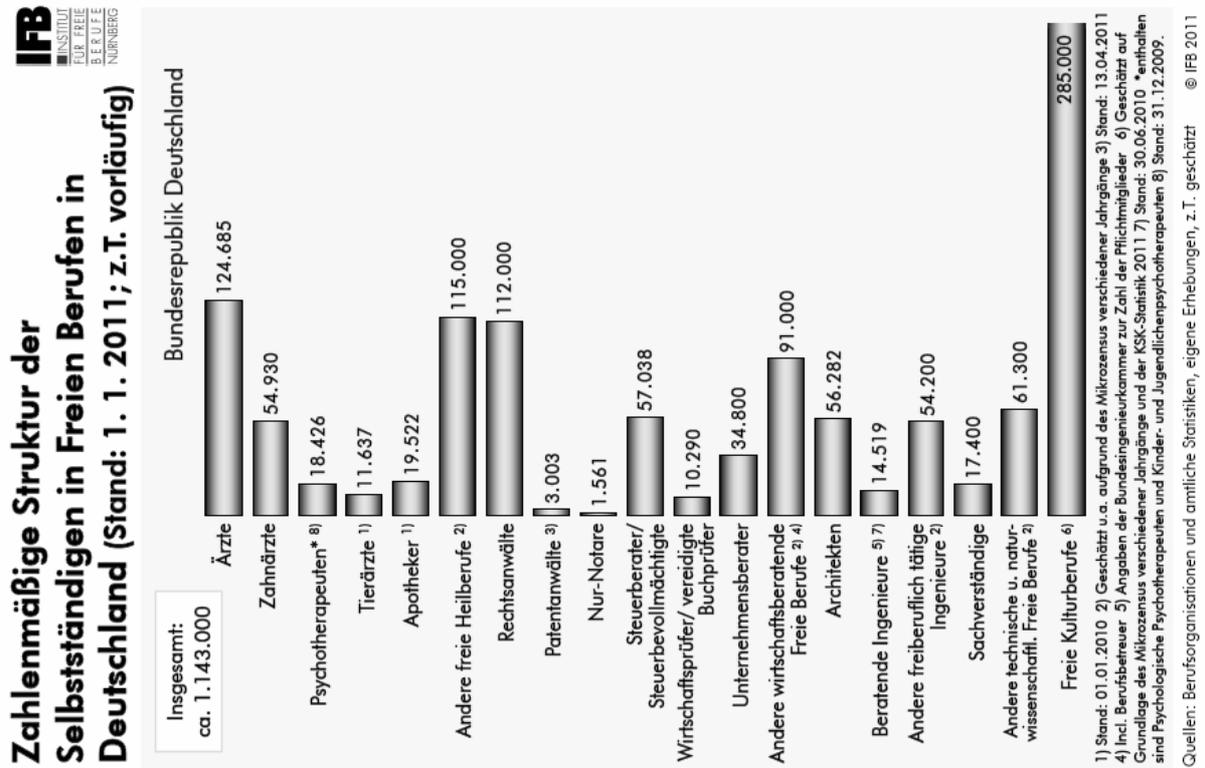


Abb. II.25: Selbstständige in Freien Berufen im Januar 2011<sup>1</sup>  
Quelle: Institut für Freie Berufe Nürnberg (2011)

## 5 Erwerbstätigkeit

Die Erwerbstätigkeit ist zentral für die Gesellschafts-, Sozial- und Wirtschaftspolitik zur Sicherung des Lebensunterhalts und der persönlichen Selbstentfaltung und bildet eine Nahtstelle zwischen der Bevölkerungsstatistik (Angebot an Arbeit[skraft]) und der Produktions- und Dienstleistungsstatistik (Nachfrage nach Arbeit[skraft]).

### Vielfältige Anforderungen

- Bevölkerungs- und erwerbsstatistische Aspekte:  
laufende Beobachtung der Erwerbsbeteiligung und ihre Veränderung im Zeitablauf
- Sozialpolitischer Aspekt:  
Anzahl der Arbeitslosen, Höhe der Stillen Reserve
- Wirtschaftspolitische Aspekte:  
Beschäftigung im Unternehmens- und Betriebszusammenhang → Aufschlüsse über Quantität, Qualität und Struktur des Produktionsfaktors 'Arbeit'; Regionale Aspekte

Arbeitsmarkt, Erwerbstätigkeit sind zentral für die individuelle Lebensgestaltung

### Veränderung der Arbeitslandschaft ('Zukunft der Arbeit', Prognose IAB)

Verschiebung in individueller Zeitverwendung [Zeitbudgeterhebung 1991/1992 des Statistischen Bundesamtes] ('Zeitverwendung in Erwerbstätigkeit und Haushaltsproduktion - Dynamische Mikroanalysen mit Paneldaten': DFG-Projekt Merz ab 1992)

### 5.1 Wichtige arbeitsmarktstatistische Erhebungen

Unterscheidung der Statistiken nach dem

- **Wohnortprinzip** (Erwerbstätige)  
Personen werden an ihrem Wohnort über ihre Erwerbstätigkeit befragt.
- **Arbeitsortprinzip** (Beschäftigte)  
Betrieb, Arbeitsplatz

Damit wird zwischen Erwerbstätigen und Beschäftigten unterschieden.

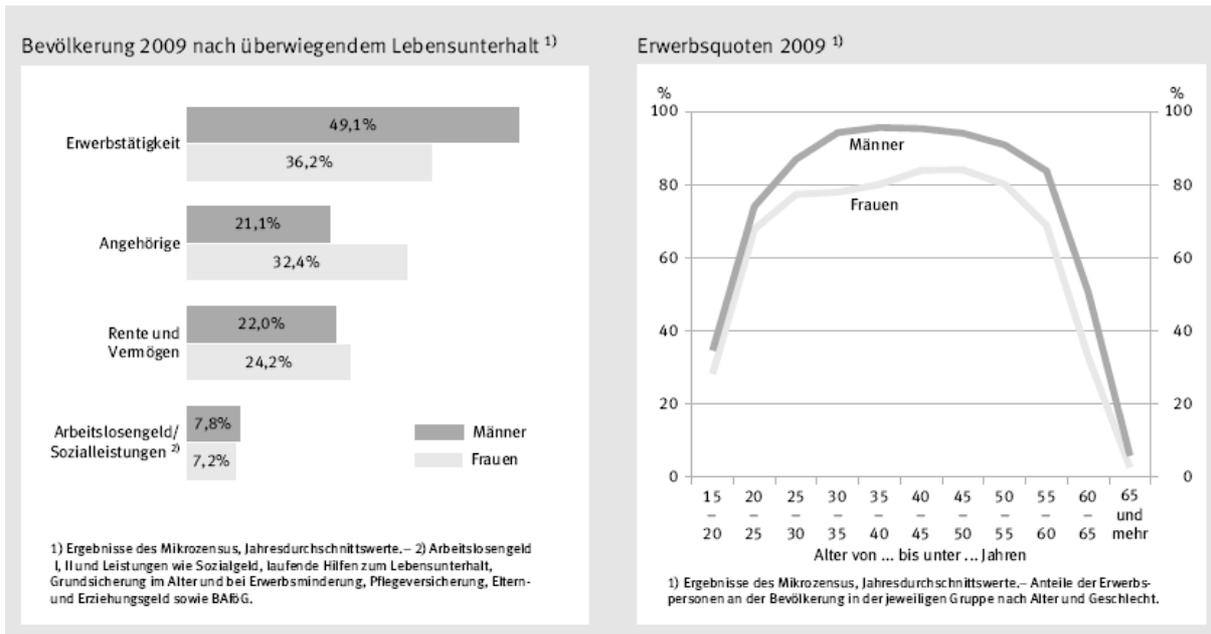
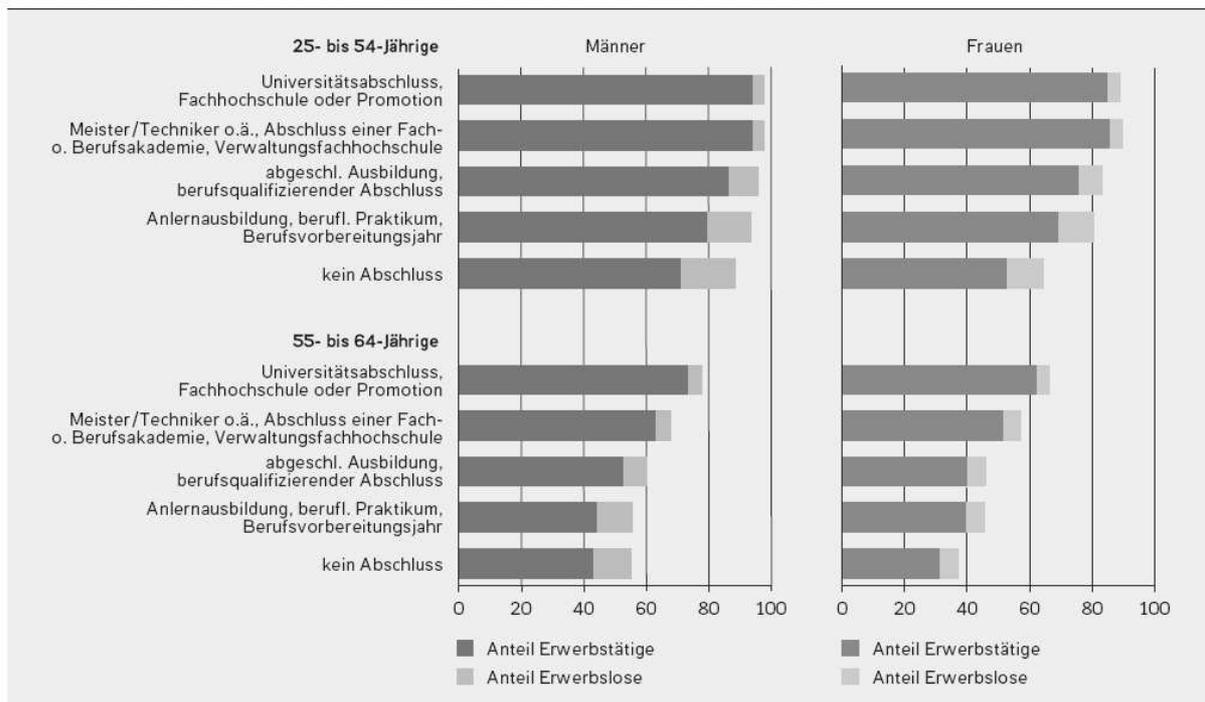


Abb. II.26a: Bevölkerung nach überwiegendem Lebensunterhalt und Erwerbsquoten 2009  
Quelle: Statistisches Jahrbuch 2008



Ergebnisse des Mikrozensus. Bevölkerung ab 15 Jahren und nicht mehr in Bildung oder Ausbildung.

Abb. II.26b: Erwerbsstatus nach dem höchsten Abschluss 2006, in %  
Quelle: Datenreport 2008, S. 114

Berichtsmonat: Juli 2008  
Deutschland

Merkmal	2008				Veränderung gegenüber Vorjahresmonat (Arbeitslosenquote Vorjahreswerte)			
	July	Juni	Mai	April	Juli		Juni	Mai
	absolut		in %		in %		in %	
	absolut	in %	absolut	in %	absolut	in %	absolut	in %
ERWERBSTÄTIGE (Monatsdurchschnitt; Inland) <sup>1)</sup>	...	40.321.000	40.233.000	40.055.000	...	...	1,4	1,4
SOZ.-VERS.PFL. BESCHÄFTIGTE (Bestand, geschätzt)	...	...	27.423.100	27.326.000	...	...	...	2,2
ARBEITSLÖSE								
- Bestand	3.210.045	3.159.811	3.283.279	3.413.921	-505.383	-13,6	-14,3	-13,9
dar.: 50,0% Frauen	1.606.571	1.564.794	1.611.365	1.659.502	-272.861	-14,5	-15,3	-14,9
50,0% Männer	1.603.463	1.595.005	1.671.907	1.754.415	-232.497	-12,7	-13,4	-12,8
11,3% Jüngere unter 25 Jahren	362.665	304.022	315.848	338.929	-63.551	-14,9	-17,0	-16,1
2,4% dar.: Jugendliche unter 20 Jahren	76.746	55.615	57.075	61.065	-14.245	-15,7	-20,2	-19,4
26,2% 50 Jahre und älter	842.087	845.088	872.202	895.345	-123.613	-12,8	-14,1	-14,3
13,2% dar.: 55 Jahre und älter	423.674	419.590	427.225	432.275	-41.582	-8,9	-12,0	-13,6
15,2% Ausländer	487.661	486.988	500.855	514.396	-60.491	-11,0	-11,8	-11,7
ARBEITSLÖSENQUOTEN bezogen auf								
- alle zivilen Erwerbspersonen insgesamt	7,7	7,5	7,8	8,1	8,9	-	8,8	9,1
- abhängige zivile Erwerbspersonen insgesamt	8,6	8,5	8,8	9,1	9,9	-	9,9	10,2
Männer	8,3	8,2	8,6	9,0	9,5	-	9,5	9,9
Frauen	8,9	8,7	9,0	9,2	10,4	-	10,2	10,5
Jüngere unter 25 Jahren	7,6	6,4	6,6	7,1	8,9	-	7,7	7,9
dar.: Jugendliche unter 20 Jahren	4,9	3,6	3,7	3,9	5,8	-	4,5	4,5
Ausländer	17,9	17,8	18,3	18,6	19,8	-	20,0	20,5
LEISTUNGSEMPFÄNGER <sup>2)</sup>								
- Arbeitslosengeld	831.561	824.164	876.092	942.555	-184.316	-18,1	-17,5	-16,8
- erwerbsfähige Hilfebedürftige (Alg II)	5.040.136	5.058.794	5.126.871	5.156.803	-240.526	-4,6	-4,8	-4,4
- nicht erwerbsfähige Hilfebedürftige (Sozialgeld)	1.910.546	1.909.947	1.929.370	1.936.470	-51.285	-2,6	-3,1	-2,7
GEMELDETE STELLEN								
- Zugang im Monat	244.219	255.517	240.389	220.099	-27.802	-10,2	3,8	2,0
dar.: ungefördert <sup>3)</sup>	188.808	187.071	178.078	155.590	-10.918	-5,5	-1,6	-1,3
- Zugang seit Jahresbeginn	1.617.734	1.373.515	1.117.998	877.609	-154.181	-8,7	-8,4	-10,8
dar.: ungefördert <sup>3)</sup>	1.191.247	1.002.439	815.368	637.290	-102.268	-7,9	-8,4	-9,8
- Bestand <sup>4)</sup>	588.420	596.269	579.180	592.168	-61.419	-9,5	-8,0	-9,9
dar.: ungefördert <sup>3)</sup>	424.446	420.764	401.787	399.236	-38.026	-8,2	-9,3	-11,9
sofort zu besetzen	523.904	518.713	515.004	527.429	-67.226	-11,4	-9,9	-10,4
TEILNEHMER AN AUSGEWÄHLTEN MASSNAHMEN AKTIVER ARBEITSMARKTPOLITIK <sup>2)</sup>								
- Bestand insgesamt	...	...	...	1.560.343	x	x	x	-3,7
dar.: Qualifizierung	227.916	247.095	245.156	256.323	25.429	12,6	8,1	3,1
Berufsberatung u. Förderung d. Berufsausbildung	...	...	...	337.562	x	x	x	-2,2
Beschäftigungsbegleitende Leistungen	364.624	366.189	361.610	355.993	-32.753	-8,2	-8,7	-9,6
Beschäftigung schaffende Maßnahmen	366.760	357.914	347.708	338.458	-699	-0,2	-4,1	-7,3
SAISONBEREINIGTE ENTWICKLUNG	July 08	Juni 08	Mai 08	April 08	März 08	Febr. 08	Jan. 08	Dez. 07
	Veränderung gegenüber Vormonat							
Erwerbstätige (Inland) <sup>1)</sup>	...	13.000	14.000	30.000	32.000	33.000	138.000	42.000
Sozialversicherungspflichtig Beschäftigte	...	...	9.000	17.000	41.000	57.000	109.000	77.000
Arbeitslose	-20.000	-36.000	4.000	-6.000	-47.000	-68.000	-85.000	-69.000
gemeldete Stellen (einschl. geförderte Stellen)	-9.000	7.000	-7.000	-10.000	-5.000	-4.000	-19.000	-6.000
ungeförderte Stellen <sup>3)</sup>	2.000	10.000	-	-5.000	-6.000	-11.000	-20.000	-8.000
- normale Stellen	...	1.000	-3.000	-4.000	-1.000	-2.000	-8.000	-2.000
Arbeitslosenquote bezogen auf alle zivilen EP	7,8	7,8	7,9	7,9	7,9	8,0	8,2	8,4
ILO Erwerbslosenquote <sup>1)</sup>	...	7,3	7,4	7,4	7,4	7,5	7,7	7,9

Datenstand: Juli 2008

© Statistik der Bundesagentur für Arbeit

1) Quelle: Statistisches Bundesamt

2) Endgültige Werte stehen erst nach einer Wartezeit fest. Am aktuellen Rand können die Daten aufgrund von Erfahrungswerten überwiegend hochgerechnet werden.

3) ohne PSA, ABM, BSI und Arbeitsgelegenheiten

4) Den Arbeitsagenturen waren im 4. Quartal 2007 laut Umfrage bei Betrieben 49 Prozent des gesamten Stellenangebotes gemeldet.

Abb. II.27: Arbeitsmarktzahlen der Bundesanstalt für Arbeit  
Quelle: Bundesagentur für Arbeit (Nürnberg), [www.arbeitsagentur.de](http://www.arbeitsagentur.de)

## **Statistiken über Erwerbstätige**

Hier werden vor allem personenbezogene Merkmale (Alter, Geschlecht, Ausbildung etc.) erfasst.

Aufwendige Erhebungen, Statistisches Bundesamt: Volkszählung, jährlicher Mikrozensus (vgl. Skriptum Kap. II.2)

Volkszählung:

Grundinformationen und Eckdaten über das Erwerbsleben

Mikrozensus:

lfd. Repräsentativstatistik der Bevölkerung und des Erwerbslebens bzw. der Erwerbsbeteiligung, überwiegende Unterhaltsquelle, Wirtschaftszweig, Stellung im Beruf, wöchentliche Arbeitszeit, Nettoeinkommen

## **Statistiken über Beschäftigte**

Überblick: 'Katalog der Statistiken zum Arbeitsgebiet der Bundesstatistik', Statistisches Bundesamt (Hrsg.), aktuellste Version

Daten über Befragungen von Betrieben → viele Sachverhalte auch branchenspezifisch

## **Arbeitsstättenzählungen**

etwa alle zehn Jahre, nun zusammen mit Volkszählung

Erhebungsbereich: alle Wirtschaftsbereiche

(Ausnahme: Land- und Forstwirtschaft mit ähnlichen eigenen Zählungen)

Vollständiger Überblick der Beschäftigten in tiefer fachlicher und regionaler Gliederung

erhebende Behörde: Statistisches Bundesamt

## **Statistiken für einzelne Wirtschaftsbereiche**

- Zensus im Produzierenden Gewerbe
- Handwerkszählung
- Handels- und Gaststättenzählung
- Landwirtschaftszählung

erhebende Behörde: Statistisches Bundesamt

## **Beschäftigten- und Entgeltstatistik (Beschäftigtenstatistik)**

Sie gilt als bedeutendste Quelle arbeitsmarktstatistischer Daten und wird seit 1973 durchgeführt.

Die Beschäftigten- und Entgeltstatistik hat eine Sonderstellung innerhalb der Statistiken der Erwerbstätigkeit und der Beschäftigten, da die von den Betrieben erhobenen Meldungen (Beschäftigungsfälle) personenbezogen zusammengeführt werden können.

Grundlage für die Beschäftigten- und Entgeltstatistik ist das Meldeverfahren zur Sozialversicherung und zur Bundesagentur für Arbeit

- Der Arbeitgeber liefert einen maschinenlesbaren Beleg über alle versicherungspflichtigen Beschäftigte bei einem Wechsel der Arbeitsstätte (Anmeldung und Abmeldung) und Meldung am Jahresende für alle Beschäftigten.
- 75 % aller Erwerbstätigen werden dadurch erfaßt.
- Nicht enthalten sind Beamte, Selbständige und mithelfende Familienangehörige.
- Bundesagentur für Arbeit (BA) führt für jeden Versicherten ein Konto.
- Personenbezogene Zusammenführung erlaubt es, die Diskrepanz zwischen der Zahl der Beschäftigungsfälle und der Zahl der Erwerbstätigen aufzuheben.

### **Arbeitsmarktstatistiken**

Bundesagentur für Arbeit, organisatorisch selbständig, eigene Behörden (lokale Arbeitsämter, Landesarbeitsämter)

Veröffentlichungen der BA: Amtliche Nachrichten der Bundesagentur für Arbeit (ANBA), Mitteilungen aus der Arbeitsmarkt- und Berufsforschung (MittAB)

Die BA führt Statistiken über

- Beschäftigte,
- Nichtarbeitslose (Arbeitsuchende),
- Arbeitslose,
- offene Stellen,
- Arbeitsvermittlungen,
- Kurzarbeiter etc.

### **Nichtamtliche Statistiken**

Aus der Umfrageforschung mit problemorientierter Vertiefung:

- z.B. Sozio-ökonomisches Panel
- Sfb 3-Nebenerwerbstätigkeitsumfrage

Überblick über erhältliche Mikrodaten: Zentralarchiv in Köln (siehe auch Skriptum Kap. II.1)

Veröffentlichungen des Statistischen Bundesamtes zur Erwerbstätigkeit: Fachserie 1 (Bevölkerung und Erwerbstätigkeit), Reihe 4 und Einzelveröffentlichungen

Beteiligung am Erwerbsleben sowie Erwerbstätige nach der Stellung im Beruf und nach dem Wirtschaftszweig sind in den Tab. II.15-17 zu finden.

Tab. II.15a: Freie Berufe und Erwerbstätige nach Stellung im Beruf in Rheinland-Pfalz und in der Bundesrepublik Deutschland im Jahr 1987

Erwerbstätige	Rheinland-Pfalz		Bundesrepublik	
	Anzahl (in Tsd.)	% <sup>2</sup>	Anzahl (in Tsd.)	%
<b>Selbständige</b>	140,9	9,0	2.296,0	8,5
davon in freien Berufen	17,8		369,6	
Freiberufleranteil	12,6 %		16,1 %	
<b>Beamte</b>	151,6	9,7	2.440,2	9,1
<b>Angestellte</b>	603,9	38,4	11.034,7	41,0
<b>Arbeiter</b>	636,4	40,5	10.658,0	39,6
<b>Erwerbstätige insgesamt <sup>1</sup></b>	1.571,4	100,0	26.907,5	100,0

Quelle: Arbeitsstättenzählung 1987, Sonderauswertungen des Statistischen Landesamtes Rheinland-Pfalz und des Statistischen Bundesamtes, eigene Berechnungen (Merz, Rauberger und Rönnau (1994), S. I.66)

1 Die Erwerbstätigen insgesamt enthalten auch die hier nicht aufgeführten Mithelfenden Familienangehörigen

2 Die Prozente werden auf Grundlage der Erwerbstätigen insgesamt ermittelt.

Tab. II.15b: Beteiligung am Erwerbsleben 1970, 1987

Land	Jahr	Erwerbstätige in 1000		Erwerbslose in 1000	
		männlich	weiblich	männlich	weiblich
Schleswig-Holstein	1970	681,9	350,6	3,3	2,5
	1987	690,5	430,9	63,7	44,4
Hamburg	1970	503,9	323,8	2,7	2,1
	1987	396,4	294,4	53,6	37,0
Niedersachsen	1970	1966,4	1038,6	9,5	6,6
	1987	1903,7	1133,6	163,8	124,2
Bremen	1970	203,3	106,6	1,1	0,7
	1987	165,9	105,6	22,7	17,2
Nordrhein-Westfalen	1970	4747,2	2209,5	21,2	11,5
	1987	4419,0	2514,0	389,4	278,0
Hessen	1970	1557,1	845,1	5,3	3,2
	1987	1539,6	951,2	80,8	70,8
Rheinland-Pfalz	1970	999,6	523,5	2,8	1,2
	1987	1004,2	567,2	61,6	50,6
Baden-Württemberg	1970	2576,5	1599,5	6,4	5,2
	1987	2651,0	1703,7	110,9	97,0
Bayern	1970	2935,7	1958,8	10,0	8,2
	1987	3090,5	2006,5	143,8	131,5
Saarland	1970	285,7	120,5	2,3	0,9
	1987	270,6	141,6	33,1	18,8
Berlin (West)	1970	546,9	412,7	5,9	4,2
	1987	524,2	403,1	54,90	38,7
Bundesgebiet	1970	17004,2	9489,3	70,5	46,1
	1987	16655,7	10251,9	1178,3	908,1

Quelle: Statistisches Bundesamt 1989, VZ 1970, 1987

Tab. II.16: Erwerbstätige nach der Stellung im Beruf 1970, 1987

Land	Jahr	insgesamt	Selbstständige	Mittel-fende Familienangehörige	Beam-te, Rich-ter, Sol-daten, usw.	Ange-stellte, Auszu-bildende (kfm./techn.)	Arbeiter, Auszu-bildende (gewerblich)
		in 1000	%				
Schleswig-Holstein	1970	1032,6	10,7	6,1	10,8	31,9	40,5
	1987	1121,4	9,2	2,1	12,5	42,0	34,2
Hamburg	1970	827,7	8,5	2,6	7,4	45,2	36,4
	1987	690,8	8,5	1,0	8,6	50,9	31,0
Niedersachsen	1970	3005,1	10,4	7,9	8,2	28,3	45,2
	1987	3037,2	8,5	2,1	10,6	39,4	39,5
Bremen	1970	309,9	7,5	1,6	8,6	42,1	40,1
	1987	271,5	6,7	0,8	10,0	45,3	37,2
Nordrhein-Westfalen	1970	6956,6	8,5	3,4	6,7	33,3	48,1
	1987	6933,1	7,9	1,3	8,8	42,0	39,9
Hessen	1970	2402,1	9,2	5,8	7,5	33,7	43,9
	1987	2490,8	8,3	1,4	8,5	44,7	37,1
Rheinland-Pfalz	1970	1523,1	11,3	8,3	8,0	27,2	45,1
	1987	1571,4	9,0	2,5	9,6	38,4	40,5
Baden-Württemberg	1970	4176,1	9,3	7,0	6,1	29,5	48,1
	1987	4354,7	8,1	1,7	8,0	40,3	42,0
Bayern	1970	4894,5	11,9	10,2	7,3	26,5	44,1
	1987	5097,0	9,9	2,6	8,7	37,8	40,9
Saarland	1970	406,2	7,6	3,3	8,8	31,0	49,4
	1987	412,3	7,5	0,9	10,3	38,5	42,8
Berlin (West)	1970	959,7	7,5	1,8	7,1	38,7	45,0
	1987	927,3	7,7	0,6	9,4	45,4	37,0
Bundesgebiet	1970	26493,5	9,7	6,2	7,3	31,1	45,6
	1987	26907,5	8,5	1,8	9,1	41,0	39,6

Quelle: Statistisches Bundesamt 1989, VZ 1970, 1987

Tab. II.17: Erwerbstätigkeit nach Wirtschaftsbereichen 1970, 1987

Land	Jahr	Land- und Forstwirtschaft, Fischerei	Produzie- rendes Gewerbe*	Handel, Verkehr und Nachrichten- übermittlung	Übrige Wirtschafts- bereiche**
Schleswig- Holstein	1970	9,4	37,0	21,1	32,5
	1987	4,9	30,3	20,9	43,2
Hamburg	1970	1,2	35,4	30,2	33,2
	1987	1,1	27,3	26,1	45,5
Niedersachsen	1970	10,9	44,6	18,2	26,2
	1987	5,0	38,2	17,8	39,1
Bremen	1970	1,6	37,4	30,9	30,1
	1987	0,8	32,8	24,8	41,6
Nordrhein- Westfalen	1970	3,5	53,8	18,5	24,3
	1987	2,0	43,6	17,7	36,8
Hessen	1970	6,3	48,6	18,4	26,6
	1987	2,1	39,9	19,9	38,2
Rheinland-Pfalz	1970	10,7	45,9	16,9	26,4
	1987	4,6	41,4	16,8	37,2
Baden- Württemberg	1970	7,9	54,7	14,3	23,1
	1987	2,7	48,1	15,3	33,8
Bayern	1970	13,2	47,2	15,8	23,8
	1987	5,1	44,0	16,4	34,5
Saarland	1970	2,4	51,4	20,0	26,2
	1987	1,0	43,0	18,0	37,7
Berlin (West)	1970	0,5	43,4	20,8	35,2
	1987	0,6	31,3	18,5	49,6
Bundesgebiet	1970	7,5	48,9	17,9	25,7
	1987	3,2	41,8	17,7	37,3

\* Energie- und Wasserversorgung, Bergbau, Verarbeitendes Gewerbe, Baugewerbe

\*\* Kreditinstitute und Versicherungsgewerbe, Dienstleistungen, soweit von Unternehmen und Freien Berufen erbracht, Organisationen ohne Erwerbszweck und Private Haushalte, Gebietskörperschaften und Sozialversicherung.

Quelle: Statistisches Bundesamt 1989, VZ 1970, 1987

## 5.2 Konzepte zur Erfassung der Erwerbstätigkeit

### Potentialkonzept

Erfassung der 'Bevölkerung im erwerbsfähigen Alter', alle Personen zwischen 15 und 65 Jahren.

Probleme: nicht effektiv erwerbstätige Bevölkerung

- Trend zu erhöhter Schulbildung (d.h. noch in Ausbildung, z.B. bei 15 Jahren)
- Nicht dem Arbeitsmarkt zur Verfügung stehende werden einbezogen, wie z.B. ausschließliche im Haushalt tätige Frauen, Kranke, 'discouraged worker' etc.
- es gibt auch noch eine beträchtliche Anzahl von Menschen, die über das 65. Lebensjahr arbeiten
- Vorruhestand etc.

### Unterhaltskonzept

Erfasst werden nur solche Personen, die ihren **Lebensunterhalt überwiegend** aus der **Erwerbstätigkeit** bestreiten.

Dies führt zu bedeutend niedrigeren Zahlen der Erwerbstätigen wegen eventueller Teilzeitbeschäftigung, Ausbildung (Elternzuschuß) etc.

### Hauptberufskonzept

Sonderfall des Unterhaltskonzeptes;

Es werden lediglich die in der Ausbildung befindlichen Personen zusätzlich mit erfaßt.

### Erwerbskonzept

Erfasst werden *'alle Personen mit Wohnsitz im Bundesgebiet (Inländerkonzept), die eine unmittelbar oder mittelbar auf Erwerb ausgerichtete Tätigkeit auszuüben pflegen, unabhängig von der Bedeutung für ihren Lebensunterhalt und ohne Rücksicht auf die von ihnen tatsächlich oder vertragsmäßig zu leistende Arbeitszeit'* (Statistisches Bundesamt, Statistisches Jahrbuch 1993, S. 109)

- **'mittelbar auf Erwerb ausgerichtete Tätigkeit'** meint z.B. auch Frauen von Selbständigen (Landwirtschaft, Einzelhandel), die helfen, ohne dafür ein Entgelt zu beziehen.
- **'auszuüben pflegen'**: es werden auch Erwerbslose erfasst.

### Labour-force-Konzept

Erfasst im wesentlichen den Personenkreis des Erwerbskonzeptes. Ausgeschlossen sind aber Personen, die aus eigenem Entschluss **weniger als 15 Stunden pro Woche arbeiten**.

Wandel bei Gebrauch dieser Konzepte:

- bis Ende der 50er Jahre: Unterhaltskonzept (Bevölkerungsstatistik)
- heute ist das **Erwerbskonzept** das gebräuchlichste Konzept

Kombinationen geben weitere Informationen über die Bevölkerungsstruktur: so wird z.B. seit der Volkszählung 1961

- die Frage nach der Erwerbsbeteiligung mit
  - der Frage nach dem überwiegenden Lebensunterhalt
- verknüpft (vgl. dazu Tab. II.18).

Merkmal	Erwerbstätige		Erwerbslose	
	in 1000	in %	in 1000	in %
männlich	20.637	53,9	1.695	8,2
weiblich	17.633	46,1	1.249	7,1
Insgesamt	38.270	100,0	2.944	7,7

Abb. II.28a: Eckdaten zu Erwerbstätigkeit und Erwerbslosigkeit 2010 – Ergebnisse des Mikrozensus  
 Quelle: Statistisches Bundesamt, Wiesbaden (2011)

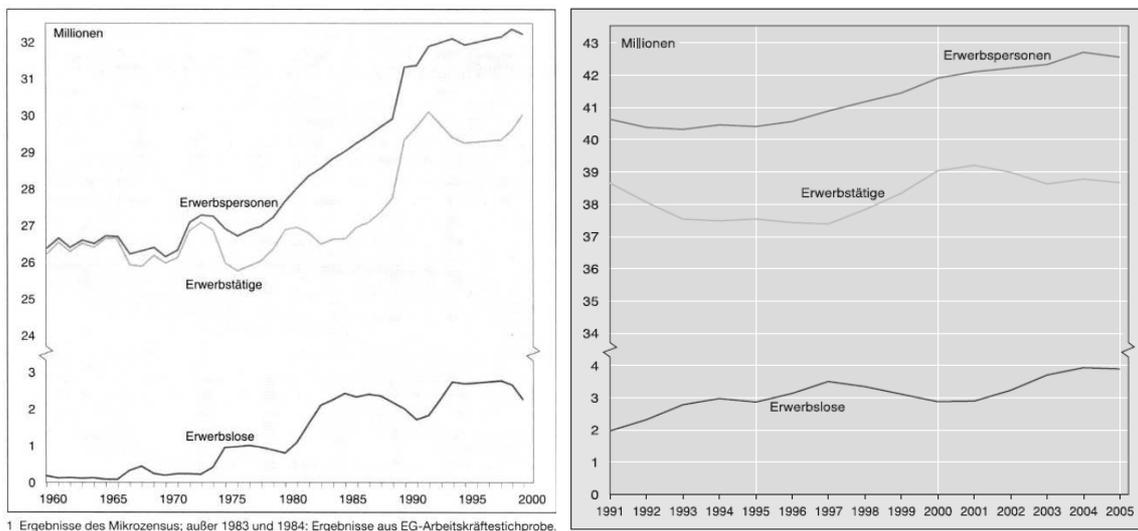


Abb. II.28b: Entwicklung von Erwerbstätigkeit und Erwerbslosigkeit seit 1960 im früheren Bundesgebiet (Ergebnisse des Mikrozensus; außer 1983 und 1984: Ergebnisse aus EG-Arbeitskräftestichprobe) und von 1991 bis 2005

Quelle: Datenreport 2002, S. 86 und Datenreport 2006, S. 86

Tab. II.18: Bevölkerung in den alten Bundesländern nach der Beteiligung am Erwerbsleben und dem überwiegenden Lebensunterhalt 2000 und 2010

Personen mit überwiegendem Lebensunterhalt durch:	Beteiligung am Erwerbsleben (Erwerbskonzept)								
	Jahr	Erwerbstätige		Erwerbslose		Nichterwerbs- personen		Insgesamt	
		in 1000	in %	in 1000	in %	in 1000	in %	in 1000	in %
Erwerbstätigkeit	2000	33.702	100,00	x		x		33.702	41,02
	2010	35.113	99,44	24	0,07	174	0,49	35.311	44,14
Arbeitslosen geld/hilfe	2000	138	4,94	2.655	95,06	x		2.793	3,40
	2010	648	22,70	2.207	77,30	1711	59,93	2.855	3,57
Renten und dgl.	2000	832	3,92	457	2,16	19.911	93,92	21.200	25,80
	2010	1045	5,06	129	0,62	19.469	94,31	20.643	25,80
Angehörige	2000	1.931	7,89	611	2,50	21.921	89,61	24.463	29,78
	2010	2.132	10,06	589	2,78	18.475	87,16	21.196	26,49
Insgesamt	2000	36.603	44,55	3.723	4,53	41.832	50,92	82.158	100,00
	2010	38.938	48,67	2.949	3,69	39.829	49,78	80.005	100,00

x = Feld aus sachlichen Gründen nicht besetzt

Quelle: Statistisches Bundesamt, Wiesbaden (2011)

### 5.3 Indikatoren und Maßzahlen

Gliederung der **Wohnbevölkerung** nach der **Beteiligung am Erwerbsleben** (Erwerbskonzept) in Abb. II.28

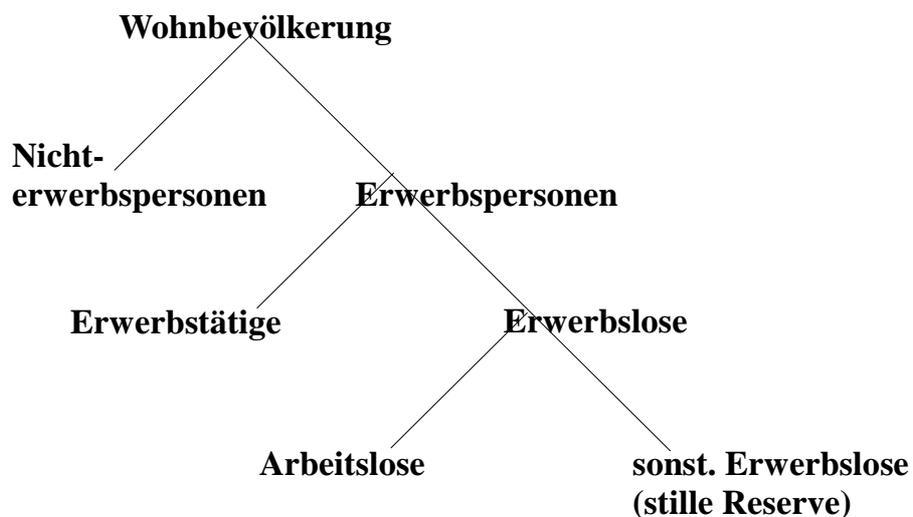


Abb. II.28: Gliederung der Wohnbevölkerung nach dem Erwerbskonzept

**Nichterwerbspersonen**

alle Personen, die keinerlei auf Erwerb gerichtete Tätigkeit ausüben oder suchen (z.B. Studenten, Schulkinder, Rentner);

**Erwerbspersonen**

siehe Erwerbskonzept: 'alle ... Arbeitszeit';  
(Selbständige, Mithelfende Familienangehörige, Abhängig Beschäftigte);

**Erwerbstätige**

Personen, die in einem Arbeitsverhältnis stehen (einschließlich Soldaten und Mithelfende Familienangehörige oder Selbständige, (darunter Freie Berufe));

**Erwerbslose**

Personen ohne Arbeitsverhältnis, die sich um eine Arbeitsstelle bemühen

**Arbeitslose:**

Personen ohne Arbeitsverhältnisse, die als Arbeitssuchende bei einer Arbeitsagentur registriert sind.

**Stille Reserve (verdeckte Arbeitslose)**

Erwerbslose ./ Arbeitslose

- Aktive stille Reserve:  
nichtbeschäftigt, arbeitsuchend ohne Arbeitsagentur
- Passive stille Reserve:  
Personen, die von der angespannten Arbeitsmarktlage entmutigt, keine Arbeit mehr suchen (discouraged worker);

Die stille Reserve ist der Teil des Erwerbspersonenpotentials, der nicht in der offiziellen Arbeitsmarktstatistik erfasst wird.

Kenntnisse über Art und Umfang der stillen Reserve sind wichtig zur besseren Beurteilung der Arbeitsmarktlage.

Zur Schätzung der stillen Reserve existieren unterschiedliche Konzepte:

**Stille Reserve - Konzept des IAB:**

Stille Reserve = Erwerbspersonenpotential bei Hochkonjunktur abzüglich Erwerbstätige und Arbeitslose,

wobei das 'Erwerbspersonenpotential bei Hochkonjunktur' regressionsanalytisch auf der Basis von makroökonomischen Daten bestimmt wird.

**Stille Reserve - Konzept des Sfb 3:**

Stille Reserve = Anzahl der Personen, die nicht erwerbstätig, nicht arbeitslos, nicht in Ausbildung, aber erwerbsbereit sind, und die Aufnahme einer Erwerbstätigkeit (Voll- oder Teilzeit) binnen Jahresfrist wünschen.

Datengrundlage: Sozio-ökonomisches Panel des Sfb 3/DIW (Mikrodaten) mit jährlichen Angaben von mehr als 12.000 Personen

In Abhängigkeit der Konjunktur lassen sich Wanderungen der Wohnbevölkerung zwischen Erwerbs- und Nicht-Erwerbspersonen beobachten.

- 'discouraged worker effect' (im weiteren Sinne [z.B. Frührentner, Vorruhestand])  
Erwerbspersonen → Nicht-Erwerbspersonen
- 'additional worker effect' (z.B. Arbeitslosigkeit eines Partners zwingt den anderen Partner zur Arbeitssuche)  
Nicht-Erwerbspersonen → Erwerbspersonen

### Maßzahlen:

Zur Beteiligung am Erwerbsleben gegliedert nach dem Geschlecht, der Stellung im Beruf und nach Wirtschaftszweigen vgl. Tab. II.15-17.

$$\text{Allgemeine Erwerbsquote} = \frac{\text{Erwerbspersonen}}{\text{Wohnbevölkerung}}$$

Für den internationalen Vergleich (z.B. mit sehr kinderreichen Staaten) ist die folgende spezifische Erwerbsquote besser geeignet:

$$\text{spezifische Erwerbsquote} = \frac{\text{Zahl der Erwerbspersonen}}{\text{Bevölkerung im Alter von 15 - 65 Jahren}}$$

Eckdaten zur Erwerbstätigkeit und Erwerbslosigkeit finden sich in Abb. II.28a.

$$\text{Arbeitslosenquote} = \frac{\text{Arbeitslose}}{\text{abhängige (!) Erwerbspersonen}}$$

Problem unterschiedlicher Quellen: Arbeitslose (BA), Erwerbspersonen (MZ)

Nach Auskunft der Bundesagentur für Arbeit in Nürnberg werden für die Bestimmung der Arbeitslosenquote die abhängigen Erwerbspersonen in folgender Weise berücksichtigt:

$$\begin{aligned} \text{Abhängige, zivile Erwerbspersonen} = & \text{ Sozialversicherungspflichtige Beschäftigte} \\ & \text{ (Arbeiter, Angestellte einschl. Auszubildende)} \\ & + \text{ Geringfügig Beschäftigte} \\ & + \text{ Beamte (ohne Soldaten)} \\ & + \text{ Arbeitslose} \end{aligned}$$

Die abhängigen, zivilen Erwerbspersonen werden für die Berechnung der Arbeitslosenquote verwendet. Die so ermittelte Arbeitslosenquote wird auch von der Bundesagentur veröffentlicht.

Intern ermittelt die Bundesagentur für Arbeit noch eine weitere Arbeitslosenquote, die einen größeren Erwerbspersonenkreis berücksichtigt:

Alle zivilen Erwerbspersonen = alle abhängig Beschäftigten (alle abh. Erwerbst.)  
+ Selbständige  
+ mithelfende Familienangehörige

Eurostat verwendet als Grundlage der europäischen Arbeitslosenstatistik die 1982 erarbeitete Definition des International Labour Office (ILO) nach der jede Person über 15 als arbeitslos gilt, die ohne Arbeit ist, bereit steht eine Arbeit innerhalb der nächsten zwei Wochen aufzunehmen und während der letzten vier Wochen eine Arbeit suchte (Quelle: Eurostat).

Das Department of Labor der USA berechnet die Arbeitslosigkeit über eine Zufallsstichprobe von 60.000 Haushalten, deren Ergebnis auf die Gesamtbevölkerung hochgerechnet wird. Nach der Definition des Bureau of Labor Statistics gelten Personen als arbeitslos, die in der Referenzwoche ohne Arbeit waren, bereit waren eine Arbeit anzunehmen und in einem vier Wochen Zeitraum vor Ende der Referenzwoche Maßnahmen ergriffen haben eine Arbeit zu bekommen (z.B. Vermittlungen kontaktiert haben). Als erwerbstätig gilt jede Person die in der Referenzwoche eine beliebige (!) Zeit einer bezahlten Arbeit nachgegangen ist oder mehr als 15 Stunden unbezahlt in einem Familienbetrieb gearbeitet haben. (Quelle: US Department of Labor, z.B. unter [http://www.bls.gov/opub/hom/homch1\\_c.htm](http://www.bls.gov/opub/hom/homch1_c.htm), 15.05.2003)

**Zur Abgrenzung von Arbeitslosen** (Quelle: Sozialgesetzbuch (SGB) Drittes Buch (III) – Arbeitsförderung, § 119 Arbeitslosigkeit, URL: [http://bundesrecht.juris.de/sgb\\_3/\\_119.html](http://bundesrecht.juris.de/sgb_3/_119.html)):

- (1) Arbeitslos ist ein Arbeitnehmer, der
  1. nicht in einem Beschäftigungsverhältnis steht (Beschäftigungslosigkeit),
  2. sich bemüht, seine Beschäftigungslosigkeit zu beenden (Eigenbemühungen) und
  3. den Vermittlungsbemühungen der Agentur für Arbeit zur Verfügung steht (Verfügbarkeit).
  
- (2) Eine ehrenamtliche Betätigung schließt Arbeitslosigkeit nicht aus, wenn dadurch die berufliche Eingliederung des Arbeitslosen nicht beeinträchtigt wird.
  
- (3) Die Ausübung einer Beschäftigung, selbständigen Tätigkeit oder Tätigkeit als mithelfender Familienangehöriger (Erwerbstätigkeit) schließt die Beschäftigungslosigkeit nicht aus, wenn die Arbeits- oder Tätigkeitszeit (Arbeitszeit) weniger als 15 Stunden wöchentlich umfasst; gelegentliche Abweichungen von geringer Dauer bleiben unberücksichtigt. Die Arbeitszeiten mehrerer Erwerbstätigkeiten werden zusammengerechnet.
  
- (4) Im Rahmen der Eigenbemühungen hat der Arbeitslose alle Möglichkeiten zur beruflichen Eingliederung zu nutzen. Hierzu gehören insbesondere
  1. die Wahrnehmung der Verpflichtungen aus der Eingliederungsvereinbarung,
  2. die Mitwirkung bei der Vermittlung durch Dritte und
  3. die Inanspruchnahme der Selbstinformationseinrichtungen der Agentur für Arbeit.
  
- (5) Den Vermittlungsbemühungen der Agentur für Arbeit steht zur Verfügung, wer
  1. eine versicherungspflichtige, mindestens 15 Stunden wöchentlich umfassende zumutbare Beschäftigung unter den üblichen Bedingungen des für ihn in Betracht kommenden Arbeitsmarktes ausüben kann und darf,
  2. Vorschlägen der Agentur für Arbeit zur beruflichen Eingliederung zeit- und ortsnahe Folge leisten kann,

3. bereit ist, jede Beschäftigung im Sinne der Nummer 1 anzunehmen und auszuüben und
4. bereit ist, an Maßnahmen zur beruflichen Eingliederung in das Erwerbsleben teilzunehmen.

Neben den bisher genannten angebotsorientierten Arbeitsmarktindikatoren existiert mit der bei der Arbeitsagentur gemeldeten '**Zahl der offenen Stellen**' ein Indikator für die Nachfrageseite. Problematisch ist in diesem Zusammenhang jedoch, dass

- offene Stellen z.T. nicht bei den Arbeitsagenturen gemeldet sind,
- Stellen mehrfach als nicht besetzt gemeldet werden.

Kombiniert man die beiden Indikatoren 'Anzahl der Arbeitslosen' und 'Zahl der offenen Stellen', läßt sich ein Index konstruieren, mit dessen Hilfe man die Anspannung auf dem Arbeitsmarkt darstellen kann:

$$\text{Anspannungsindex} = \frac{\text{Zahl der Arbeitslosen}}{\text{Zahl der offenen Stellen}}$$

Probleme können sich aus den Teilindizes kumulieren.

## 6 Einkommens- und Verbrauchsstatistik, Preise

### 6.1 Einkommens- und Verbrauchsstatistik

Grundlagen für Einkommens- und Einkommensverteilungsanalysen

Erwerbs- und Vermögenseinkommen

1. aus unselbständiger Arbeit
2. aus Unternehmertätigkeit
3. aus Vermögen (Vermieten, Verpachten)
4. aus staatlichen und privaten Übertragungen

#### - funktionelle Einkommensverteilung

nach Produktionsfaktoren Arbeit und Kapital (Lohnquote, Gewinnquote)

#### - personelle Einkommensverteilung

Individual- und Haushaltseinkommen nach sozioökonomischen Gruppen

**Traditionelle Einkommenserhebungen** wie

- lfd. Verdiensterhebungen (Lohnsummenverfahren)
- Arbeitsstättenzählung (Lohnsummenverfahren)
- Lohn- und Gehaltsstrukturserhebungen
- Einkommen- und Lohnsteuerstatistik\*
- Tariflohnstatistik

sind (bis auf \*) nicht zur Berechnung von Einkommensschichtungen geeignet.

**Wirtschaftsrechnungen privater Haushalte**

(Aufzeichnungen von Einnahmen und Ausgaben in Haushaltsbüchern)

- lfd. Wirtschaftsrechnungen
- Einkommens- und Verbrauchsstichproben (EVS)

### **lfd. Wirtschaftsrechnungen**

#### Altes System:

- Haushaltstyp 1: 2-Personen-Haushalte von Renten- und Sozialhilfeempfängern mit geringem Einkommen
- Haushaltstyp 2: 4-Personen-Arbeitnehmerhaushalte mit mittlerem Einkommen (alleinvertienend, Ehepaare mit zwei Kindern)
- Haushaltstyp 3: 4-Personen-Haushalte von Beamten und Angestellten mit höherem Einkommen (Ehepaare mit zwei Kindern)

Diese Haushaltstypen sind nicht repräsentativ für alle Haushalte (nur 5-10 % aller privaten Haushalte), umfassen aber wirtschafts- und sozialpolitisch bedeutende Gruppen. Es fehlen bspw. die Selbständigen. Die Stichproben umfassen nur 100-500 Haushalte.

#### Neues System (seit 1999):

Befragung von 6000 Privathaushalten (außer Anstaltshaushalten und Haushalten von Selbstständigen), wobei jeden Monat 2000 Haushalte wechselnd befragt werden, so dass jeder Haushalt pro Quartal einmal befragt wird

### **Einkommens- und Verbrauchsstichproben (EVS)**

#### Altes System:

fast 50.000 Haushalte, freiwillig geführte Haushaltsbücher (Feinanschreibung 1 Monat (Oktober), Hauptanschreibung 11 Monate, Ausgewähltes in Gruppen); Grund- und Schlußinterview für Sozioökonomie, Vermögen

#### Neues System (seit 1998):

Ca. 74.000 Haushalte als Quotenstichprobe, jeweils 18.500 führen im Rotationsverfahren ein Quartal ein Haushaltsbuch (Feinaufschreibung für jeweils 20 % der Haushalte für einen Monat im jeweiligen Quartal), Aufzeichnungsperiode fünf Jahre, Einganginterview, erfasst werden alle privaten Haushalte außer Anstaltshaushalten und Haushalten mit einem Haushaltsnettoeinkommen über 17.895 EUR, dargestellt werden: die Ausstattung mit langlebigen Gebrauchsgütern, Einnahmen, Ausgaben, Sach- und Geldvermögen, Preise und Mengen von Nahrungs- und Genussmitteln  
Die Ergebnisse werden über den Mikrozensus gewichtet

Zu den Ausgabefähigen Einkommen und Einnahmen privater Haushalte 2005 vgl. Abb. II.30.

	Deutschland	Früheres Bundesgebiet	Neue Länder (einschl. Berlin-Ost)
Erfasste Haushalte (Anzahl)	7 579	6 095	1 484
Hochgerechnete Haushalte (in 1 000)	35 555	28 838	6 717
Durchschnitt je Haushalt und Monat in EUR			
<b>Haushaltsbruttoeinkommen</b>	<b>3 496</b>	<b>3 665</b>	<b>2 766</b>
abzüglich:			
Einkommens-, Kirchensteuer und Solidaritätszuschlag	353	385	217
Pflichtbeiträge z. Sozialversicherung	376	387	329
<b>Haushaltsnettoeinkommen</b>	<b>2 766</b>	<b>2 893</b>	<b>2 220</b>
zuzüglich:			
Einnahmen aus dem Verkauf von Waren und sonstige Einnahmen	50	53	35
<b>Ausgabefähige Einkommen u. Einnahmen</b>	<b>2 816</b>	<b>2 946</b>	<b>2 255</b>

Ergebnisse der LWR.

Abb. II.30: Ausgabefähige Einkommen und Einnahmen privater Haushalte 2005 (Ergebnisse der LWR)

Quelle: Datenreport 2008

Tiefgegliederte Individualanalysen werden möglich!

Ausgabenmodell: z.B. Merz (1980a), Die Ausgaben privater Haushalte - Ein mikroökonomisches Modell für die Bundesrepublik Deutschland, Frankfurt a.M./ New York.

Einkommensverteilung und Äquivalenzzahlen: Merz und Faik (1992), Equivalence Scales Based on Revealed Preference Consumption Expenditure Microdata, FFB-Diskussionspapier Nr. 3, Lüneburg

Zur Einkommensverteilung in Deutschland 2006 vergleiche Abb. II.31

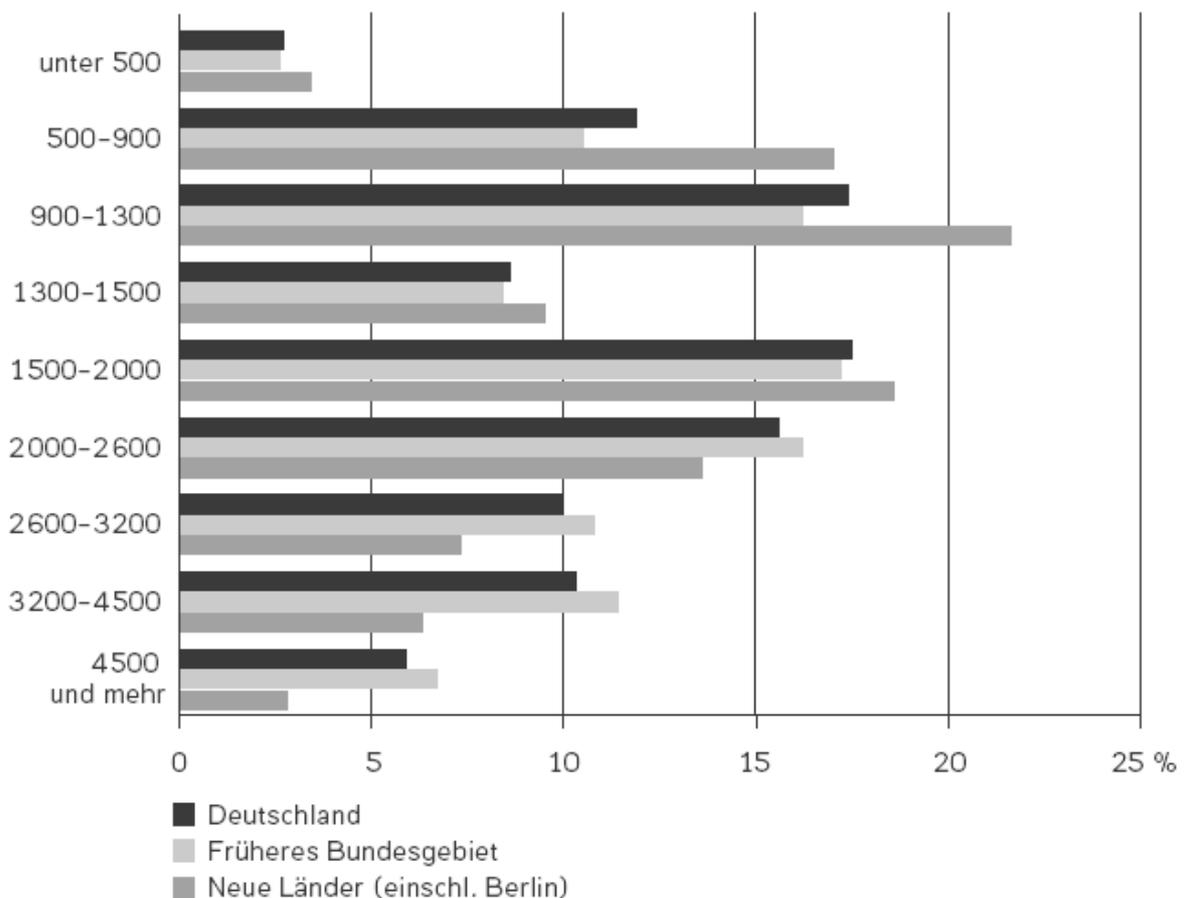


Abb. II.31: Einkommensverteilung 2006 – monatliches Haushalts-Nettoeinkommen, in EUR  
 Ergebnisse des Mikrozensus – Bevölkerung in Privathaushalten am Haupt- und Nebenwohnsitz. Ohne  
 Selbstständige in der Landwirtschaft und ohne Haushalte, die keine Angabe zum Einkommen machten  
 Quelle: Datenreport 2008

Weitere Informationen: Rinne (1994), Kap. 7 und Datenreport 2002, Kap. 4

## 6.2 Preisstatistik

Preise sind ein wichtiges Steuerungsinstrument vor allem in einer marktwirtschaftlich orientierten Wirtschaft.

Nominale Größen

Preissteigerungen, Inflation

Reale Größen

(von Preisbewegungen bereinigt)

- **Einzelpreise** (Mittlere Preise)
- **Preisindex** für die Kennzeichnung der Preisentwicklung von Gütergesamtheiten

Ein Preisindex beantwortet die Frage, in welchem Maße Preisveränderungen für die Veränderung der Wertsumme einer Gütergesamtheit verantwortlich sind (Preisindex nach Paasche und nach Laspeyres, siehe Skriptum Kap.V).

z.B. Preisindex für die Lebenshaltung für unterschiedliche Haushaltstypen.  
 Basis: Warenkorb (ca. 800 Güter), vgl. Tab. II.19.

Tab. II.19: Verbraucherpreisindex und Veränderungsraten für Deutschland

Jahr	Verbraucherpreisindex	Veränderung zum Vorjahr
	2005 = 100	in %
1991	75,9	.
1992	79,8	5,1
1993	83,3	4,4
1994	85,6	2,8
1995	87,1	1,8
1996	88,3	1,4
1997	90,0	1,9
1998	90,9	1,0
1999	91,4	0,6
2000	92,7	1,4
2001	94,5	1,9
2002	95,9	1,5
2003	96,9	1,0
2004	98,5	1,7
2005	100,0	1,5
2006	101,6	1,6
2007	103,9	2,3
2008	106,6	2,6
2009	107,0	0,4
2010	108,2	1,1

Quelle: Statistisches Jahrbuch 2010

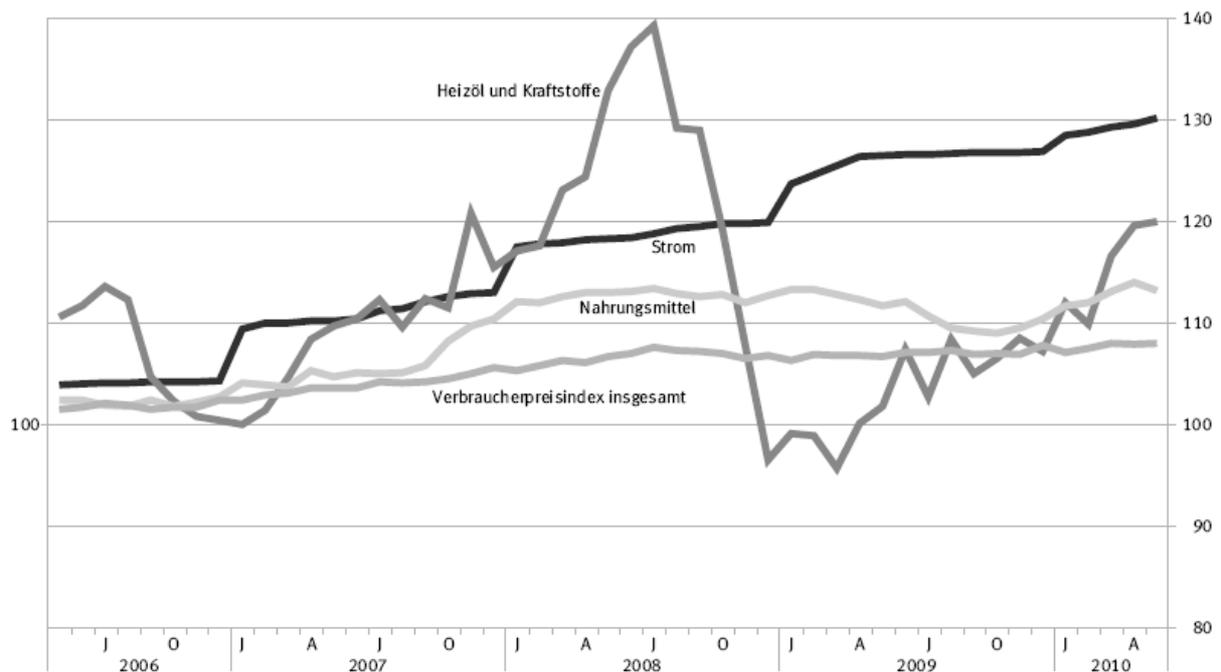


Abb. II.32: Verbraucherpreisindex für Deutschland nach Hauptgruppen (2005 = 100)

Quelle: Statistisches Jahrbuch 2010

## 7 Umwelt

Zunehmende Umweltbelastungen als Folge der industriellen und technischen Entwicklung (wie Luft- (Klima, Ozon) und Wasserverschmutzung, Lärmbelästigung) haben Natur- und Umweltschutz in den letzten Jahren vermehrt öffentliche Aufmerksamkeit verschafft.

Institutionen/Programme/Rechtsvorschriften: Umweltprogramme, Umweltbundesamt, Sachverständigenrat für Umweltfragen, Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (1987), Bundes-Emissionsgesetz, Abwasserabgabegesetz, Abfallgesetz.

Aktuelle Diskussion: Umwelt als Staatsziel

Statistisches Bundesamt: Seit 1975 in mehrjährigen Abständen Statistiken über Wasserversorgung, Abwasser- und Abfallversorgung

Waldschadenerhebung seit 1992, naturwissenschaftlich-technische Daten, Meinungsumfragen

Wirtschaft: quantitatives → qualitatives Wachstum  
zunehmend Umweltschutz als Wirtschafts- und Beschäftigungsfaktor

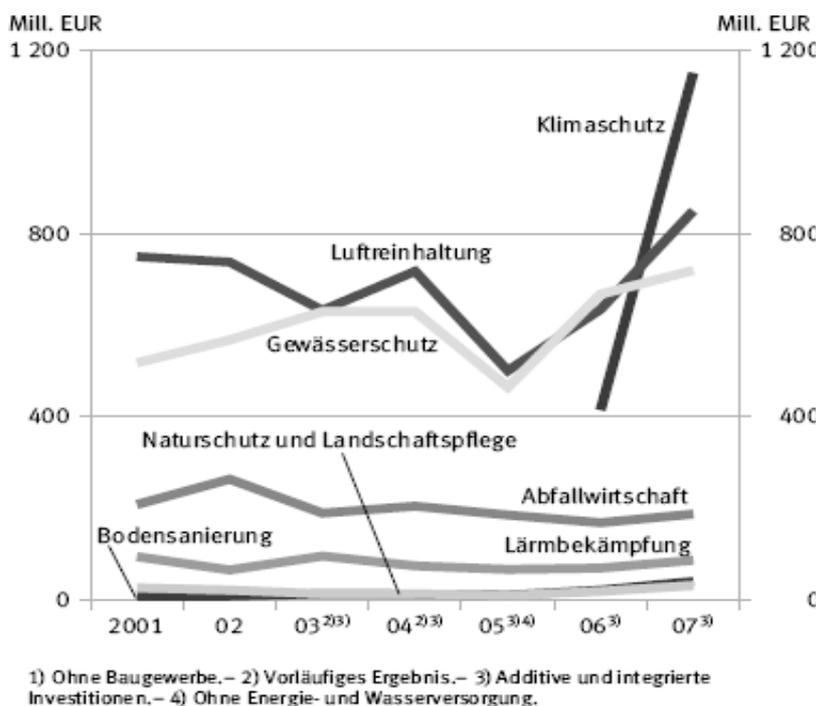


Abb. II.33a: Investitionen für Umweltschutz im Produzierenden Gewerbe<sup>1</sup>  
Quelle: Statistisches Jahrbuch 2010

## Umwelt im Datenreport 2002:

Umweltökonomische Trends

Umweltbelastungen

Luftbelastung

Verpackungen

Wassergewinnung und Abwasserbeseitigung

Dünger, Pflanzenschutzmittel und Streusalz

Unfälle mit wassergefährdenden Stoffen

Umweltzustand

Waldschäden

Gefährdete Tiere und Pflanzen

Gesundheitsbeeinträchtigung der Bevölkerung

Umweltschutz

Internationales Umweltrecht

Umweltschutzeinrichtungen

Ausgaben für Umweltschutz

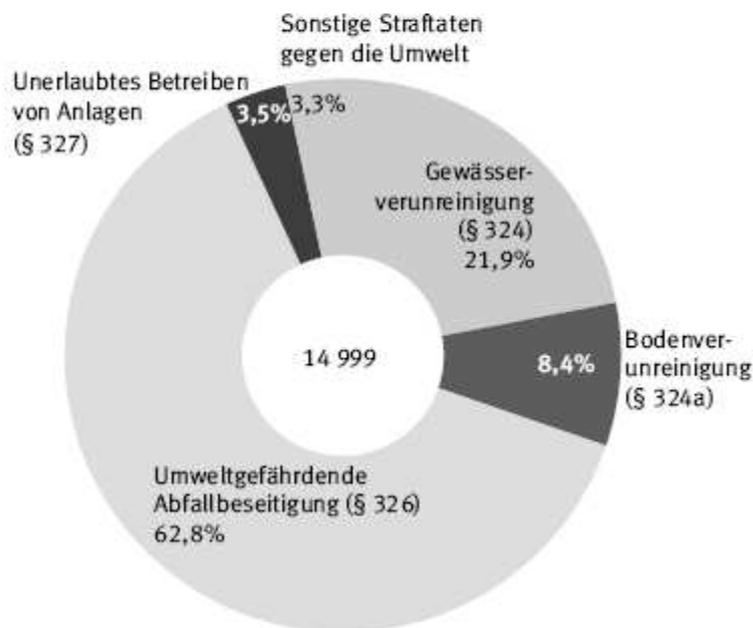
Schadstoffreduzierte Personenkraftwagen

Natur- und Landschaftsschutzrecht

Wahrgenommene Umweltbeeinträchtigungen und –gefährdungen

Zufriedenheit mit dem Zustand der Umwelt und Sorge um den Schutz der Umwelt

Umweltrelevante Einstellungen und Verhaltensweisen



1) Straftaten nach dem Strafgesetzbuch.

Abb. II.33b: Straftaten gegen die Umwelt 2008

Quelle: Statistisches Jahrbuch 2010

	Jahr	Gesamt- emission Mio t	Davon entfielen auf:			
			Haushalte	Kraft und Fernheiz- werke	Industrie	Verkehr
Schwefeldioxid SO <sub>2</sub>	1975	3,35	15,2	52,7	28,0	4,0
	1980	3,20	10,5	58,8	27,2	3,4
	1985	2,40	9,7	62,9	23,8	3,7
	1989	0,96	14,1	34,8	43,4	7,5
	1990	0,94	14,2	34,3	43,1	8,3
	1999	0,83	9,1	49,2	34,6	3,7
Stickoxid NO <sub>2</sub>	1975	2,55	5,4	26,0	17,0	51,7
	1980	2,95	4,9	27,2	13,5	54,5
	1985	2,95	4,7	26,0	10,2	59,1
	1989	2,70	4,0	18,0	9,7	68,3
	1990	2,60	4,2	12,9	9,7	73,1
	1999	1,64	5,0	15,3	13,8	63,9
Kohlenmonoxid CO	1975	14,00	10,3	0,3	16,9	72,6
	1980	12,00	9,3	0,4	16,9	73,4
	1985	8,90	11,3	0,5	17,2	71,0
	1989	8,25	8,5	0,5	17,3	73,8
	1990	8,20	8,5	0,6	16,3	74,8
	1999	4,95	16,1	2,1	23,9	56,4
Kohlendioxid CO <sub>2</sub>	1975	716,00	25,0	32,8	26,1	16,0
	1980	783,00	22,8	35,0	24,6	17,5
	1985	722,00	23,5	34,5	22,3	19,7
	1989	688,00	19,3	35,9	21,5	23,2
	1990	708,00	19,9	36,0	20,3	23,7
	1999	859,00	14,6	38,0	19,8	22,3
Staub	1975	0,81	11,9	19,6	40,5	7,5
	1980	0,69	9,3	18,5	37,7	9,3
	1985	0,58	8,7	15,5	32,7	12,2
	1989	0,46	6,8	5,1	33,2	15,8
	1990	0,45	6,6	5,1	31,8	16,9
	1999	0,26	12,0	7,7	40,7	20,8
Flüchtige organische Verbindungen	1975	2,80	3,9	0,6	11,9	43,0
	1980	2,75	3,2	0,7	7,5	47,6
	1985	2,60	3,4	0,5	5,4	48,4
	1989	2,55	2,6	0,5	5,3	50,2
	1990	2,55	2,6	0,5	5,3	50,4
	1999	1,65	3,2	0,4	8,9	23,5

Abb. II.33c: Gesamtemission ausgewählter Schadstoffe nach Verursacherguppen für das frühere Bundesgebiet

Quelle: Statistisches Jahrbuch 1993, S. 741-742 und Statistisches Jahrbuch 2002, S. 691

**Integrierte Volkswirtschaftliche und Umweltgesamtrechnung**

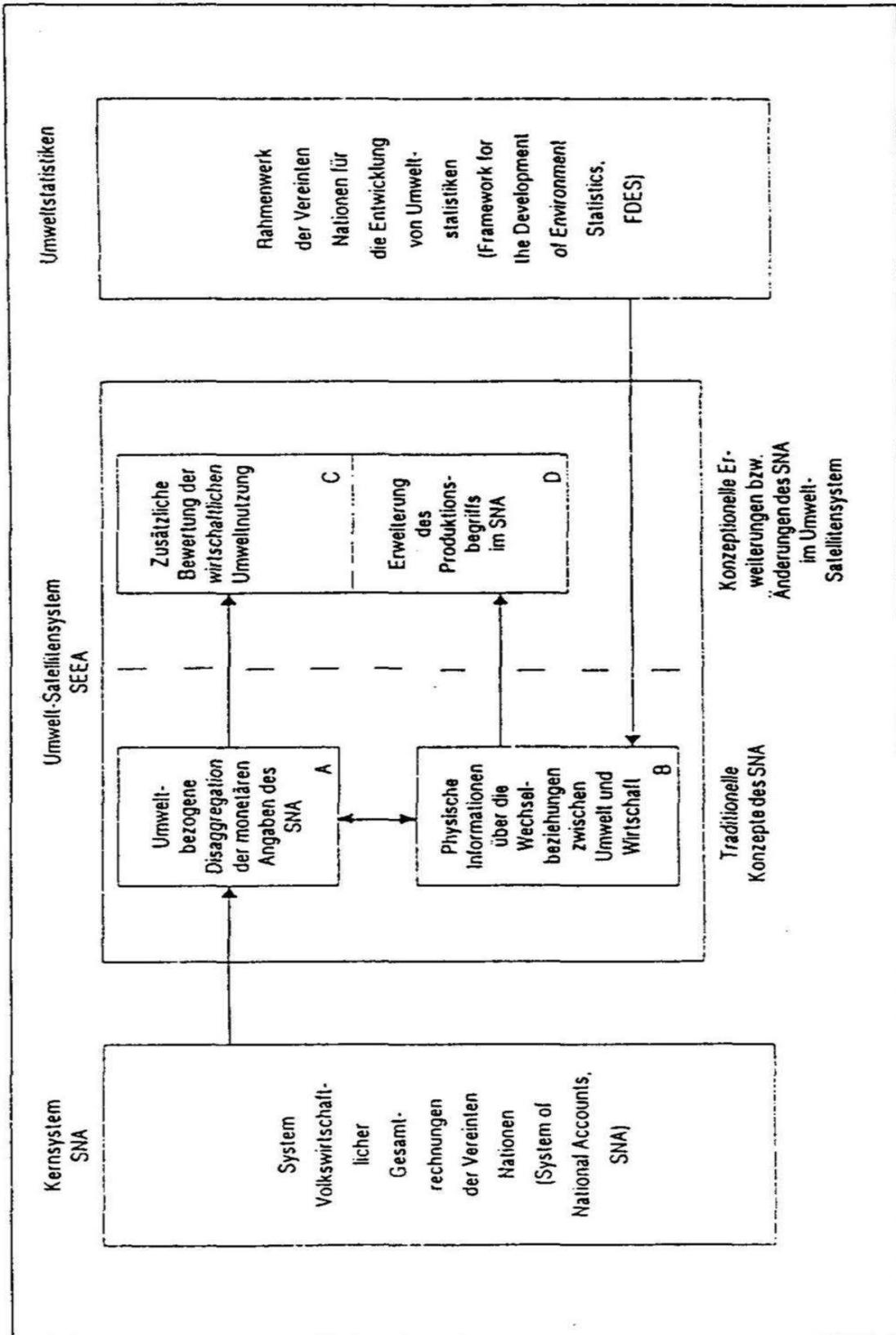
Mengen- und Wertgerüst (Statistisches Bundesamt: Stahmer (1992), Hamer und Stahmer (1992)), vgl. Abb. II.34.

Zum Umweltsatellitensystem zu den Volkswirtschaftlichen Gesamtrechnungen vergleiche Abb. II.34.

Internalisierung externer Kosten: Umweltverbrauch/-nutzung berücksichtigen (Hamer und Stahmer (1992), S. 111).

Neue Bewertung durch Ökoinlandsprodukt statt Bruttosozialprodukt.

- Produktionswert (Wert der produzierten Güter)
- Vorleistungen (Verbrauch von Gütern)
- = **Bruttoinlandsprodukt zu Marktpreisen**  
(Bruttowertschöpfung zu Marktpreisen)
- Abschreibungen (Wertminderung des reproduzierbaren Sachkapitals)
- = **Nettoinlandsprodukt zu Marktpreisen**  
(Nettowertschöpfung zu Marktpreisen)
- Wertminderung des nicht-produzierten Naturvermögens als Folge der wirtschaftlichen Aktivitäten
  - a) Quantitative Verminderung, z.B. der Bodenschätze und des Bestandes an Pflanzen und Tieren
  - b) Qualitative Verschlechterung der natürlichen Umwelt (Umweltmedien, Landschaft, Ökosysteme) durch Änderungen der Landschaftsnutzung und durch Abgabe von Rest- und Schadstoffen
- = **Ökoinlandsprodukt**



1) SEEA, System for Integrated Environmental and Economic Accounting (Integrierte Volkswirtschaftliche und Umweltgesamtrechnung). — A, B, C, D: Bausteine des SEEA.

Abb. II.34: SEEA als Umwelt-Satellitensystem zu den Volkswirtschaftlichen Gesamtrechnungen der Vereinten Nationen  
Quelle: Stahmer (1992), S. 579

## 8 Volkswirtschaftliche Gesamtrechnung, Soziale Indikatoren

Berichtssysteme für Volkswirtschaft und Gesellschaft

### 8.1 Volkswirtschaftliche Gesamtrechnung (VGR)

VGR = Gesamtwirtschaftliches Rechnungssystem

Zur Beobachtung und Analyse gesamtwirtschaftlicher Entwicklungen

Vorläufer: Kreislaufanalyse und Volkseinkommensberechnung, William Petty (1623-1687); 'Tableau économique', François Quesnay (1694-1774)

International

- System of National Accounts (SNA) der UN
- Material Product System (MPS) des RGW
- Europäisches System der Volkswirtschaftlichen Gesamtrechnungen (ESVG) der EU

Kreislaufdarstellung der VGR in

- Kontenform
- Matrizen-(Tabellen-)form
- Gleichungsform
- Kreislaufdiagramm

Für die Sektoren

- Unternehmen (Produkt, Kredit, Versicherung)
- Staat (Gebietskörperschaften, Sozialversicherung)
- Private Haushalte (Private Haushalte, Private Organisationen)

gibt es jeweils acht Konten:

- 0 Zusammengelegtes Güterkonto
- 1 Produktion
- 2 Einkommensentstehung
- 3 Einkommensverteilung
- 4 Einkommensumverteilung
- 5 Einkommensverwendung
- 6 Vermögensänderung
- 7 Finanzierung
- 8 Zusammengelegtes Konto der übrigen Welt

Zur Messung der Leistung einer Volkswirtschaft durch das Bruttoinlandsprodukt und das Bruttonsozialprodukt vgl. Abb. II.35.

Bruttoinlandsprodukt (BIP)		
geleistete Einkommen an die übrige Welt	In der Inlandsproduktion entstandene Einkommen von Ausländern	Empfangene Einkommen aus der übrigen Welt
Bruttoinländerprodukt oder Bruttonsozialprodukt (BSP)		

Abb. II.35: Bruttoinlands- (BIP) und Bruttonsozialprodukt (BSP)

Quelle: Rinne (1994), S. 439

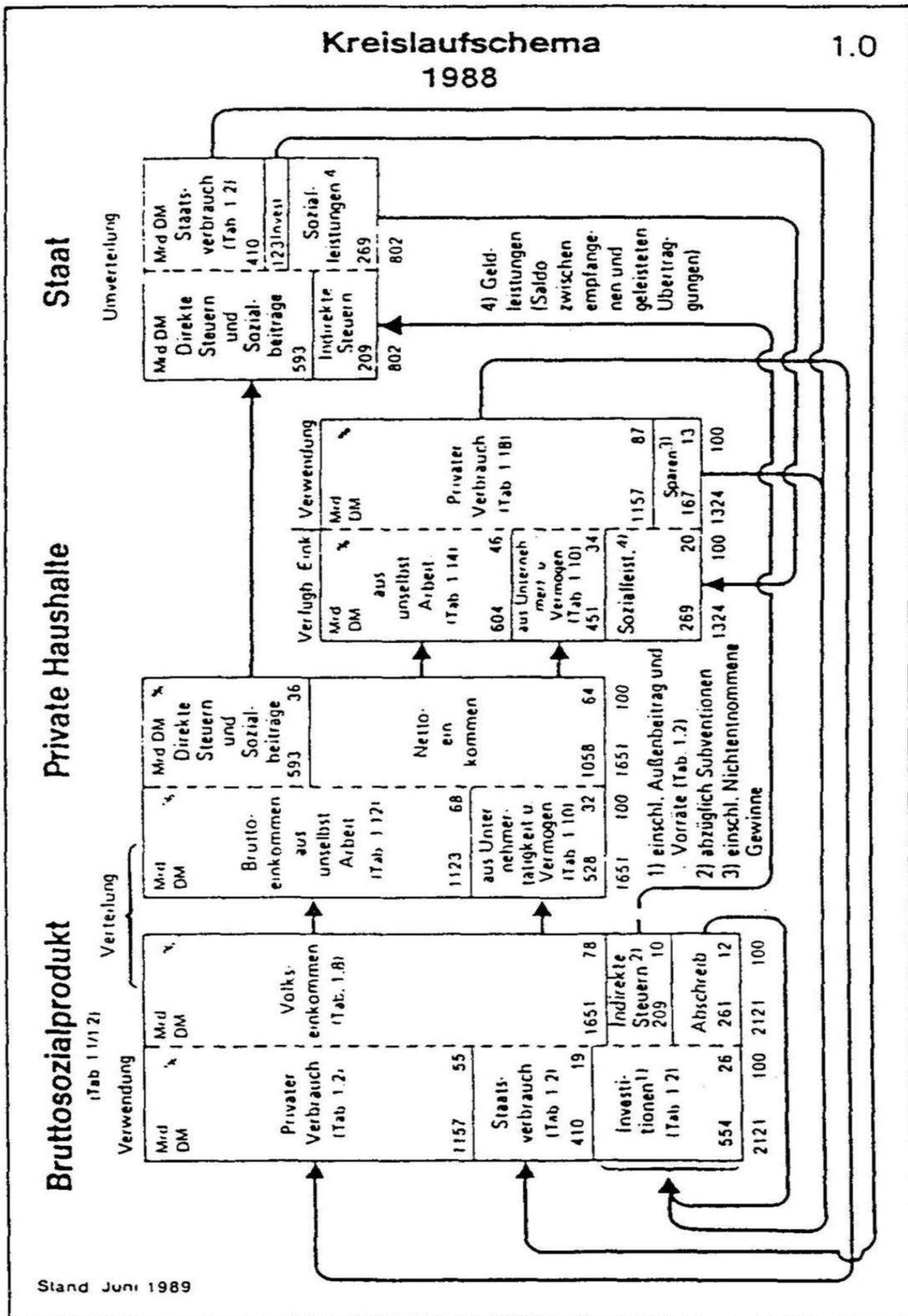


Abb. II.36: Kreislaufschema  
Quelle: Statistisches Taschenbuch 1989

Gegenstand der Nachweisung (Entstehung)	2005	2009
<b>Bruttowertschöpfung</b>	<b>2.027,52</b>	<b>2.150,80</b>
Land- und Forstwirtschaft und Fischerei	17,80	18,10
Produzierendes Gewerbe ohne Baugewerbe	523,62	473,80
Baugewerbe	78,10	98,20
Handel, Gastgewerbe und Verkehr	365,40	378,20
Finanzierung, Vermietung und Unternehmensdienstleister	590,20	666,50
Öffentliche und private Dienstleister	452,40	516,00
+ <b>Gütersteuern abzgl. Gütersubventionen</b>	<b>218,00</b>	<b>256,50</b>
= <b>Bruttoinlandsprodukt (BIP)</b>	<b>2.245,50</b>	<b>2.407,20</b>
Gegenstand der Nachweisung (Verteilung)	2005	2009
<b>Volkseinkommen</b>	<b>1.683,90</b>	<b>1.815,00</b>
Arbeitnehmerentgelt	1.128,80	1.223,90
Unternehmens- und Vermögenseinkommen	555,10	591,10
+ <b>Produktions- und Importabgaben an den Staat abzgl. Subventionen vom Staat</b>	<b>237,70</b>	<b>270,40</b>
+ <b>Abschreibungen</b>	<b>327,70</b>	<b>365,00</b>
- <b>Saldo der Primäreinkommen aus der übrigen Welt</b>	<b>3,80</b>	<b>43,20</b>
= <b>Bruttoinlandsprodukt (BIP)</b>	<b>2.245,50</b>	<b>2.407,20</b>
Gegenstand der Nachweisung (Verwendung)	2005	2009
<b>Konsum</b>	<b>1.746,91</b>	<b>1.887,81</b>
Privater Verbrauch	1.329,73	1414,69
Staatsverbrauch	417,18	473,12
+ <b>Bruttoinvestitionen</b>	<b>386,50</b>	<b>409,46</b>
Bruttoanlageinvestitionen	384,67	430,64
Anlagen	153,90	158,16
Bauten	205,55	245,16
Sonstige Anlagen	25,22	27,32
Vorratsänderungen	1,85	-21,18
= <b>Inländische Verwendung</b>	<b>2.133,43</b>	<b>2297,27</b>
+ <b>Außenbeitrag</b>	<b>112,07</b>	<b>109,93</b>
Exporte	901,69	982,25
Abzgl. Importe	789,62	872,32
= <b>Bruttoinlandsprodukt (BIP)</b>	<b>2.245,50</b>	<b>2.407,20</b>

Alle Angaben in Mrd. EUR

Abb. II.37: VGR - Entstehung, Verteilung und Verwendung des BIP  
Quelle: Statistisches Bundesamt, Statistisches Jahrbuch (2006, 2010)

Inlands- und Inländerprodukt werden durch

- Entstehungsrechnung
- Verteilungsrechnung
- Verwendungsrechnung

definiert (vgl. Abb. II.34).

VGR: Entstehungs- Verteilungs- und Verwendungsdaten sind für die BRD 2005 und 2009 in Abb. II.37 wiedergegeben.

Weitere Hinweise: Statistisches Bundesamt, Statistisches Jahrbuch (2010), Kap. 24

## 8.2 Soziale Indikatoren

Die Kritik am BSP (BIP) als eindimensionalem Wohlstandsmaß einer Gesellschaft (z.B. Leiper (1975, 1989)), in dem z.B. auch Schäden den Wohlstand vermehren, führte zu mehrdimensionalen Konzepten der nicht nur Wirtschafts- sondern umfassender: Sozialberichterstattung mit **sozialen Indikatoren**. Im Gegensatz zu der auf rein objektiven Kriterien beruhenden volkswirtschaftlichen Gesamtrechnung bedient sich die Sozialberichtserstattung einer Kombination objektiver Faktoren (z.B. finanzielle Ausstattung der Haushalte) und der Analyse subjektiven Empfindens (z.B. des 'Wohlfühlens' eines Individuums in seiner aktuellen Situation).

Eine Kurzübersicht über die hinter der Idee der Sozialberichtserstattung stehenden Konzepte (wie z.B. „sustainable development“) erhält man beispielsweise über die Unterabteilung „System Sozialer Indikatoren“ des Vereins GESIS- Leibniz-Institut für Sozialwissenschaften (<http://www.gesis.org/unser-angebot/daten-analysieren/soziale-indikatoren/system-sozialer-indikatoren/>)

Soziale Indikatoren für Bereiche gesellschaftspolitischer Relevanz zur (vgl. Zapf (1977)):

- Wohlfahrtsmessung
- Dauerbeobachtung des sozialen Wandels
- Steuerung und Prognose gesellschaftspolitischer Prozesse

Berichterstattung	amtlich		nicht-amtlich	
			Wissenschaftliche Einrichtungen	andere
<b>Allgemein: umfassend, bereichsübergreifend</b>	Gesellschaftl Daten; Datenreport Teil I; Sozialbericht des BMA.		System sozialer Indikatoren; Datenreport Teil II; Sozialreport; Recent Social Trends	
Speziell: für einzelne Lebensbereiche und/oder Bevölkerungsgruppen	Berufsbildungsbericht; Familienbericht; Daten zur Umwelt; Erster Altenbericht; Unfallverhütungsbericht; Rentenanpassungsbericht		Altenreport; Frauenreport	Armutsbericht DGB/Parit. Wohlfahrtsverband; Caritas-Armutsstudie; Shell-Jugendstudien

Abb. II.38 Klassifikation ausgewählter Sozialberichte in Deutschland; Quelle: Noll (1997), S. 10

**Internationale Systeme sozialer Indikatoren:**

- das System der EG (ab 1977)
- das System der UN
- das System der OECD

**In Deutschland:**

SPES-Projekt (Sozialpolitisches Entscheidungs- und Indikatorensystem für die Bundesrepublik Deutschland) der Sozialpolitischen Forschergruppe der Universitäten Frankfurt/Mannheim

**Zum SPES-Indikatorensystem vergleiche:**

- Zapf, W. (Hg.) (1977c), Lebensbedingungen in der Bundesrepublik Deutschland, Sozialer Wandel und Wohlfahrtsentwicklung, Frankfurt/New York
- Krupp, H.-J. und Zapf, W. (1977), Sozialpolitik und Sozialberichterstattung, Frankfurt/New York
- Wohlfahrtssurveys seit 1978 (SPES/Sfb 3/WZB)
- Sfb 3 (Sonderforschungsbereich 3 'Mikroanalytische Grundlagen der Gesellschaftspolitik') der Universitäten Frankfurt und Mannheim
- Glatzer, W. und Zapf, W. (Hg.) (1984), Lebensqualität in der Bundesrepublik Deutschland, Frankfurt/New York

Theoretisches Konzept der Wohlfahrtsproduktion mit objektiven und subjektiven Komponenten und Indikatoren.

Zentrale regelmäßige Sozialreports finden sich im Überblick in Abb. II.39.

**Systembereiche Sozialer Indikatoren:**

- Ausbildung/Lernen
- Beruf/Erwerbstätigkeit
- ökonomische Situation der Institutionen
- Gesundheit, Gesundheitswesen
- Physische Umwelt
- Persönliche Sicherheit und öffentliche Ordnung
- Zeit und Freizeit
- Soziale Sicherheit, Systeme der Sozialen Sicherheit
- Mobilität und soziale Schichtung

Für jeden dieser Systembereiche wird über ideale Indikatoren (mit Bewertungen aus Parteien, Verbände, öffentliche Meinung) für eine Vielzahl von Dimensionen schließlich ein operabler Indikator entwickelt (vgl. z.B. SPES-Indikatorentableau in Zapf (1977), Diewald (1984))

Land Herausgeber	Titel	Jahr der Veröffentlichung	Periodizität
<b>BR Deutschland</b> Statistisches Bundesamt 1985ff mit Sfb 3; 1992 mit WZB und ZUMA	Datenreport	1983	alle 2 Jahre
<b>Dänemark</b> Danmarks Statistics/ Socialforskningsinstituttet	Levevikar i Danmark	1976	alle 4 Jahre
<b>Frankreich</b> Institut Nationale de la Statistique et des Economiques	Données Sociales	1973	alle 3 Jahre
<b>Großbritannien</b> Central Statistical Office	Social Trends	1970	jährlich
<b>Italien</b> Istituto nazionale di Statistica	Vita Sociale Italiana	1990	regelmäßig
<b>Niederlande</b> Social and Cultural Planning Office	Social and Cultural Report	1974	alle 2 Jahre
<b>Österreich</b> Österreichs Statistisches Zentrallamt	Sozialstatistische Daten	1977	alle 4/5 Jahre
<b>Portugal</b> Instituto Nacional de Estatística	Portugal Social 1985-1990	1992	regelmäßig (offen)
<b>Schweden</b>	(verschiedene)		regelmäßig
<b>Spanien</b> Instituto Nacional de Estatística	Panorámica Social Indicadores Sociales	1974 1991	regelmäßig (offen)
<b>Ungarn</b> Tarki: Társadomkutatósi Informatikai Egyesülés	Social Report	1990	regelmäßig (offen)
<b>Internationale Comprehensive Reports</b>			
<b>EUROSTAT</b>	Social Indicators for the European Community Sozialportrait Europa	1977 1991	1980, 1984 alle 3 Jahre
<b>The World Bank</b>	World Development Report Social Indicators of Development	1978 1987	jährlich jährlich
<b>United Nations Develop- ment Programme</b>	Human Development Report	1990	jährlich

Abb. II.39: Zentrale regelmäßige Sozialreports

Quelle: Habich, R. und H.-H. Noll unter Mitarbeit von W. Zapf (1993), S. 73

Ausgewählte Indikatoren der Lebensbedingungen in den Ländern der Europäischen Union finden sich in Abb. II.40.

	Materieller Wohlstand		Bildung		Erwerbstätigkeit		Gesundheit		Sozial-schutz		Verkehr		Umwelt	
	BIP/Kopf <sup>1</sup>	Privat-Vertr./Kopf <sup>2</sup>	Ausgaben <sup>3</sup> in % des BIP	Beteiligung Sek. IF	Erw. quote <sup>4</sup> Frauen	Arb. losen- quote <sup>5</sup>	Säugl. sterblich-keit <sup>6</sup>	Ziga- retten/ <sup>7</sup> Kopf <sup>8</sup>	Aus- gaben/ <sup>9</sup> Kopf <sup>10</sup>	Eisen- bahn- netz <sup>11</sup>	Ver- kehrs- opter <sup>12</sup>	Emissionen	CO <sub>2</sub> <sup>12</sup>	SO <sub>2</sub> <sup>13</sup>
	KKS	KKS	%	%	%	%	Zahl/ <sup>100.000</sup>	jährl. Zahl	KKS	km	Zahl/ <sup>100.000</sup>	kg/Kopf	1997	1997
A	17032	9539	6,3	68,3	62,7	3,9	4,4	2323	5803	67,2	644	7,4	7,4	6,0
B	16676	10647	5,2	84,2	56,0	8,8	5,3	1472	5343	110,0	706	11,4	11,4	21,6
D	16170	9131	4,6	57,3	62,9	8,6	4,6	2256	5715	106,8	616	10,1	10,1	16,6
DK	17911	9367	8,3	61,6	76,1	5,2	4,2	2003	7390	51,9	182	12,1	12,1	20,8
E	12585	7497	4,5	71,6	48,5	15,9	4,9	2678	3206	24,4	374	6,2	6,2	36,2
F	18946	10062	6,0	59,9	62,2	11,2	4,8	1712	6205	58,3	301	6,1	6,1	13,0
FIN	15545	7724	6,2	79,6	71,2	10,2	3,6	1139	5948	17,4	184	11,4	11,4	19,3
GR	9297	6925	3,5	56,6	48,5	11,6	5,9	3355	2973	19,0	339	7,9	7,9	50,6
I	15834	9644	4,9	62,0	45,6	11,3	5,1	1888	5503	53,3	521	7,0	7,0	17,8
IRL	16274	8031	4,5	47,7	54,3	5,6	5,5	2355	3702	27,7	357	9,9	9,9	45,4
L	25795	14014	4,3	37,6	50,4	2,4	4,7	2621	8146	91,3	369	20,3	20,3	14,3
NL	16785	9616	4,9	58,4	64,4	3,4	5,2	1296	5451	68,5	314	10,9	10,9	7,6
P	10295	6739	5,7	50,4	63,0	4,5	5,4	2007	2641	30,4	667	4,8	4,8	33,6
S	16089	7947	8,0	83,6	74,6	7,2	2,9	867	6555	27,1	247	5,8	5,8	5,8
UK	15855	10281	4,9	88,2	67,3	6,1	5,8	1674	4196	69,6	572	9,0	9,0	27,9

Anmerkungen und Datenquellen: (1) Bruttoinlandsprodukt pro Kopf in Preisen und Kaufkraftstandards von 1990; Europäische Kommission, Volkswirtschaftliche Gesamtrechnung ESWS, Aggregatdaten 1970-1997; (2) Private Verbrauchsausgaben pro Kopf in Preisen und Kaufkraftstandards von 1990; Statistisches Bundesamt, Statistisches Jahrbuch für das Ausland 1999; (3) L: 1997; OECD, Education at a Glance; (4) Schüler/Studenten der Sekundarstufe II oder höher als Anteil der 15-24jährigen; Europäische Kommission, Statistisches Jahrbuch; (5) GR: 1996; Eurostat, New Cronos; (6) Eurostat, Structural Indicators; (7) Anzahl der Lebendgeborenen, die vor Erreichung des 1. Lebensjahres sterben, pro 1000 Lebendgeborene; Europäische Kommission, Statistisches Jahrbuch; (8) Jährlicher Zigarettenkonsum pro 15jähriger und älterer Person; L: 1994, NL: 1996; WHO, Health for all Database; (9) Öffentliche Sozialausgaben in Preisen und Kaufkraftstandards von 1995; OECD Social Expenditure Database; (10) Länge des Schienennetzes in Kilometer/1000 Quadratkilometer; Europäische Kommission, Statistisches Jahrbuch; (11) Bei Verkehrsunfällen verletzte Personen/100.000 Einwohner, NL: 1997; WHO, Health for all Database; (12) Europäische Kommission, Statistisches Jahrbuch; (13) E. P.: 1996; WHO, Health for all Database.

Abb. 11.40: Ausgewählte Indikatoren der Lebensbedingungen in den Ländern der Europäischen Union

Quelle: Berger-Schmitt (2002), S. 2

Weitere Informationen zu sozialen Indikatoren: Habich und Noll unter Mitarbeit von Zapf (1993), Noll (1990).

Aus disaggregierten Informationen wieder aggregierter Index:

z.B. **Human-Development Index der UN**, berechnet aus drei Berichtsindikatoren:

- Brutto-Inlandsprodukt pro Kopf
- Lebenserwartung
- Analphabetismusrate bzw. Anzahl der Schuljahre

(vgl. United Nations Development Program (UNDP), Human Development Report, lfd., Oxford)

### **Auszug aus dem Human Development Report 2002 (S. 34f):**

“The human development index (HDI) is a simple summary measure of three dimensions of the human development concept: living a long and healthy life, being educated and having a decent standard of living (see technical note). Thus it combines measures of life expectancy, school enrolment, literacy and income to allow a broader view of a country’s development than using income alone – which is too often equated with well-being. Since the creation of the HDI in 1990 three supplementary indices have been developed to highlight particular aspects of human development: the human poverty index (HPI), gender-related development index (GDI) and gender empowerment measure (GEM). [...]

#### **Human poverty index**

While the HDI measures overall progress in a country in achieving human development, the human poverty index (HPI) reflect the distribution of progress and measures the backlog of deprivations that still exists. The HPI measures deprivation in the same dimensions of basic human development as the HDI.

#### *HPI-1*

The HPI-1 measures poverty in developing countries. It focuses on deprivations in three dimensions: longevity, as measured by the probability at birth of not surviving to age 40; knowledge, as measured by the adult illiteracy rate; and overall economic provisioning, public and private, as measured by the percentage of people not using improved water sources and the percentage of children under five who are underweight.

#### *HPI-2*

Because human deprivation varies with the social and economic conditions of a community, a separate index, the HPI-2, has been devised to measure human poverty in selected OECD countries, drawing on the greater availability of data. The HPI-2 focuses on deprivation in the3 same three dimensions as the HPI-1 and one additional one, social exclusion. The indicators are the probability at birth of not surviving to age 60, the adult functional illiteracy rate, the percentage of people living below the income poverty line (with disposable income household income less than 50% of the median) and the long-term unemployment rate (12 months or more).

#### **Gender-related development index**

The gender-related development index (GDI) measures achievements in the same dimensions and using the same indicators as the HDI, but captures inequalities in achievement between women and men, It is simply the HDI adjusted downward for gender inequality. The greater is the gender disparity in basic human development, the lower is a country’s GDI compared with its HDI.

#### **Gender empowerment measure**

The gender empowerment measure (GEM) reveals whether women can takre active part in economic and political. It focuses on participation, measuring gender inequality in keay areas of ecomical and political participation and decision-making. It tracks the percentages of women in parliament, among legislators, senior officials and managers and among professional and technical workers – and the gender disparity in earned income, reflecting economic independence. Differing from the GDI, it exposes inequality in opportunities in selected areas.”

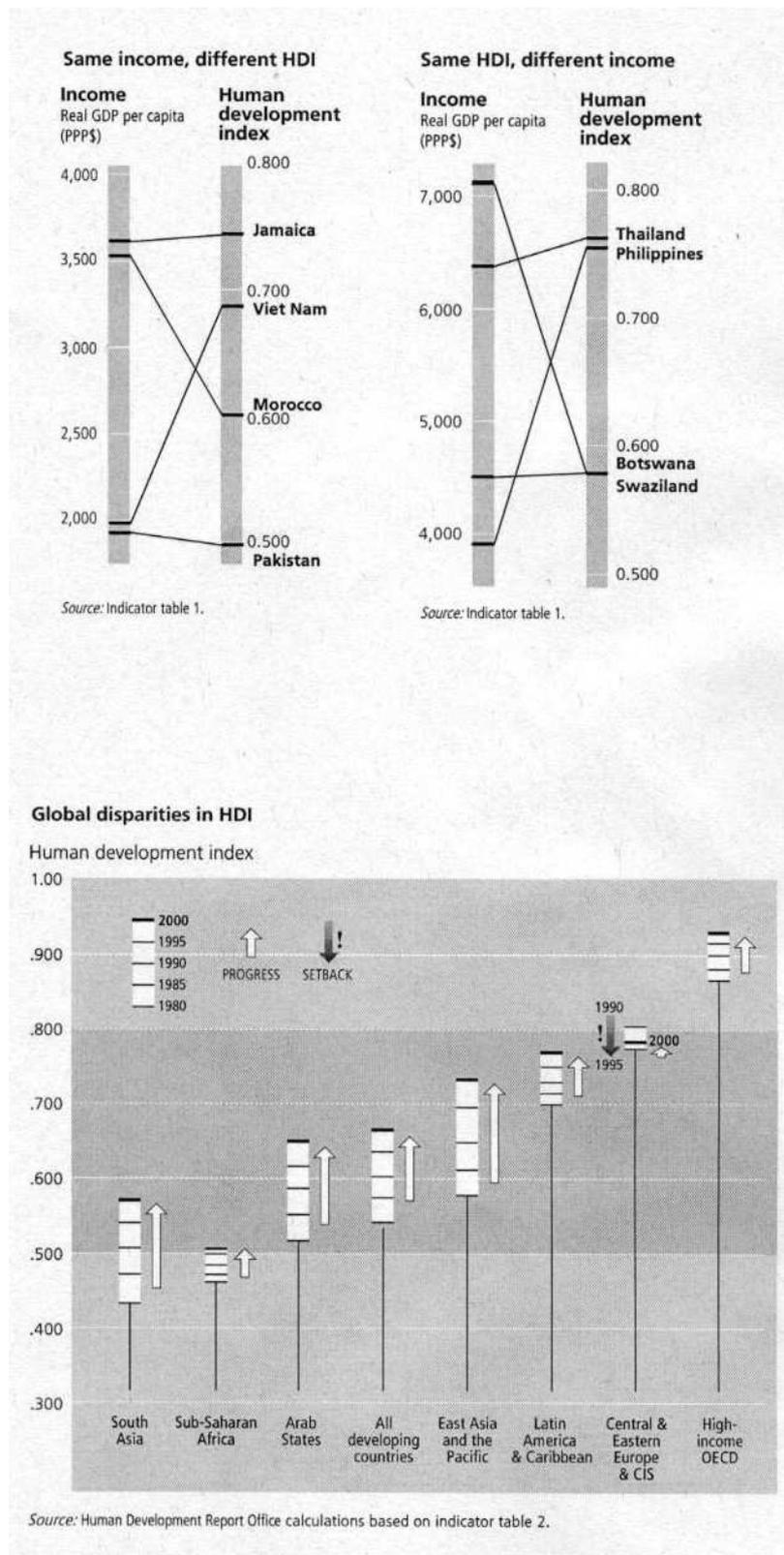


Abb. II.41: Human Development Index im Vergleich  
 Quelle: Human Development Report 2002, S. 35

	Human Development Index (HDI) value 2010	Life expectancy at birth (years) 2010	Mean years of schooling (years) 2010	Expected years of schooling (years) 2010	Real GDP per capita (PPP 2008 \$) 2010	GNI per capita rank minus HDI rank	Nonincome HDI value	GNI per capita rank minus HDI rank
1 Norway	0,938	81,0	12,6	17,3	58.810	2	0,954	2
2 Australia	0,937	81,9	12,0	20,5	38.692	11	0,989	11
3 New Zealand	0,907	80,6	12,5	19,7	25.438	30	0,979	30
4 United States	0,902	79,6	12,4	15,7	47.094	5	0,917	5
5 Ireland	0,895	80,3	11,6	17,9	33.078	20	0,936	20
6 Liechtenstein	0,891	79,6	10,3	14,8	81.011	-5	0,861	-5
7 Netherlands	0,890	80,3	11,2	16,7	40.658	4	0,911	4
8 Canada	0,888	81,0	11,5	16,0	38.668	6	0,913	6
9 Sweden	0,885	81,3	11,6	15,6	36.936	8	0,911	8
10 Germany	0,885	80,2	12,2	15,6	35.308	9	0,915	9
11 Japan	0,884	83,2	11,5	15,1	34.692	11	0,915	11
12 Korea, Republic of	0,877	79,8	11,6	16,8	29.518	16	0,918	16
13 Switzerland	0,874	82,2	10,3	15,5	39.849	-1	0,889	-1
14 France	0,872	81,6	10,4	16,1	34.341	9	0,898	9
15 Israel	0,872	81,2	11,9	15,6	27.831	14	0,916	14
16 Finland	0,871	80,1	10,3	17,1	33.872	8	0,897	8
17 Iceland	0,869	82,1	10,4	18,2	22.917	20	0,928	20
18 Belgium	0,867	80,3	10,6	15,9	34.873	3	0,888	3
19 Denmark	0,866	78,7	10,3	16,9	36.404	-1	0,883	-1
20 Spain	0,863	81,3	10,4	16,4	29.661	6	0,897	6
21 Hong Kong, China	0,862	82,5	10,0	13,8	45.090	-11	0,860	-11
22 Greece	0,855	79,7	10,5	16,5	27.580	8	0,890	8
23 Italy	0,854	81,4	9,7	16,3	29.619	4	0,882	4
24 Luxembourg	0,852	79,9	10,1	13,3	51.109	-18	0,836	-18
25 Austria	0,851	80,4	9,8	15,0	37.056	-9	0,859	-9
.								
.								
.								
162 Liberia	0.300	59.1	3.9	11.0	320	5	0,509	-38
163 Chad	0.295	49.2	1.5	6.0	1.067	-9	0,298	-11
164 Guinea-Bissau	0.289	48.6	2.3	9.1	538	1	0,362	-19
165 Mozambique	0.284	48.4	1.2	8.2	854	-5	0,300	-15
167 Burundi	0.282	51.4	2.7	9.6	402	0	0,400	-11
168 Niger	0.261	52.5	1.4	4.3	675	-3	0,285	-39
169 Congo, DR of the	0.239	48.0	3.8	7.8	291	0	0,390	-14
170 Zimbabwe	0.140	47.0	7.2	9.2	176	0	0,472	-25

Abb. II.42: Human development index: Die ersten 25 und letzten acht Länder im Vergleich  
Quelle: Human Development Report 2010, S. 143-146

***Keyconcepts***

*Bevölkerungsstatistik*

*Volkszählung*

*Mikrozensus*

*Produktion*

*Dienstleistung*

*Erwerbstätigkeit*

*Erwerbskonzept*

*Einkommens- und Verbrauchsstatistik*

*Preise*

*Umwelt*

*Volkswirtschaftliche Gesamtrechnung*

*Soziale Indikatoren*

*Human Development Report*



### III Statistische Analyse eines einzelnen Merkmals



Informationskomprimierung mit Hilfe von Häufigkeitsverteilungen und Kennzahlen zur Lage, Streuung und Konzentration eines einzelnen Merkmals.

Die statistische Analyse *eines* Merkmals konzentriert sich auf die Analyse/Beschreibung *einer* Dimension von Merkmalsträgern (z.B. Einkommen von Haushalten, Umsätze mehrerer Geschäftsvorgänge/Filialen/Jahre)

Komprimierende Beschreibung über:

- Häufigkeitsverteilung
- Lageparameter
- Streuungsmaße bzw. -parameter
- Konzentration der Verteilung

#### 1 Eindimensionale Häufigkeitsverteilungen und ihre Darstellung

Von Interesse:     - ein Merkmal einer statistischen Masse  
                           - beobachtete Häufigkeiten der Merkmalsausprägungen

Darstellungsformen:                     Tabellen, Grafiken

Es ist zu unterscheiden zwischen qualitativen und quantitativen Größen.

##### 1.1 Häufigkeitsverteilung nominalskaliert (qualitativer) Merkmale

Statistische Masse mit

- $n$              statistische Einheiten
- $A$              qualitatives Merkmal
- $A_i$            *i*-te Merkmalsausprägung ( $i = 1, \dots, k$ ) mit
- $n(A_i) = n_i$    absolute Häufigkeit des Merkmals in der Klasse  $i$

##### Absolute Häufigkeit

Anzahl der statistischen Einheiten der Klasse  $i$ :  $n(A_i) = n_i$

$$\text{Es gilt: } \sum_{i=1}^k n(A_i) = \sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = n$$

Für Vergleichszwecke geeignet:

**Relative Häufigkeit**

$$h(A_i) = h_i = \frac{n_i}{n} = \frac{n(A_i)}{n}$$

$$h_i = \frac{\text{Absolute Häufigkeit der Klasse } i}{\text{Umfang der statistischen Masse}}$$

$$\text{Es gilt: } \sum_{i=1}^k h(A_i) = \sum_{i=1}^k h_i = 1$$

**Häufigkeitstabelle:** Tabellarische Darstellung

Tab. III.1: Häufigkeitstabelle: Allgemeiner Aufbau

Merkmalsausprägung	Absolute Häufigkeit	relative Häufigkeit
$A_1$	$n_1$	$h_1$
$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$
$A_i$	$n_i$	$h_i$
$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$	$\cdot$
$A_k$	$N_k$	$h_k$
Summe	$N$	1

**Beispiel:**

Erwerbstätige untergliedert nach der Stellung im Beruf

Tab. III.2: Erwerbstätige nach der Stellung im Beruf in der BRD 1987 (in 100.000)

Ausprägung $A_i$ des Merkmals 'Stellung im Beruf'	Absolute Häufigkeit $n(A_i) = n_i$	Relative Häufigkeit $h(A_i) = \frac{n_i}{n}$	bzw. [%]
Selbständige	( $A_1$ ) 23	0,085	8,5
Mithelfende Familien- angehörige	( $A_2$ ) 5	0,018	1,8
Beamte	( $A_3$ ) 24	0,091	9,1
Angestellte/Auszubildende (kfm./techn.)	( $A_4$ ) 110	0,410	41,0
Arbeiter/Auszubildende (gewerblich)	( $A_5$ ) 107	0,396	39,6
Insgesamt	269 $\left(\sum_{i=1}^5 n_i\right)$	1,000	100,0

Quelle: Statistisches Bundesamt 1989, Volkszählung 1987

In der BRD waren 1987 somit von den insgesamt  $269 \cdot (100.000) = 26,9$  Mio. Erwerbstätigen 8,5 % Selbständige und 39,6 % (10,7 Mio.) Arbeiter.

### Häufigkeitsverteilung

Die **Häufigkeitsverteilung qualitativer Merkmale** heißt die Funktion  $h(A_i)$ , die jeder Merkmalsausprägung  $A_i$  den Anteil der statistischen Einheiten mit dieser Merkmalsausprägung (relative Häufigkeit)

$$h(A_i) = \frac{n(A_i)}{n} = \frac{n_i}{n} \quad (i = 1, \dots, k)$$

zuordnet. Die Häufigkeitsverteilung ist also die Gesamtheit der relativen Häufigkeiten.

### Grafische Darstellung

- Kreisdiagramm ('pie-chart'): Die Kreisfläche wird in entsprechende Anteile aufgeteilt.

Hinweis zur Berechnung der Anteile an der Kreisfläche:  $\sum_{i=1}^k h(A_i) = 100 \% = 360^\circ$ .

- Balkendiagramme

### Beispiele:

- **Kreisdiagramm**

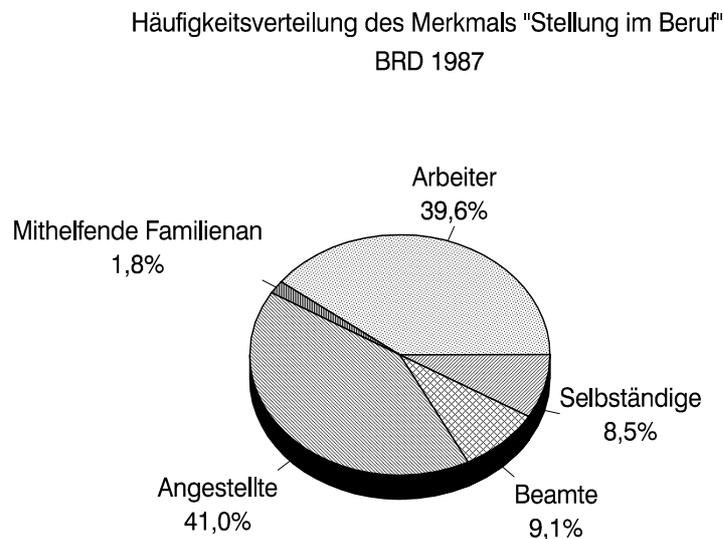


Abb. III.1: Häufigkeitsverteilung des Merkmals 'Stellung im Beruf' in der BRD 1987

Quelle: Statistisches Bundesamt 1989, Volkszählung 1987

- **Balkendiagramm** ('bar-chart'): Balken proportional zu den entsprechenden Einheiten

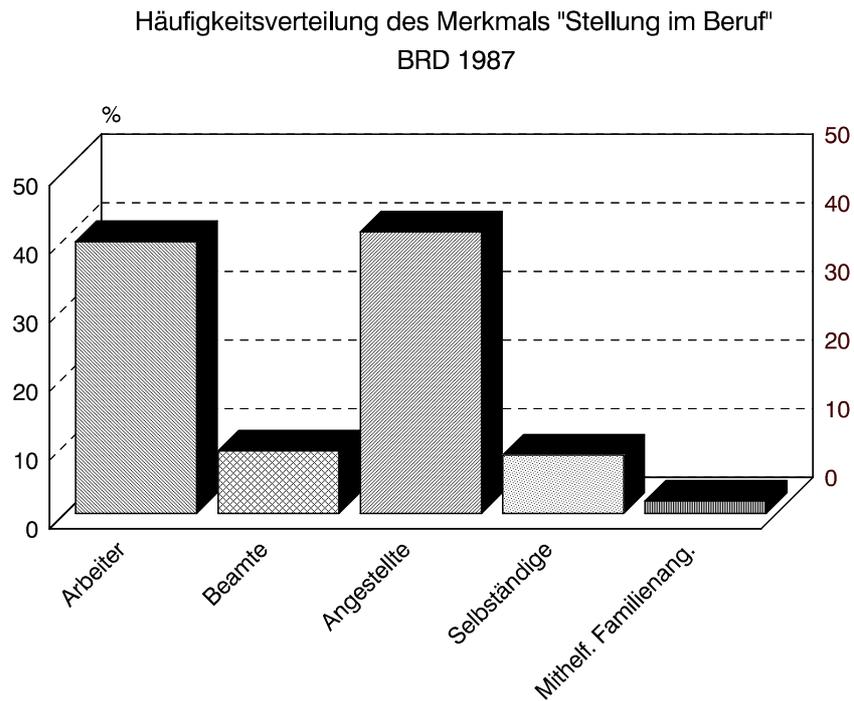
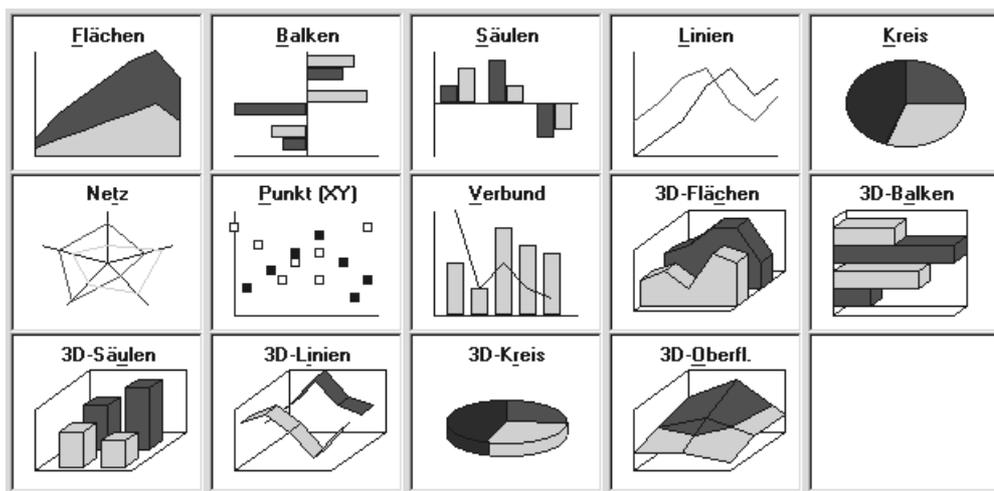


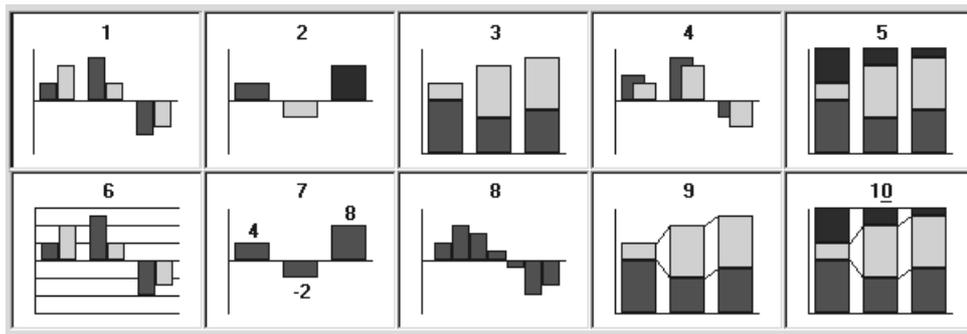
Abb. III.2: Häufigkeitsverteilung des Merkmals 'Stellung im Beruf' in der BRD 1987  
Quelle: Statistisches Bundesamt 1989, Volkszählung 1987

- **Grafische Darstellungsmöglichkeiten in einem Tabellenkalkulationsprogramm z.B. Microsoft Excel**

Hauptmenü:



## Untermenü Säulen:



- 1 Einfaches Säulendiagramm
- 2 Säulendiagramm für eine Datenreihe mit unterschiedlichen Mustern
- 3 Gestapelt
- 4 Überlappend
- 5 100 % gestapelt
- 6 Mit horizontalen Gitternetzlinien
- 7 Mit Wertebeschriftungen
- 8 Stufendiagramm (Rubriken ohne Zwischenraum)
- 9 Gestapelt mit Linien, die die Daten in derselben Datenreihe verbinden
- 10 100 % gestapelt mit Linien, die die Daten in derselben Datenreihe verbinden

Abb. III.3: Grafische Darstellungsmöglichkeiten in dem Tabellenkalkulationsprogramm Microsoft Excel

- **ET, Econometrics Toolkit**

z.B.: 20 Studentinnen/Studenten werden nach der Augenfarbe befragt.

Jede Beobachtung (= Proband), jede statistische Einheit erhält für das Merkmal Augenfarbe eine Merkmalsausprägung (Code):

Codes:     1 = blaue Augen  
               2 = grüne Augen  
               3 = rote Augen  
               4 = gelbe Augen

- ET: 1 Data entry and manipulation (Main Menu)
- Data (Read or edit data)  
*variable name (eyes)* ↵  
*input of data*
  - 5 Histograms, plots, descriptive statistics
  - Histogram for individual or frequencies  
*variable name (eyes)* ↵

Main Menu	
1 Data entry and mani	
2 Current sample and	
3 Management and disp	
4 File system and out	
5 Histograms, plots,	Histograms, plots, desc. stats
6 Regression model es	
7 Probability,Matrice	
8 Tests: t, F, fits,	
9 Editor for text and	
System: Give DOS comm	Command Keys Function
Graphics: set screen	Describe D 1 Descriptive statistics
QUIT: leave ET (Save	Histogram H 2 Histogram for ind. or freq
Data: Rows= 200 ( 1	Scatter P 3 Plot variables in scatter diag.
Cols=100,Observations	Identify B 4 Box-Jenkins time series ID
	ARIMA A 5 ARIMA and ARMAX time series
	Stepwise S 6 Stepwise linear regression
	Crosstab C 7 Cross tabulation for 2 vars.
	XCorrel X 8 Cross correlation, time series
	Give command by letter or #, or ↑/↓ and ↵.
	Use PgUp and PgDn for other command groups.
	Press ESC to clear and return to main menu.

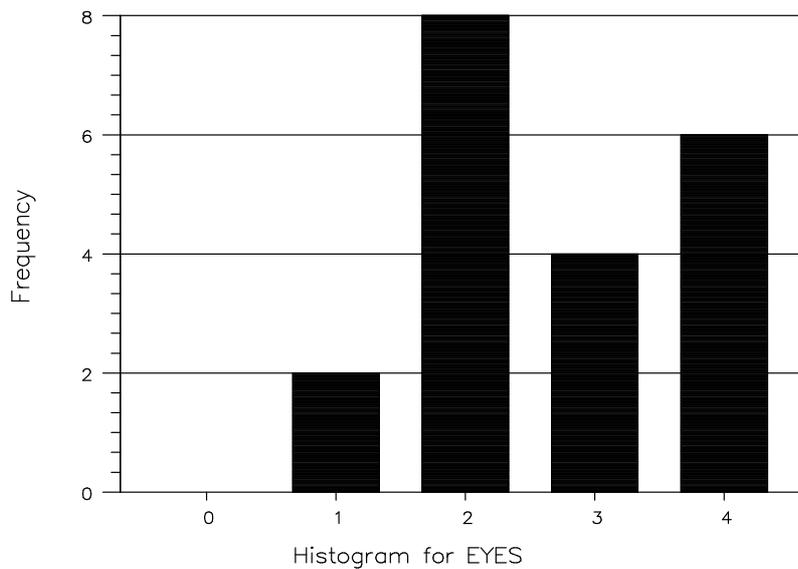
F1=HELP F2=Variables F3=Namelists F4=Matrices F5=Scalars F6=List all  
 F7=Output(Full/Basic)B F8=Option MenusN F9=Color/Mono F10=Mode(Menu/Command) M

DATA LISTING (Current sample) ↵/ESC Press ESC to interrupt list.

Observation	EYES
1	4.0000
2	2.0000
3	3.0000
4	1.0000
5	3.0000
6	3.0000
7	2.0000
8	2.0000
9	2.0000
10	4.0000
11	4.0000
12	2.0000
13	4.0000
14	1.0000
15	2.0000
16	2.0000
17	4.0000
18	2.0000
19	3.0000
20	4.0000

Histogram for EYES computed using 20 observations  
 Obs. out of range: too low= 0, too high= 0  
 Individual data Mean= 2.700, std.dev.= 1.031

	Frequency		Cumulative	
	Lower Limit	Upper Limit	Total	Relative
0	-.500	.500	0	.0000
1	.500	1.500	2	.1000
2	1.500	2.500	8	.4000
3	2.500	3.500	4	.2000
4	3.500	4.500	6	.3000



### - Verbale Häufigkeitsdarstellung:

"Wenn wir für eine Minute schweigen sollten für jeden Menschen, der 1982 an Hunger starb, wären wir nicht in der Lage, den Beginn des 21. Jahrhunderts zu feiern, weil wir dann immer noch still sein müßten."

Kubas Staatspräsident Fidel Castro, 1983

## 1.2 Häufigkeitsverteilung metrisch skaliertes, diskreter Merkmale

Ein metrisch skaliertes (quantitatives), diskretes Merkmal nimmt nur bestimmte Zahlenwerte (aus den reellen Zahlen) an: meist nichtnegative ganze Zahlen: 0,1,2,...; z.B.:

- Anzahl der Personen in einem Haushalt
- Anzahl der Verkäufe eines bestimmten Produktes
- Anzahl der Räume in privaten Wohnungen

$x_i$  Merkmalswert für die  $i$ -te Merkmalsausprägung des Merkmals  $x$

Wie für nominalskalierte Größen erhält man durch Auszählen der jeweiligen statistischen Einheiten die:

### Absolute Häufigkeit

Anzahl der statistischen Einheiten mit dem Merkmalswert  $x_i$ :  $n(x_i) = n_i$

### Relative Häufigkeit

$$h(x_i) = \frac{n_i}{n} = \frac{n(x_i)}{n}$$

$h(x_i)$  kann als diskrete Funktion aufgefaßt werden:

**Häufigkeitsfunktion**  $h(x) = h(x_i)$  mit  $i = 1, \dots, k$  Merkmalsausprägungen

Bei metrisch skalierten Merkmalen mißt die Differenz zweier Merkmalswerte ihren Abstand. Daher: Berechnung von Anteilen (relative Häufigkeit) für mehrere Merkmalswerte.

Häufig: Berechnung des Anteilwertes für Merkmalswert kleiner oder gleich  $x_i$ :  
kumulierte Häufigkeiten

### Kumulierte absolute Häufigkeit

$$n(x \leq x_i) = n_1 + n_2 + \dots + n_i = \sum_{j=1}^i n_j$$

### Kumulierte relative Häufigkeit

$$h(x \leq x_i) = h(x_1) + h(x_2) + \dots + h(x_i) = \sum_{j=1}^i h(x_j)$$

### Verteilungsfunktion

Die Verteilungsfunktion ist die Funktion  $F(x_i)$ , die jedem Merkmalswert  $x_i$  den Anteilwert aller statistischen Einheiten zuordnet, die einen Merkmalswert kleiner oder gleich  $x_i$  ( $x \leq x_i$ ) haben.

$$F(x_i) = h(x \leq x_i) = \sum_{j=1}^i h(x_j) = \sum_{j=1}^i \frac{n_j}{n}$$

**Beispiel:** \_\_\_\_\_

Größe der Privathaushalte in der BRD 1987

---

### Tabellarische Darstellung: Häufigkeitstabelle

Tab. III.3: Größe der Privathaushalte in der BRD 1987 (in 100.000)

Anzahl der Personen $x_i$	Absolute Häufigkeit $n(x_i) = n_i$	Relative Häufigkeit $h(x_i) = \frac{n_i}{n}$	Verteilungsfunktion $F(x_i)$
$x_1 = 1$	88	0,34	0,34
$x_2 = 2$	74	0,28	0,62
$x_3 = 3$	46	0,18	0,80
$x_4 = 4$ und mehr	54	0,20	1,00
Insgesamt	262	1,00	

Quelle: Statistisches Bundesamt 1989, Volkszählung 1987

Aus der Häufigkeitsverteilung ergibt sich z.B., dass 18 % aller Privathaushalte aus drei Personen bestehen.

Aus der Verteilungsfunktion ergibt sich z.B., dass 80 % aller Privathaushalte drei oder weniger Personen hatten.

### Grafische Darstellung

Häufigkeitsfunktion: als Stabdiagramm (es gibt keine Zwischenwerte); Stablänge = relative Häufigkeit

Verteilungsfunktion: als Treppenfunktion (Summe der bisherigen Stablängen)

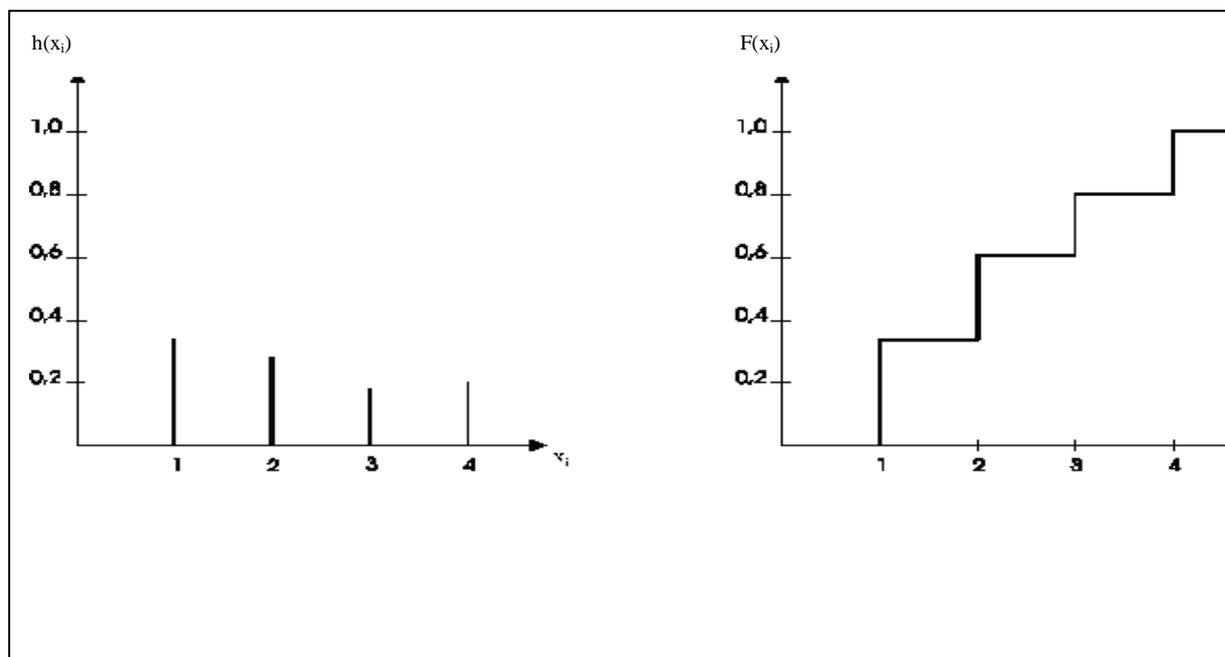


Abb. III.4: Häufigkeitsfunktion und Verteilungsfunktion: Grafische Darstellung der Größe der Privathaushalte in der BRD 1987 (vgl. Tab. III.3)

Wie groß ist der Anteil der Privathaushalte mit mehr als 1 aber weniger als 4 Personen?

$$\begin{aligned} h(1 < x < 4) &= h(x \leq 3) - h(x = 1) \\ &= F(3) - F(1) \\ &= 0,80 - 0,34 = 0,46 \end{aligned}$$

46 % aller Haushalte sind 2- oder 3-Personenhaushalte.

### 1.3 Häufigkeitsverteilung metrisch skaliertes (quantitativer) stetiger Merkmale

Merkmalswerte können sich in allen reellen Zahlen ausprägen. Da jeder Merkmalswert in der Regel nur einmal beobachtet wird (bei beliebig genauer Messung) ist zur Darstellung erst eine **Klassenbildung** notwendig.

#### Klasseneinteilung

- Möglichst gleiche Klassenbreiten, bei großem Variationsbereich der Daten auch unterschiedliche Klassenbreiten verwenden;
- Anzahl der Klassen ( $k$ ) nicht zu groß ( $\sqrt[3]{n}, \dots, \sqrt[2]{n}$  ( $n = 100, 5 - 10$  Klassen));
- Der häufigste Wert der Urliste sollte die Klassenmitte der Klasse mit der größten Häufigkeit bilden;
- Für einen Vergleich mit anderen Verteilungen: gleiche Klassen bilden

Generelles Ziel: Struktur des Ausgangsmaterials klar und unverfälscht herausarbeiten!

Nach Klasseneinteilung: Auszählen der Merkmalswerte je Klasse ergibt Häufigkeiten für die einzelnen Klassen

#### Häufigkeitsverteilung metrisch skaliertes, stetiger Merkmale

ist die Funktion

$$h(x_i) = h_i = h(x_i^u \leq x < x_i^o),$$

die jeder Klasse  $i$  ( $i = 1, \dots, k$ ) eine relative Häufigkeit  $\frac{n_i}{n}$  zuordnet, wobei

$x_i^u$  Untergrenze der Klasse  $i$ ,  $x_i^o$  Obergrenze der Klasse  $i$

$n_i$  Zahl der beobachteten Merkmalswerte im Intervall  $[x_i^u, x_i^o]$

Die klassierte Häufigkeitsverteilung wird mit Rechtecken als **Histogramm** grafisch dargestellt. Damit die Rechteckflächen auch bei unterschiedlichen Klassenbreiten proportional den (relativen) Häufigkeiten sind, werden die relativen Häufigkeiten über die Dichtefunktion (= Höhe der Rechtecke) normiert.

**Dichtefunktion**

normierte relative Häufigkeiten (vor allem für unterschiedliche Klassenbreiten)

$$f(x_i) = \frac{n_i}{n \cdot \Delta x_i}, \text{ wobei } \Delta x_i = \text{Breite der } i\text{-ten Klasse}$$

Aus Rechteckfläche: Höhe  $f(x_i) \cdot \text{Breite } \Delta x_i = \text{Anteilswert}$

$$\text{also } f(x_i) \cdot \Delta x_i = \frac{n_i}{n} \Rightarrow f(x_i) = \frac{n_i}{n \cdot \Delta x_i} \text{ Häufigkeitsdichte}$$

**Beispiele:**

Monatliches Haushaltsnettoeinkommen der Haushalte in der BRD im Jahre 2009

**Tabellarische Darstellung**

Tab. III.4: Monatl. Haushaltsnettoeinkommen der Haushalte in Deutschland im Jahre 2009

Einkommens- klasse $x_i^u \leq x < x_i^o$	Klassen- breite	absolute Häufigkeit $n_i$	relative Häufigkeit $h_i =$ $h(x_i^u \leq x < x_i^o)$	Verteilung s-funktion. in den Klassen- obergrenzen	Dichte- funktion $f(x_i) = \frac{n_i}{n \cdot \Delta x_i}$
(in EUR)		(=Anz. HH in 1000)	$h_i = \frac{n_i}{n}$	$F(x_i^o)$	
unter 500	500	1.400	0,037	0,037	7,46468E-05
500 - unter 1000	500	6.200	0,165	0,203	3,30579E-04
1000 - unter 1500	500	7.500	0,200	0,403	3,99893E-04
1500 - unter 2000	500	7.200	0,192	0,595	3,83898E-04
2000 - unter 2500	500	4.800	0,128	0,722	2,55932E-04
2500 - unter 3000	500	3.800	0,101	0,824	2,02613E-04
3000 - unter 3500	500	2.400	0,064	0,888	1,27966E-04
3500 - unter 4000	500	1.700	0,045	0,933	9,06425E-05
4000 - unter 4500	500	860	0,023	0,956	4,58544E-05
4500 - unter 5000	500	740	0,020	0,976	3,94561E-05
5000 - unter 5500	500	270	0,007	0,983	1,43962E-05
5500 - unter 7000	1500	640	0,017	1,000	1,13747E-05
		37.510	1,000		

Haushalte mit einem Einkommen über EUR 7.000 nicht in der Tabelle erfasst, ihr Anteil beträgt 1,8 % der Stichprobe ( $n = 9708$ ).

Quelle: Sozio-ökonomisches Panel (Welle Z (26)), 2009), eigene Berechnungen

### Grafische Darstellung als Histogramm mit unterschiedlichen Klassenbreiten:

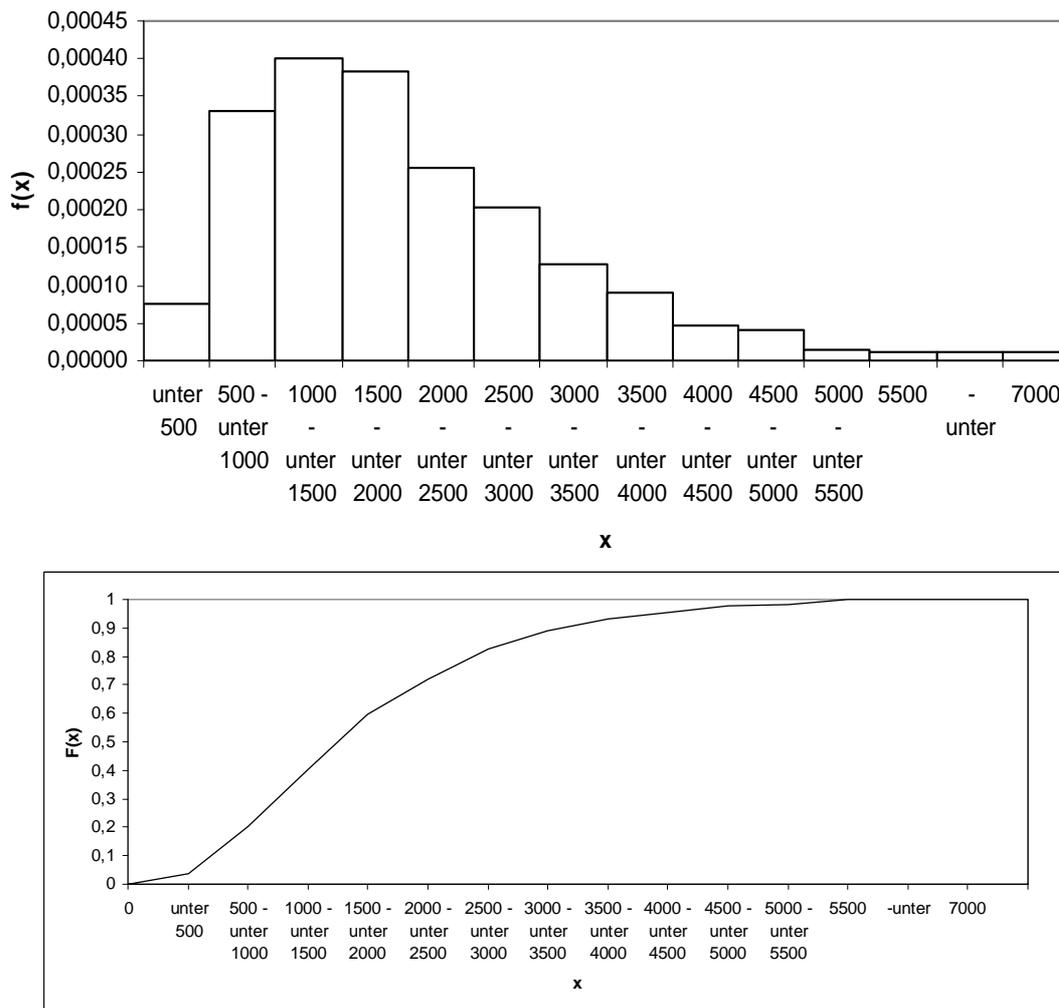


Abb. III.6: Histogramm und Verteilungsfunktion des monatlichen Haushaltsnettoeinkommens der Haushalte in der BRD 2009

Quelle: Sozio-ökonomisches Panel (Welle Z (26), 2009), eigene Berechnungen

#### Zur Berechnung der Anteilswerte innerhalb einer Klasse

Wieviel Prozent der statistischen Einheiten besitzen einen Merkmalswert kleiner oder gleich  $x$ ?

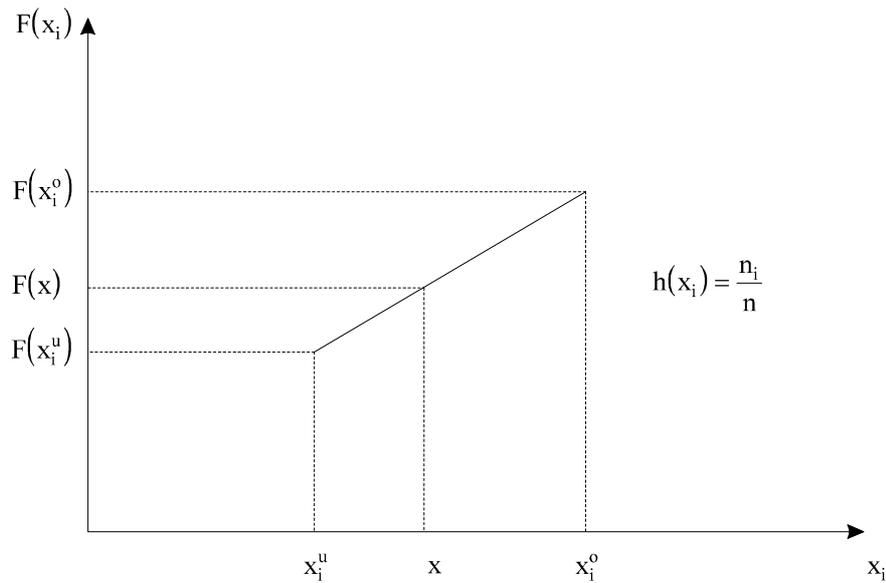
Durch lineare Interpolation (Annahme: Gleichverteilung in der Klasse):

Es gilt:

$$\frac{F(x) - F(x_i^u)}{h(x_i)} = \frac{x - x_i^u}{\Delta x_i} \quad \text{bzw.} \quad \frac{F(x) - F(x_i^u)}{x - x_i^u} = \frac{h(x_i)}{\Delta x_i}$$

Daraus folgt:

$$F(x) = F(x_i^u) + \frac{x - x_i^u}{\Delta x_i} \cdot h(x_i)$$




---

**Beispiel:**

Monatliches Haushaltsnettoeinkommen (Tabelle III.4)

$$F(x = 1800) = 0,403 + \frac{1800 - 1500}{500} \cdot 0,192 = 0,518$$

Etwa 52 % der Haushalte haben weniger als 1.800 EUR Nettoeinkommen pro Monat.

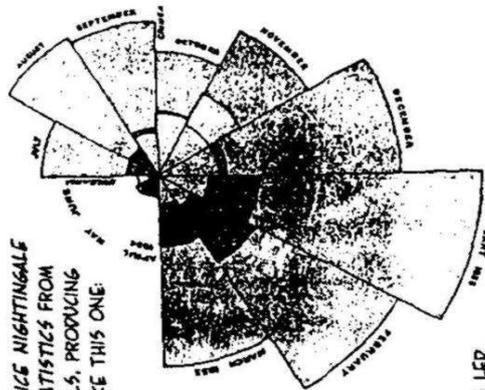
---

## 1.4 Computergestützte grafische Darstellung

PC-Programme wie

- WORD
- EXCEL
- CHART
- 3D-Boeing Graph, Sunrise
- Harvard Graphics
- in Verbindung mit Analyseprogrammen:  
SPSS-PC, SAS-PC, ET, LIMDEP, GAUSS, MICRO-TSP, SYSTAT, STATA...

Beispiele: siehe PC-Vorführung in der Vorlesung

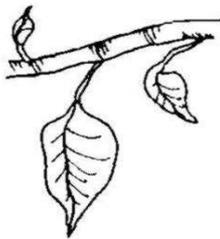


CRUSADING NURSE FLORENCE NIGHTINGALE COMPILED MORTALITY STATISTICS FROM BRITISH MILITARY HOSPITALS, PROVOKING SHOCKING HISTOGRAMS LIKE THIS ONE: THE RADIAL AXIS INDICATES DEATHS—IN HOSPITALS AS WELL AS ON THE BATTLEFIELD—OF BRITISH SOLDIERS IN THE CRIMEAN WAR.

HER STATISTICAL EFFORTS LED DIRECTLY TO IMPROVED HOSPITAL CONDITIONS AND A REDUCTION IN THE DEATH RATE.



THE STATISTICIAN JOHN TUKEY INVENTED A QUICK WAY TO SUMMARIZE DATA AND STILL KEEP THE INDIVIDUAL DATA POINTS. IT'S CALLED THE STEM-AND-LEAF DIAGRAM.



FOR THE WEIGHT DATA, THE STEM IS A COLUMN OF NUMBERS, CONSISTING OF THE WEIGHT DATA COUNTED BY TENS (I.E., WE LEAVE OFF THE LAST DIGIT).

9	:	
10	:	
11	:	829
12	:	0155009
13	:	090015
14	:	09
15	:	0
16	:	
17	:	
18	:	
19	:	
20	:	
21	:	

I.E., 90 POUNDS, 100 POUNDS, ETC.

NOW ADD THE FINAL DIGIT OF EACH WEIGHT IN THE APPROPRIATE ROW:

STEM : LEAVES	
9 :	
10 :	829
12 :	0155009
13 :	090015
14 :	09
15 :	0
16 :	
17 :	
18 :	
19 :	
20 :	
21 :	

MEANING THERE ARE WEIGHTS OF 116, 112, 118, 120, ETC.

FILLED IN, IT LOOKS LIKE THIS:

9	:	5
10	:	200
11	:	829855040
12	:	01553095525
13	:	0900950400159
14	:	09509580902
15	:	5093705109505090909000
16	:	090004
17	:	095000
18	:	0900
19	:	00900
20	:	
21	:	5

AND FINALLY, PUT THE 'LEAVES' IN ORDER.

9	:	5
10	:	200
11	:	027576498
12	:	00013255555
13	:	00000793555409
14	:	00002555558
15	:	000000000099555555557
16	:	000049
17	:	000095
18	:	0009
19	:	00009
20	:	
21	:	5

ALL THOSE ZEROS AND FIVES CLEARLY SHOW THE STUDENT'S REPORTING BIAS!

## 2 Lageparameter

Häufigkeitsverteilungen waren eine erste Stufe der Informationsverdichtung. Weitergehende und stärkere Verdichtung ist durch Maßzahlen möglich wie:

Lageparameter: verschiedene Mittelwerte

Streuungsmaße: Varianzen etc.

Konzentration: Gini, Lorenzkurve

### 2.1 Häufigster Wert (Modus)

Modus oder Modalwert = Wert, der am häufigsten vorkommt.

Der **Modalwert D** eines metrisch skalierten diskreten Merkmals ist derjenige Merkmalswert  $x$ , für den die relative Häufigkeit  $h(x)$  ihr Maximum annimmt.

Die Klasse mit der größten Häufigkeitsdichte  $f(x_i) = \frac{n_i}{n \cdot \Delta x_i}$  heißt **modale Klasse**, ihre Klassenmitte definiert man als Modalwert D (metrisch skaliert, stetig).

Der Modalwert ist nur dann aussagekräftig, wenn die Verteilung eingipflig (unimodal) ist. Bei einer mehrgipfligen (u-förmig, etc.) Verteilung ist er wenig sinnvoll.

#### Beispiele:

---

- Privathaushalte in der BRD 2010  
häufigster Haushaltstyp: 1-Personenhaushalt, Modus D = 1 (diskretes Merkmal)
  - Einkommensverteilung in der BRD 1992  
Modus D = 2500,- (größte Dichte  $f(x)$ , stetiges Merkmal)
- 

### 2.2 Median (Zentralwert)

Der **Median oder Zentralwert** halbiert das Datenmaterial, d.h. 50 % aller Einheiten liegen oberhalb und 50 % aller Einheiten liegen unterhalb dieses Wertes (Median = '50-Prozentpunkt')

Dazu ist es notwendig, die Einheiten nach der Größe ihrer Merkmalswerte zu ordnen.

#### Ungruppiertes Material

Gegeben sind:  $n$  beliebige Merkmalswerte  $x_1, x_2, \dots, x_n$ ;

geordnet nach Größe:  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ;

$x_{(i)}$  : i-ter Merkmalswert der geordneten Reihe

Als **Median Z** wird definiert

$$Z = x_{\left(\frac{n+1}{2}\right)}, \text{ falls } n \text{ ungerade}$$

$$Z = \frac{1}{2} \left[ x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right], \text{ falls } n \text{ gerade}$$

---

**Beispiel:**

Monatsgehälter in der Fa. DALLES & CO.

Männer: 1650, 2030, 1840, 1520, 1670;      n = 5  
 Frauen: 1710, 1960, 2570, 1490                      ; n = 4

Geordnete, sortierte Werte:

Männer: 1520, 1650, 1670, 1840, 2030

Frauen: 1490, 1710, 1960, 2570

M+F: 1490, 1520, 1650, 1670, 1710, 1840, 1960, 2030, 2570

$$Z_M = x_{\left(\frac{n+1}{2}\right)} = x_{(3)} = 1670 \quad (n \text{ ungerade})$$

$$Z_F = \frac{1}{2} \left[ x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right] = \frac{1}{2} (x_{(2)} + x_{(3)}) = \frac{1}{2} (1710 + 1960) = 1835 \quad (n \text{ gerade})$$

$$Z_{M+F} = x_{\left(\frac{n+1}{2}\right)} = 1710$$

Im allgemeinen kann man den Median einer zusammengefaßten Grundgesamtheit **nicht** aus den Medianen der Teilgesamtheiten berechnen!

---

**Grafische Darstellung als 'Box and Whisker'-Plots**

'Box and Whisker'-Plots beschreiben eine Verteilung mit dem Median und dem Bereich um den Median, in dem 50 % aller (geordneten) Werte liegen. Damit liegen 25 % unterhalb und 25 % aller Daten oberhalb der Box.

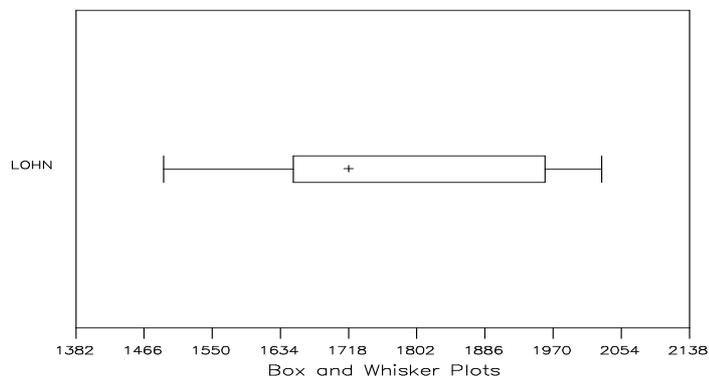
**Beispiel:**

Abb. III.7: ET-'Box and Whisker'-Plot der Monatsgehälter der Fa. BRUCH, DALLES &amp; Co.

Interpretation:

50 % der Beobachtungen liegen zwischen 1650 und 1960; der Median liegt nicht im Zentrum, damit sind die Daten nicht symmetrisch verteilt.

Vertikale Striche: kleinster ( $\pm 1,5 \cdot$ Boxbreite) bzw. größter Wert

**Gruppiertes Material**

Falls das Urmaterial nur gruppiert vorliegt, erhält man den Median **nur approximativ mit Hilfe der Verteilungsfunktion**  $F(x)$ .

Da 50 % der Merkmalswerte einen kleineren Merkmalswert als den Median  $Z$  haben, gilt:

$$h(x \leq Z) = F(Z) = 0,5$$

Lineare Interpolation

$$Z = x_i^u + \frac{F(z) - F(x_i^u)}{f(x_i)} = x_i^u + \frac{F(z) - F(x_i^u)}{n_i / n} \cdot \Delta x_i$$

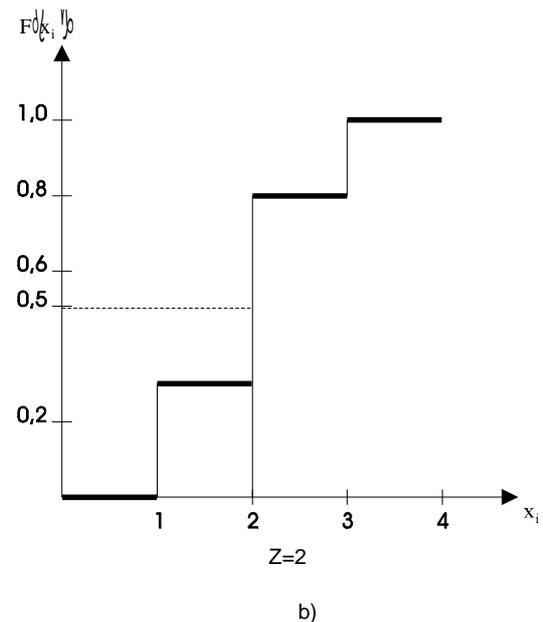
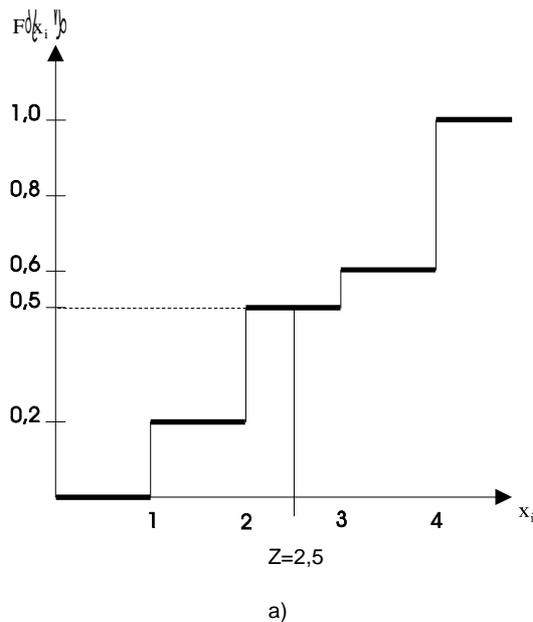
mit  $x_i^u$  als Klassenuntergrenze der Klasse mit  $F(Z) = 0,50$

Bei **metrisch skalierten, diskreten Merkmalen erhält man den Median  $Z$**  durch Ablesen der Stelle  $x$  aus der Treppenfunktion an der Stelle  $F(x) = 0,5 \Rightarrow Z = x$ .

Bestimmung des Medians aus der Verteilungsfunktion bei diskreten Merkmalen

$$F(Z) = 0,50$$

- a) Falls  $F(z)$  auf einer Treppenstufe den Wert 0,5 annimmt, dann ist der Median gleich den Abzissen des Mittelwertes dieser Treppenstufe.
- b) Falls  $F(z)$  den Wert 0,5 nicht annimmt, ist der Median gleich dem kleinsten Merkmalswert, an dem die Verteilungsfunktion größer als 0,5 ist.

**Beispiel:**

Privathaushalte in der BRD 1987:  $F(2) = 50\% \Rightarrow Z = 2$

d.h. ca. 50 % aller Privathaushalte sind 1 oder 2-Personenhaushalte (approximativ).

$Z$  ist hier die Stelle, an der die Verteilungsfunktion  $F(x)$  den Wert 0,5 erstmals überschreitet.

Bei **metrisch skalierten, stetigen Merkmalen** halbiert der Median die Fläche des Histogramms. Bestimmung gegebenenfalls durch lineare Interpolation.

Eine Verallgemeinerung des Median ( $Z = 50\%$  Quantil) ist das Konzept der  $p$ -Quantile, vgl. 3.2.

**Beispiel:**

Monatliches Haushaltsnettoeinkommen BRD 2009

Der Median liegt innerhalb der Einkommensklasse 1500-2000 EUR.

Lineare Interpolation

$$Z = x_i^u + \frac{F(z) - F(x_i^u)}{\frac{n_i}{n}} \cdot \Delta x_i$$

$$\text{hier: } Z = 1500 + \frac{0,50 - 0,403}{0,192} \cdot 500 = 1.753,82$$

d.h. 50 % aller Haushalte haben in der BRD 2009 weniger als 1.753,82 EUR (approximativ) monatliches Nettoeinkommen verdient.

## 2.3 Arithmetisches Mittel

Das **arithmetische Mittel**  $\bar{x}$  gibt an, welchen Merkmalswert **jede** statistische Einheit haben würde, wenn die gesamte Merkmalssumme gleichmäßig auf alle statistischen Einheiten verteilt wäre (Ersatzwert).

### Ungruppiertes Material

Das arithmetische Mittel ( $\bar{x}$ ) ist der Durchschnitt ( $\emptyset$ ) aus den Merkmalswerten aller statistischen Einheiten.

Gegeben:  $n$  beliebige Merkmalswerte  $x_1, x_2, \dots, x_n$

### Arithmetisches Mittel

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Über Umformung ergibt sich die Merkmalssumme:  $n \cdot \bar{x} = \sum_{i=1}^n x_i$

### Beispiel:

- a) Durchschnittliche Körpergröße von Studentinnen und Studenten in der Vorlesung Statistik I an der Uni Lüneburg:

$$\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{1}{5}(x_1 + x_2 + \dots + x_5) = \frac{1}{5}(172 + 178 + 164 + 167 + 171) = \frac{1}{5} \cdot 852 = 170,4 \text{ cm}$$

- b)

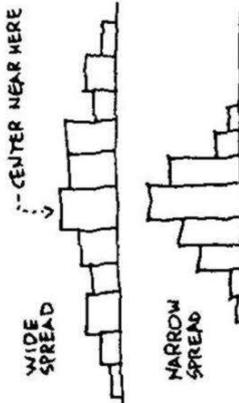
$$\begin{array}{c|ccc} i & 1 & 2 & 3 \\ \hline x_i & 3 & 4 & 5 \end{array} \quad \bar{x} = \frac{1}{3}(3+4+5) = \frac{1}{3} \cdot 12 = 4$$

→  $\bar{x}$  kann, muß aber nicht einen der  $x_i$ -Werte annehmen!

# SUMMARY STATISTICS

NOW WE MOVE FROM PICTURES TO FORMULAS. OUR OBJECT IS TO GET SOME SIMPLE MEASUREMENTS OF THE CRUDEST CHARACTERISTICS OF A SET OF DATA...

ANY SET OF MEASUREMENTS HAS TWO IMPORTANT PROPERTIES: THE CENTRAL OR TYPICAL VALUE, AND THE SPREAD ABOUT THAT VALUE. YOU CAN SEE THE IDEA IN THESE HYPOTHETICAL HISTOGRAMS.



A SMALL SET OF  $n = 5$  DATA POINTS MAKES THE BOOKKEEPING EASY. SUPPOSE, FOR EXAMPLE, WE ASK FIVE PEOPLE HOW MANY HOURS OF TELEVISION THEY WATCH IN A WEEK... AND GET THE FOLLOWING ARRAY:

OBSERVATION	1	2	3	4	5
DATA VALUE	5	7	3	30	7

THEN  $x_1 = 5$ ,  $x_2 = 7$ ,  $x_3 = 3$ ,  $x_4 = 30$ , AND  $x_5 = 7$ .



WHAT'S THE 'CENTER' OF THESE DATA? THERE ARE ACTUALLY SEVERAL DIFFERENT WAYS TO MEASURE IT. WE'LL LOOK AT JUST TWO OF THEM.

## THE MEAN (OR 'AVERAGE')

THE MEAN OR AVERAGE VALUE IS REPRESENTED BY  $\bar{x}$ ... IT'S OBTAINED BY ADDING ALL THE DATA AND DIVIDING BY THE NUMBER OF OBSERVATIONS:

$$\bar{x} = \frac{\text{SUM OF DATA}}{n}$$

$$= \frac{x_1 + x_2 + \dots + x_n}{n}$$

FOR OUR EXAMPLE,

$$\bar{x} = \frac{5 + 7 + 3 + 30 + 7}{5} = \frac{60}{5}$$

= 12 HOURS

WE CAN GO A LONG WAY WITH A LITTLE NOTATION. SUPPOSE WE'RE MAKING A SERIES OF OBSERVATIONS...  $n$  OF THEM, TO BE EXACT... THEN WE WRITE

$$x_1, x_2, x_3, \dots, x_n$$

AS THE VALUES WE OBSERVE. THUS,  $n$  IS THE TOTAL NUMBER OF DATA POINTS, AND  $x_4$  (SAY) IS THE VALUE OF THE FOURTH DATA POINT.

AN ARRAY IS A TABLE OF DATA:

OBSERVATION	1	2	3	4	...	$n$
DATA VALUE	$x_1$	$x_2$	$x_3$	$x_4$	...	$x_n$

READ AS "X-ONE, X-TWO," ETC.

## Gruppiertes Material

Nach Zusammenfassung des Datenmaterials in  $k$  Größenklassen ergibt sich für jede Klasse ein arithmetisches Mittel  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ .

Die Berechnung des arithmetischen Mittels des gesamten Datenmaterials (über alle Klassen) ergibt sich als

**gewichtetes (gewogenes) arithmetisches Mittel (Additionssatz für Mittelwerte):**

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \dots + \bar{x}_k n_k}{n_1 + n_2 + \dots + n_k} = \frac{1}{n} \sum_{i=1}^k \bar{x}_i n_i = \sum_{i=1}^k \bar{x}_i \cdot \frac{n_i}{n} = \sum_{i=1}^k \bar{x}_i \cdot h(x_i)$$

Gewichte: relative Häufigkeiten

Bei **unbekanntem Gruppenmittel** werden die Klassenmitten  $x_i^*$  anstelle von  $\bar{x}_i$  verwendet:

$$\bar{x} \approx \frac{1}{n} \sum_{i=1}^k x_i^* \cdot n_i = \sum_{i=1}^k x_i^* \cdot h(x_i) \quad \text{['je gleichverteilter desto besser']}$$

**Beispiel:**

Klausurnoten Statistik I

Note	$n_i$	$n_i/n$	$\bar{x}_i$
1	4	0,4	1
2	2	0,2	2
3	3	0,3	3
4	1	0,1	4
5	-	0,0	
	10	1,0	

$$\bar{x}_i = \text{Note}_i, \text{ z. B.: } \bar{x}_1 = \frac{1}{4}(1+1+1+1) = 1$$

gewogenes arithmetisches Mittel:

$$\bar{x} = \sum_{i=1}^k \bar{x}_i \cdot h(x_i) = 1 \cdot 0,4 + 2 \cdot 0,2 + 3 \cdot 0,3 + 4 \cdot 0,1 = 2,1$$

ungruppiertes arithmetisches Mittel:

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{1}{10} (1+1+1+1+2+2+3+3+3+4) = \frac{1}{10} \cdot 21 = 2,1$$

FOR THE  $n=92$  STUDENT WEIGHTS, WE CAN FIND THE MEDIAN FROM THE ORDERED STEM-AND-LEAF DIAGRAM: JUST COUNT TO THE 46<sup>TH</sup> OBSERVATION. THE MEDIAN IS

9 : 5	10 : 289
11 : 002556688	12 : 00012355555
13 : 0000019555688	14 : 00007255555 8
15 : 000000000003555555555	16 : 000045
17 : 000055	18 : 00005
19 : 00005	20 :
21 : 5	

$$\frac{x_{46} + x_{47}}{2} = \frac{145 + 145}{2} = 145 \text{ POUNDS}$$

WE HAVE A SHORTHAND FOR THAT  $x_1 + x_2 + \dots + x_n$  USING THE GREEK CAPITAL LETTER SIGMA, FOR SUMMATION:

FOR THE SUM  $x_1 + x_2 + \dots + x_n$  WE WRITE

$$\sum_{i=1}^n x_i$$

AND READ IT AS "THE SUM OF  $x_i$  AS  $i$  GOES FROM 1 TO  $n$ ."

SAY IT TEN TIMES AND YOU'LL NEVER FORGET IT...

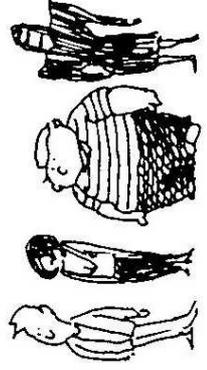


SO... TO REPEAT, THE AVERAGE, OR MEAN, OF A SET OF DATA  $x_i$  IS

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{OR} \quad \sum_{i=1}^n \frac{x_i}{n}$$

IN THE CASE OF OUR 92 PENN STATE STUDENTS, THE MEAN WEIGHT IS

$$\sum_{i=1}^{92} \frac{x_i}{92} = \frac{13,354}{92} = 145.15 \text{ POUNDS}$$



WHY MORE THAN ONE MEASURE OF THE CENTER? EACH HAS ADVANTAGES. FOR EXAMPLE, THE MEDIAN IS NOT SENSITIVE TO OUTLIERS, OR EXTREME VALUES NOT TYPICAL OF THE REST OF THE DATA. SUPPOSE IN OUR SMALL TV-WATCHING GROUP, ONE PERSON WATCHES 200 HOURS PER WEEK. THEN OUR DATA ARE 3, 5, 7, 7, 200. THE MEDIAN, 7, IS UNCHANGED, BUT THE MEAN IS NOW  $\bar{x} = 45.8!$

IN 1984 THE UNIVERSITY OF VIRGINIA ANNOUNCED THAT ITS DEPARTMENT OF RHETORIC AND COMMUNICATIONS GRADUATES' MEAN STARTING SALARY WAS \$59,000. THE OUTLIER, THE SALARY OF N.B.A. CENTER RALPH SAMPSON, DID NOT REPRESENT THE EARNING POWER OF A B.A. IN SPEECH FROM U. OF V. (THE MEDIAN SALARY WASN'T PUBLISHED.)

### Formale Eigenschaften des arithmetischen Mittels $\bar{x}$

- Die Summe der Abweichungen der Merkmalswerte von  $\bar{x}$  ist Null.

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n \cdot \bar{x} = \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0 \quad \text{q.e.d.}$$

- Die Summe der quadrierten Abweichungen der Merkmalswerte von  $\bar{x}$  ist ein Minimum.

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \min$$

Beweis:

$$d(a) = \sum_{i=1}^n (x_i - a)^2 \quad a \text{ beliebig}$$

Extremproblem: 1. Ableitung wird Null gesetzt  $\rightarrow$  Extrema (min, max) möglich

$$\begin{aligned} \frac{d(a)}{da} &= -\sum_i 2(x_i - a) \cdot 1 && \text{Kettenregel} \\ &= -2\sum_i (x_i - a) = 0 \end{aligned}$$

Also:

$$\begin{aligned} \sum_i (x_i - a) &= 0 \\ \sum_i x_i - n \cdot a &= 0 \\ \Rightarrow a &= \frac{1}{n} \sum_i x_i = \bar{x} \quad \text{q.e.d.} \end{aligned}$$

### Fechnersche Lageregel zum Vergleich von arithmetischem Mittel $\bar{x}$ , Zentralwert (Median) Z und Modus D

a) symmetrische Verteilung:  $\bar{x} = Z = D$

b) asymmetrische Verteilung:  $\bar{x} \neq Z \neq D$

Links- und rechtssteile Verteilungen (Abb. III.8)

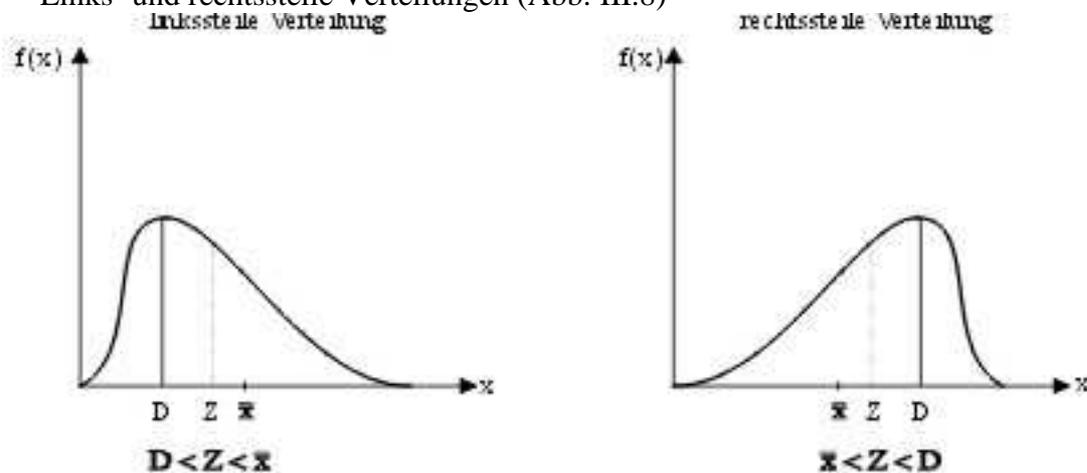


Abb. III.8: Fechnersche Lageregel und links- und rechtssteile Verteilung

## 2.4 Geometrisches Mittel

Das geometrische Mittel ist sinnvoll bei der Mittlung von Wachstumsraten oder anderen multiplikativ verknüpften Merkmalswerten.

Gegeben: positive Merkmalswerte  $x_1, x_2, \dots, x_n$

### Geometrisches Mittel

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (x_i > 0)$$

oder

$$\log GM = \frac{1}{n} (\log x_1 + \log x_2 + \dots + \log x_n) = \frac{1}{n} \sum_{i=1}^n \log x_i$$

### Beispiel:

---

Umsätze der Firma F.I.R.M.A. von 2006 - 2010 in Mio. EUR

Jahr	Umsatz	Zuwachsrate in %	Wachstums- faktor
2006	2,0		
2007	2,4	+20,00	1,2000
2008	2,9	+20,83	1,2083
2009	2,7	- 6,89	0,9310
2010	3,1	+14,81	1,1481

Wie groß ist der durchschnittliche relative Umsatzzuwachs (Zuwachsrate) pro Jahr?

$$GM = \sqrt[4]{1,2 \cdot 1,2083 \cdot 0,9310 \cdot 1,1481} = 1,11579$$

Durchschnittliche Zuwachsrate pro Jahr:  $(1,11579 - 1) \cdot 100\% = 11,5791\%$

Jahr	Umsatz	$\cdot 1,11579$
2006	2,0	2,2316
2007	2,2316	2,4900
2008	2,4900	2,7783
2009	2,7783	3,1000
2010	3,1000	---

---

## 2.5 Harmonisches Mittel

Das harmonische Mittel wird bei der Mittlung von Brüchen mit konstantem Zähler angewandt (z.B.: Geschwindigkeit dividiert durch die Zeit, Preise, Verhältniszahlen).

Gegeben: positive Merkmalswerte  $x_1, x_2, \dots, x_n$

### Harmonisches Mittel

$$\text{HM} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

#### Beispiel:

---

Fertigungszeiten: Vier Arbeiter sind acht Stunden (= 480 min) lang mit der Herstellung eines Einzelteils beschäftigt:

Arbeiter	Fertigungszeit je Stück in Min.
A	2,3
B	3,0
C	3,4
D	3,7

Wie hoch ist die durchschnittliche Fertigungszeit?

Das arithmetische Mittel ergäbe

$$\bar{x} = \frac{1}{4}(2,3 + 3,0 + 3,4 + 3,7) = 3,1 \text{ min}$$

Besser harmonisches Mittel: Denn Fertigungszeiten je Stück sind arithmetische Mittel, die aus konstanter Arbeitszeit (480 Min.) und der Stückzahl  $n_i$  als  $\bar{x}_i = 480/n_i$  berechnet sind

$$\text{HM} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{4}{\frac{1}{2,3} + \frac{1}{3,0} + \frac{1}{3,4} + \frac{1}{3,7}} = \frac{4}{1,332} = 3,0$$

Die durchschnittliche Fertigungszeit je Stück beträgt also drei Minuten.

---

## 3 Streuungsmaße

Mittelwerte bzw. Lageparameter sind in der Praxis wichtige Verteilungsparameter. Sie liefern aber noch unvollständige Beschreibungen einer Häufigkeitsverteilung, denn es läßt sich keine Aussage über die Größe der Abweichungen der einzelnen Merkmalswerte vom Mittelwert machen. Maßzahlen dafür sind die **Streuungsmaße** (vgl. Abb. III.9).

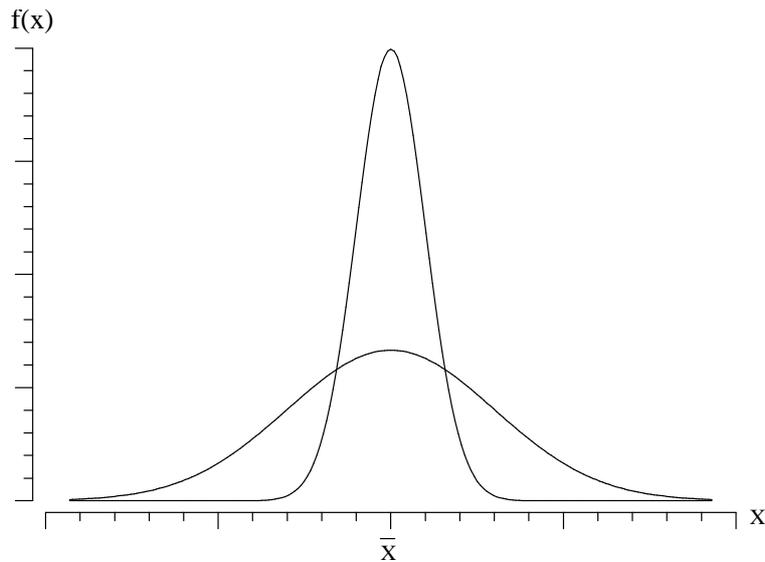


Abb. III.9: Häufigkeitsverteilungen mit gleichem Mittelwert  $\bar{x}$  aber verschiedenen Streuungen

### 3.1 Spannweite

Die Spannweite entspricht dem Konzept des häufigsten Wertes (Modus D).

#### Spannweite (Range) R

Differenz zwischen dem größten ( $x_{\max}$ ) und dem kleinsten Merkmalswert ( $x_{\min}$ )

$$R = x_{\max} - x_{\min} = x_{(n)} - x_{(1)}$$

Bei gruppierten Daten werden die Klassengrenzen (Klassenmitten) der Randklassen verwendet.

Der Nachteil der Spannweite R besteht in der Verwendung der extremen Werte (sog. Ausreißer).

#### Beispiele:

---

a) Körpergrößen: 172, 178, 164, 167, 171 [cm]

$$x_{\max} = x_{(5)} = x_2 = 178$$

$$x_{\min} = x_{(1)} = x_3 = 164$$

$$R = x_{\max} - x_{\min} = 178 - 164 = 14 \text{ cm}$$

b) Temperaturen: 7, 13, -6, 25 [°C]

$$x_{\max} = 25$$

$$x_{\min} = -6$$

$$R = 25 - (-6) = 31 \text{ °C}$$

### 3.2 Quartilsabweichung und p-Quantile

Ein Streuungsmaß auf der Basis des Mediankonzeptes ist der Quartilsabstand, dessen allgemeine Grundlage die p-Quantile sind.

Für den Median  $Z$  ( $Z = x_p$  mit  $p = 0,50$  – Quantil) gilt, dass 50 % aller Merkmalsträger einen kleineren oder gleich großen ( $\leq$ ) Merkmalswert haben.  $Z$  halbiert somit die Fläche der Häufigkeitsdichte/bzw. Häufigkeitsverteilung).

Verallgemeinert gilt für ein p-Quantil, dass p % aller Merkmalsträger einen kleineren oder gleich großen Merkmalswert haben. Damit wird mit einem p-Quantil auch p % der Fläche unter der Häufigkeitsdichte abgetrennt (vgl. Abb. III.10).

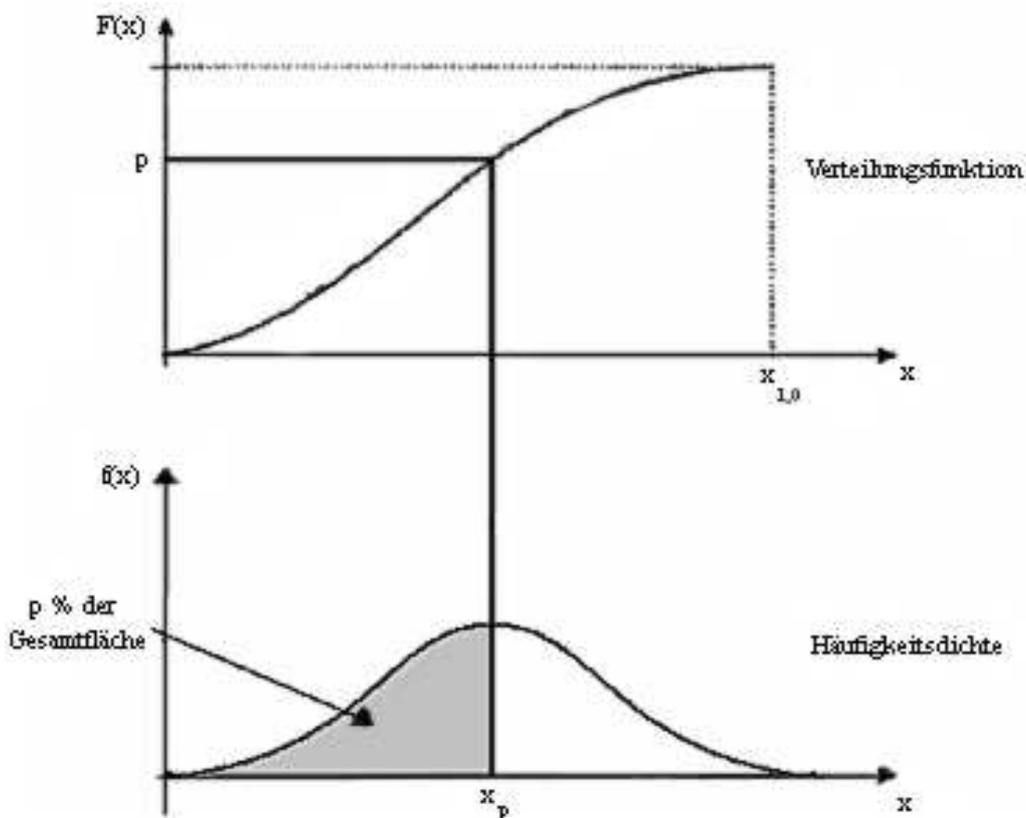


Abb. III.10: p-Quantile, Häufigkeitsdichte und Verteilungsfunktion

Ist also  $x_{(1)} \leq \dots \leq x_{(n)}$  die geordnete Merkmalsreihe, dann ist allgemein das p-Quantil  $x_p$  ( $0 < p < 1$ )

$$x_p = \begin{cases} x_{(K)} & \text{,falls } n \cdot p \text{ keine ganze Zahl ist} \\ & \text{(K ist dann die auf } n \cdot p \text{ folgende ganze Zahl)} \\ \frac{1}{2}(x_{(K)} + x_{(K+1)}) & \text{,falls } n \cdot p \text{ eine ganze Zahl ist} \\ & \text{(dann ist } K = n \cdot p \text{)} \end{cases}$$

Der Median ist somit ein spezielles p-Quantil, nämlich das mit  $p = 50\%$ . Weitere spezielle p-Quantile sind das 0,25-Quantil (unteres Quartil) und das 0,75-Quantil (oberes Quartil) oder die Dezile mit  $p = 0,10$ ,  $p = 0,20$  etc.

### p-Quantil der Verteilungsfunktion F(x)

$$F(x_p) = p \quad \text{bzw.} \quad x_p = F^{-1}(p)$$

Bei **gruppiertem Datenmaterial** erfolgt die Berechnung von  $x_p$  nach der Interpolationsformel:

$$x_p = x_i^u + \frac{F(x_p) - F(x_i^u)}{f(x_i)} = x_i^u + \frac{F(x_p) - F(x_i^u)}{n_i/n} \cdot \Delta x_i$$

### Quartilsabweichung

$p = 0,25$  und  $0,75$

$$QA = \frac{1}{2}(x_{0,75} - x_{0,25}) \quad F(x_{0,25}) = 0,25; F(x_{0,75}) = 0,75; F(x_{0,50}) = F(Z) = 0,50$$

In dem Bereich der QA liegen die mittleren 50 % aller Merkmalswerte. Die QA ist nicht von Extremwerten abhängig, sie ist als durchschnittliche Streuung zu interpretieren. Interquartile Spannweiten und 'Box and Whisker'-Plots beschreiben die Bereiche mit 25 % bzw. 50 % der Daten.



**Beispiele:**

a) Monatsgehälter in der Fa. DALLES & Co. (siehe Median)

geordnete Werte:

Männer: 1520, 1650, 1670, 1840, 2030 (n = 5)

Frauen: 1490, 1710, 1960, 2570 (n = 4)

Quantile Männer

$$x_{0,25} : n \cdot p = 5 \cdot 0,25 = 1,25 \text{ keine ganze Zahl, } K = 2, x_{0,25} = x_{(2)} = 1650 \text{ EUR}$$

25 % aller Monatsgehälter liegen unter 1650 EUR.

$$x_{0,75} : n \cdot p = 5 \cdot 0,75 = 3,75 \text{ keine ganze Zahl, } K = 4, x_{0,75} = x_{(4)} = 1840 \text{ EUR}$$

$$\text{Quartilsabweichung: } QA_M = \frac{1}{2}(x_{0,75} - x_{0,25}) = \frac{1}{2}(1840 - 1650) = 95$$

Die 50 % Merkmalswerte um den Median streuen um  $\pm 95$  um  $Z_M = 1670 \text{ EUR}$ .

Quantile Frauen

$$x_{0,25} : n \cdot p = 4 \cdot 0,25 = 1 \text{ ganze Zahl, } K = 1$$

$$x_{0,25} = \frac{1}{2}(x_{(1)} + x_{(2)}) = \frac{1}{2}(1490 + 1710) = 1600 \text{ EUR}$$

$$x_{0,75} : n \cdot p = 4 \cdot 0,75 = 3 \text{ ganze Zahl, } K = 3$$

$$x_{0,75} = \frac{1}{2}(x_{(3)} + x_{(4)}) = \frac{1}{2}(1960 + 2570) = 2265 \text{ EUR}$$

$$\text{Quartilsabweichung: } QA_F = \frac{1}{2}(x_{0,75} - x_{0,25}) = \frac{1}{2}(2265 - 1600) = 332,5$$

b) Monatliches Haushaltsnettoeinkommen BRD 2009, Quartilsabweichung

Aus  $F(0,75)$  folgt:

$$x_{0,75} = x_i^u + \frac{0,75 - F(x_i^u)}{n_i/n} \cdot \Delta x_i = 2500 + \frac{0,75 - 0,722}{0,101} \cdot 500 = 2.635,86 \text{ EUR}$$

Aus  $F(0,25)$  folgt:

$$x_{0,25} = 1000 + \frac{0,25 - 0,203}{0,200} \cdot 500 = 1.118,50 \text{ EUR}$$

$$QA = \frac{1}{2}(x_{0,75} - x_{0,25}) = \frac{1}{2}(2.635,86 - 1.118,50) = 758,68$$

Im Mittel weichen die Haushaltseinkommen um  $QA = 758,68 \text{ EUR}$  vom Median  $Z = 1.753,82 \text{ EUR}$  ab.

- c) Monatliches Haushaltsnettoeinkommen BRD 2009, Percentile und Spannweite, berechnet auf Grundlage der nicht klassierten Werte der Stichprobe

Tab. III.5: Monatliches Haushaltsnettoeinkommen BRD 2009, Percentile, Quartilsabweichung und Spannweite in EUR

Min	150	Spannweite:	
10 %	850	$R = x_{\max} - x_{\min}$	29.850,0
20 %	1.140	1. Quartil (25 %)	1.275,0
30 %	1.400	2. Quartil (50 %)	1.924,0
40 %	1.650	3. Quartil (75 %)	2.800,0
Median	1.924	Quartilsabweichung QA	762,5
60 %	2.200		
70 %	2.600		
80 %	3.000		
90 %	4.000		
Max	30.000		

Quelle: Sozio-ökonomisches Panel (Welle Z (26), 2009), eigene Berechnungen

### 3.3 Mittlere absolute Abweichung

Die mittlere absolute Abweichung ist ein Streuungskonzept hinsichtlich des arithmetischen Mittels.

Da  $\sum_{i=1}^n (x_i - \bar{x}) = 0$ , wird die durchschnittliche absolute Abweichung  $d$  gewählt. Positive und negative Abweichungen  $(x_i - \bar{x})$  heben sich somit nicht (!) auf.

#### Mittlere absolute Abweichung $d$ (Mean absolute deviation = MAD)

für ungruppiertes Material:

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

für gruppiertes Material:

$$d = \frac{1}{n} \sum_{i=1}^k |x_i^* - \bar{x}| \cdot n_i = \sum_{i=1}^k |x_i^* - \bar{x}| \cdot \frac{n_i}{n} = \sum_{i=1}^k |x_i^* - \bar{x}| \cdot h_i,$$

wobei  $x_i^*$  die Klassenmitte der Klasse  $i$  ist.

#### Beispiel:

Temperaturen:  $-6^\circ, 18^\circ, 12^\circ, 3^\circ$  [ $^\circ\text{C}$ ]

$$\bar{x} = \frac{1}{4}(-6 + 18 + 12 + 3) = \frac{27}{4} = 6,75^\circ\text{C}$$

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| = \frac{1}{4} \{ |-6 - 6,75| + |18 - 6,75| + |12 - 6,75| + |3 - 6,75| \}$$
$$= \frac{1}{4} (12,75 + 11,25 + 5,25 + 3,75) = \frac{1}{4} (33) = 8,25$$

Die mittlere absolute Abweichung beträgt 8,25 °C.

---

### 3.4 Mittlere quadratische Abweichung: Varianz und Standardabweichung

Konzept des arithmetischen Mittels: Streuungsmaß mit Abweichungen von  $\bar{x}$

Gegeben: Merkmal  $x$  mit den Ausprägungen  $x_1, x_2, \dots, x_n$

#### Varianz

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

#### Standardabweichung (standard deviation)

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{s^2} = \sqrt{\text{var}(x)}$$

Die Standardabweichung ist gebräuchlicher, da sie die gleiche Dimension wie die Merkmalswerte aufweist.

Gegenüber der mittleren absoluten Abweichung gewichtet die mittlere quadratische Abweichung (Varianz  $s^2$ ) größere Abweichungen durch die Quadrierung stärker als kleinere Abweichungen (gebräuchlicher als mittlere absolute Abweichung)

Die Varianz bzw. Standardabweichung wird auch als **empirische Varianz** bzw. **empirische Standardabweichung** bezeichnet.

**Vereinfachte Berechnung von  $s^2$  bzw.  $s$ :**

$$\begin{aligned}
s^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n} 2\bar{x} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \bar{x}^2 \\
&= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n} 2\bar{x} n\bar{x} + \frac{1}{n} n\bar{x}^2 \\
&= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x}^2 + \bar{x}^2 \\
&= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2
\end{aligned}$$

Bei nicht-vereinfachter Berechnung:

2 Durchgänge durch den Datensatz

Bei vereinfachter Berechnung:

**nur 1 Durchgang** durch den Datensatz!

Dies spart computing-costs bei großen Datenmengen (z.B. EVS mit über 40000 Haushalten)

**Beispiel:**

a) Temperaturen

i	$x_i$	$x_i^2$
1	-6	36
2	18	324
3	12	144
4	3	9
$\Sigma$	27	513

$$\bar{x} = 27/4 = 6,75$$

Vereinfachte Berechnung:

Varianz

$$s^2 = \frac{1}{4} \cdot 513 - (6,75)^2 = 128,25 - 45,5625 = 82,6875$$

Standardabweichung

$$s = \sqrt{s^2} = 9,0933$$

Nicht vereinfachte Berechnung ( $\bar{x} = 6,75$ ):Mit  $\bar{x} = 6,75$  ergibt sich:

i	$x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	-6	-12,75	162,5625
2	18	11,25	126,5625
3	12	5,25	27,5625
4	3	-3,75	14,0625
$\Sigma$	27		330,75

Varianz

$$s^2 = \frac{1}{4} \cdot 330,75 = 82,6875$$

Standardabweichung

$$s = \sqrt{s^2} = 9,0933$$

Zum Vergleich: mittlere absolute Abweichung  $d = 8,25$

b) Computerprogramm zur Varianzberechnung z.B. in FORTRAN:

```
DO 10 I=1,N
  SUMX=SUMX+X(I)
10 SUMX2=SUMX2+X(I)**2
VAR=SUMX2/N-(SUMX/N)**2
SD=SQRT(VAR)
```

---

BUT A SPREAD MEASURE SHOULD HAVE THE SAME UNITS AS THE ORIGINAL DATA. IN THE EXAMPLE OF WEIGHTS, THE VARIANCE  $s^2$  IS MEASURED IN POUNDS SQUARED. COOPS!



THE OBVIOUS THING TO DO IS TO TAKE THE SQUARE ROOT, AND SO WE DO... TO DEFINE:



# STANDARD DEVIATION

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

WHICH, FOR OUR SIMPLE DATA SET, IS

$$s = \sqrt{214} = 14.63$$

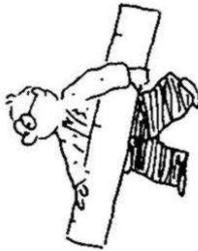


EVEN FOR SMALL DATA SETS, THE ARITHMETIC CAN BE TEDIOUS! SO NOWADAYS, WE JUST HIT THE  $\Sigma$  BUTTON ON THE HAND CALCULATOR, OR CONSULT THE DATA REPORT GENERATED BY A COMPUTER SOFTWARE PACKAGE.

THE STANDARD MEASURE OF SPREAD IS THE

# STANDARD DEVIATION

UNLIKE THE IQR, WHICH IS BASED ON MEDIANS, THE STANDARD DEVIATION MEASURES THE SPREAD FROM THE MEAN. YOU CAN THINK OF IT, ROUGHLY SPEAKING, AS THE AVERAGE DISTANCE OF THE DATA FROM THE MEAN  $\bar{x}$ .

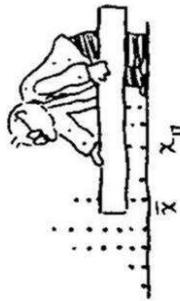


EXCEPT THAT WE USE THE SQUARES OF THE DISTANCES INSTEAD. THAT IS, IF THE SQUARED DISTANCE OF POINT  $x_i$  TO  $\bar{x}$  IS  $(x_i - \bar{x})^2$ , THEN

$$\text{AVERAGE SQUARED DISTANCE} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

FOR TECHNICAL REASONS, WE USE  $n-1$  IN THE DENOMINATOR RATHER THAN  $n$ , AND DEFINE THE SAMPLE VARIANCE  $s^2$  AS

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$



FOR THE DATA SET  $\{3 \ 5 \ 7 \ 7 \ 30\}$ , WITH  $\bar{x} = 12$  AND  $n = 5$  WE CALCULATE THE VARIANCE:

$$s^2 = \frac{(3-12)^2 + (5-12)^2 + (7-12)^2 + (7-12)^2 + (30-12)^2}{(5-1)}$$

$$= \frac{81 + 49 + 25 + 25 + 676}{4}$$

$$= 214$$

THE LARGE VARIANCE HERE REFLECTS THE WIDE SPREAD IN THE DATA...



### Varianz bei gruppiertem Datenmaterial

Die Berechnung erfolgt nun über die Klassenmitte  $x_i^*$ :

$$s^2 = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot n_i$$

vereinfacht:

$$s^2 = \frac{1}{n} \sum_{i=1}^k (x_i^*)^2 \cdot n_i - \bar{x}^2$$

#### Beispiel:

Monatliches Haushaltsnettoeinkommen BRD 2009,

Tab. III.6: Monatliches Haushaltsnettoeinkommen,  
Mittelwerte und Streuungsermittlung

$x_i^*$	$n_i$	$x_i^{*2}$	$x_i^{*2} \cdot n_i$
250	1.400	62.500	87.500.000
750	6.200	562.500	3.487.500.000
1250	7.500	1.562.500	11.718.750.000
1750	7.200	3.062.500	22.050.000.000
2250	4.800	5.062.500	24.300.000.000
2750	3.800	7.562.500	28.737.500.000
3250	2.400	10.562.500	25.350.000.000
3750	1.700	14.062.500	23.906.250.000
4250	860	18.062.500	15.533.750.000
4750	740	22.562.500	16.696.250.000
5250	270	27.562.500	7.441.875.000
6250	640	39.062.500	25.000.000.000
	37.510		204.309.375.000

$$s^2 = \frac{1}{n} \sum_i (x_i^*)^2 \cdot n_i - \bar{x}^2 = \frac{204.309.375.000}{37.510} - 1.999,13^2$$

$$= 5.446.797,67 - 3.996.520,80 = 1.450.276,8$$

$$s = \sqrt{1.450.276,8} = 1204,27 \text{ EUR}$$

**Interpretation:** Das Haushaltseinkommen weicht im Durchschnitt um 1204,27 EUR vom mittleren Haushaltseinkommen  $\bar{x} = 1.999,13 \text{ EUR}$  ab.

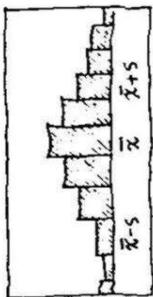
Für die Quartilsabweichung ergab sich ein Wert von 762,50 EUR.

### 3.5 Variationskoeffizient

Für den Vergleich verschiedener Häufigkeitsverteilungen wird eine 'relative Streuung' (Streuungsmaß/Lagemaß) definiert mit

## Properties of $\bar{X}$ and S

THE MEAN AND STANDARD DEVIATION ARE VERY GOOD FOR SUMMARIZING THE PROPERTIES OF FAIRLY SYMMETRICAL HISTOGRAMS WITHOUT OUTLIERS—I.E., HISTOGRAMS SHAPED LIKE MOUNDS.



A SHAPE TO REMEMBER!

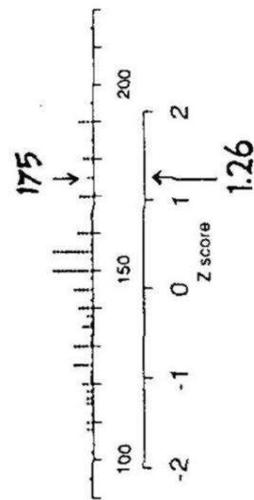


IT'S OFTEN USEFUL TO KNOW HOW MANY STANDARD DEVIATIONS A DATA POINT IS FROM THE MEAN. WE DEFINE Z-SCORES, OR STANDARDIZED SCORES, AS DISTANCE FROM  $\bar{x}$  PER STANDARD DEVIATION.

$$z_i = \frac{x_i - \bar{x}}{s} \text{ FOR EACH } i.$$



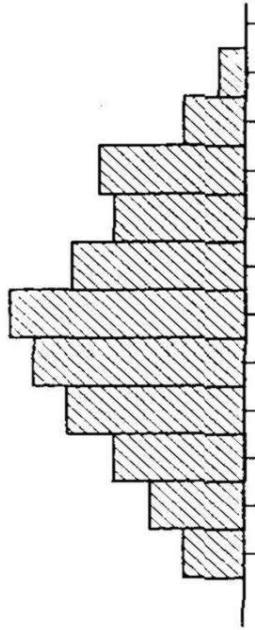
A Z-SCORE OF +2 MEANS THAT AN OBSERVATION IS TWO STANDARD DEVIATIONS ABOVE THE MEAN. FOR THE WEIGHT DATA ( $\bar{x} = 145.2$  AND  $s = 23.7$ ), WE CAN PLOT THE DATA ON THE ORIGINAL X-AXIS IN POUNDS AND THE Z-SCORE AXIS SIMULTANEOUSLY.



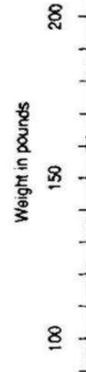
A STUDENT WEIGHING 175 POUNDS HAS A Z-SCORE OF  $\frac{175 - 145.2}{23.7} = 1.26$

## an EMPIRICAL RULE:

FOR NEARLY SYMMETRIC MOUND-SHAPED DATA SETS, APPROXIMATELY 68% OF THE DATA IS WITHIN ONE STANDARD DEVIATION OF THE MEAN AND 95% OF THE DATA IS WITHIN TWO STANDARD DEVIATIONS OF THE MEAN.



FOR THE WEIGHTS, OUR EMPIRICAL RULE HOLDS UP PRETTY WELL: 64% (= 59/92) OF THE WEIGHTS ARE WITHIN ONE STANDARD DEVIATION OF THE MEAN, AND 97% (= 89/92) OF THE WEIGHTS ARE WITHIN TWO STANDARD DEVIATIONS OF THE MEAN.



CUTE LIL' OUTLIER!



AND NOW FOR A REST FROM NUMBER CRUNCHING!

59 points

89 points

92 points



### Variationskoeffizient

$$V = \frac{s}{\bar{x}} \cdot 100(\%) = \frac{\text{Standardabweichung}}{\text{Mittelwert}} \cdot 100$$

**Problem:** Bei positiven und negativen Merkmalswerten kann  $\bar{x}$  nahe Null sein. Dadurch entstehen 'beliebig' große Werte von V.

### Beispiel:

---

a) Temperaturbeispiel

$$s = 9,09 [^{\circ}\text{C}]$$

$$\bar{x} = 6,75 [^{\circ}\text{C}]$$

$$V = \frac{s}{\bar{x}} = \frac{9,09 [^{\circ}\text{C}]}{6,75 [^{\circ}\text{C}]} = 1,35 \quad (\text{dimensionslos !})$$

b) Monatliches Haushaltsnettoeinkommen BRD 2009

$$s = 1204,27 [EUR]$$

$$\bar{x} = 1.999,13 [EUR]$$

$$V = \frac{s}{\bar{x}} = \frac{1.204,27}{1.999,13} = 0,60$$

Interpretation: Relative Streuung von b) < a)

Aber: Vorsicht bei der inhaltlichen Interpretation! Es ist besser, Äpfel mit Äpfeln und Birnen mit Birnen zu vergleichen!

---

## 3.6 Konzept der Momente, Schiefe und Exzeß

Momente sind Verallgemeinerungen des Varianzkonzeptes.

### Momente

Durchschnittliche **potenzierte** Abweichungen der Merkmalswerte von einem Bezugspunkt (a).

Bezugspunkt Null:

Momente um Null

Bezugspunkt arithmetisches Mittel: Momente um das arithmetische Mittel

### Zentrale Momente

$$m_r^a = \frac{1}{n} \sum_{i=1}^n (x_i - a)^r \quad (\text{zentrale Momente})$$

**Das r-te Moment um Null**

$$m_r^0 = \frac{1}{n} \sum_{i=1}^n (x_i - 0)^r = \frac{1}{n} \sum_{i=1}^n x_i^r \quad (\text{ungruppierte Daten})$$

$$m_r^0 = \frac{1}{n} \sum_{i=1}^k (x_i^* - 0)^r \cdot n_i = \frac{1}{n} \sum_{i=1}^k (x_i^*)^r \cdot n_i \quad (\text{gruppierte Daten})$$

$$(\text{für } r = 1: \quad m_r^0 = \bar{x})$$

**Das r-te Moment um das arithmetische Mittel  $\bar{x}$** 

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r \quad (\text{ungruppierte Daten})$$

$$m_r = \frac{1}{n} \sum_{i=1}^K (x_i^* - \bar{x})^r \cdot n_i \quad (\text{gruppierte Daten})$$

Momente höherer Ordnung ergeben die Schiefe ( $r = 3$ ) und den Exzeß (Kurtosis, Wölbung) ( $r = 4$ ).

**Schiefe**

Das 3. zentrale Moment ( $r = 3$ ) gibt Auskunft über die Symmetrie, bzw. Asymmetrie einer Verteilung.

**3. zentrales Moment:**

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

Das 3. zentrale Moment allein ist jedoch kein sehr geeignetes Maß für die Unsymmetrie, da seine Größe von der Streuung und der Maßeinheit der Variable beeinflusst wird. Daher wird das folgende Schiefemaß (skewness) verwendet:

**Schiefe:**

$$sm_3 = \frac{m_3}{s^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[ \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right]^3}$$

Symmetrie:  $sm_3 = 0$

Asymmetrie: je stärker negativ die Maßzahl, desto rechtssteiler (linksschiefer) ist die Verteilung.

je stärker positiv die Maßzahl, desto linkssteiler (rechtsschiefer) ist die Verteilung.

Sinnvoll bei Eingipfligkeit der Verteilung (unimodal).

ET verwendet z.B.  $sm_3$  mit  $(n-1)$  statt  $n$ .

**Wölbung (Kurtosis, Exzeß)**

Auskunft über den Grad der Wölbung oder Spitzigkeit einer Verteilung gibt das 4. zentralen Moment ( $r = 4$ ).

**4. zentrales Moment:**

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4$$

Das 4. zentrale Moment ist für jede Verteilung positiv.

*Kleinerer* Werte des 4. zentralen Moments deuten auf eine *spitzere/steiler gewölbte* Verteilung.

*Große* Werte weisen auf eine *flachere* Verteilung hin.

Um ein maßstabs- und streuungsunabhängiges Maß zu erhalten verwendet man die Kurtosis als Maß für die Wölbung ('thickness of the distribution tails').

**Kurtosis:**

$$sm_4 = \frac{m_4}{s^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[ \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right]^4}$$

ACHTUNG: Durch die relativ großen  $s^4$ -Werte dreht sich die Interpretation von  $sm_4$  gegenüber  $m_4$  um:

Nun steht ein *kleinerer* standardisierter Exzeß-Wert für eine *flachere* Verteilung, *große* Werte weisen auf eine *spitzere* Verteilung hin.

Da für die Normalverteilung  $sm_4 = 3$  ist, erfolgt auch hier manchmal eine Normierung mit

$$sm_4^* = sm_4 - 3$$

Ist  $sm_4^* > 0$ , dann ist die Verteilung *spitzer/steiler gewölbt* als die Normalverteilung (bei gleicher Varianz und Mittelwert).

Ist  $sm_4^* < 0$ , dann ist die Verteilung *flacher* als die Normalverteilung (bei gleicher Varianz und Mittelwert).

**Beispiele:**

a) Monatliches Haushaltsnettoeinkommen, BRD 2009

Bisher:

$$\bar{x} = 1.999,13 \text{ EUR}$$

$$s = 1.204,27 \text{ EUR}$$

gruppierte Daten:  $m_r = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^r \cdot n_i$

Tab. III.7: Monatliches Haushaltsnettoeinkommen, BRD 2009, Hilfswerte für Schiefe und Exzeß

i	Einkommens- klasse	$n_i$	$x_i^*$	$(x_i^* - \bar{x})^3 n_i$	$(x_i^* - \bar{x})^4 n_i$
1	unter 500	1.400	250	-7,492E+12	1,310E+16
2	500 - unter 1000	6.200	750	-1,208E+13	1,509E+16
3	1000 - unter 1500	7.500	1.250	-3,153E+12	2,362E+15
4	1500 - unter 2000	7.200	1.750	-1,113E+11	2,774E+13
5	2000 - unter 2500	4.800	2.250	7,579E+10	1,901E+13
6	2500 - unter 3000	3.800	2.750	1,609E+12	1,208E+15
7	3000 - unter 3500	2.400	3.250	4,697E+12	5,876E+15
8	3500 - unter 4000	1.700	3.750	9,125E+12	1,598E+16
9	4000 - unter 4500	860	4.250	9,807E+12	2,207E+16
10	4500 - unter 5000	740	4.750	1,540E+13	4,238E+16
11	5000 - unter 5500	270	5.250	9,276E+12	3,016E+16
12	5500 - unter 7000	640	6.250	4,916E+13	2,090E+17
		37.510		7,631E+13	3,572E+17

Quelle: Sozio-ökonomisches Panel (Welle Z (26), 2009), eigene Berechnungen

Schiefe:

$$m_3 = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^3 \cdot n_i$$

$$= 2,034E+09$$

$$sm_3 = \frac{m_3}{s^3} = \frac{2,034E+09}{1,747E+09} = 1,165$$

Exzeß (Wölbung):

$$m_4 = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^4 \cdot n_i$$

$$= 9,524E+12$$

$$sm_4 = \frac{m_4}{s^4} = \frac{9,524E+12}{2,103E+12} = 4,528$$

$$sm_4^* = m_4 / s^4 - 3 = 4,528 - 3 = 1,528$$

Damit ist die Normalverteilung flacher

b) Einkommensvergleich BRD (2009) und U.K. (1979-80)

Tab. III.8: Einkommensverteilung für U.K. (1979 - 80)

	Mittelwert $\bar{x}$	Varianz $m_2 = s^2$	Schiefe $sm_3$	Exzeß $sm_4$
Haushaltsnettoeinkommen BRD [EUR]	1.999,13	1,450E+06	1,165	4,528
Personal income* U.K. [£]	3.700,-	$4,8 \cdot 10^6$	1,45	7,48

\*Quelle: Spanos (1986), S. 24 ff, Sozio-ökonomisches Panel (Welle Z (26), 2009), eigene Berechnungen

Es liegen zwar verschiedene Zeitpunkte und verschiedene Währungen (EUR und £) sowie Haushalts- bzw. Personenkonzepte vor, dennoch:

jeweils U.K. **größerer Mittelwert** als BRD  
 U.K. **größere Streuung** als BRD  
 U.K. **Verteilung linkssteiler** als BRD.  
 U.K. **flachere, breitere Verteilung** als BRD

c) ET: F7, F8 Descriptives, Histogramm

für Körpergrößen von 20 Studentinnen (HWOMEN) und 20 Studenten (HMEN)

- Listing
- Descriptive Statistics
- Stem and leaf Plots
- Percentile
- Box and Whisker-Plots
- Frequency Tables
- Histograms

```
DATA LISTING (Current sample)      ↓/ESC          Press ESC to interrupt list.
Observation      HMEN              HWOMEN
   1             182.00            158.00
   2             182.00            164.00
   3             180.00            174.00
   4             187.00            178.00
   5             179.00            163.00
   6             184.00            168.00
   7             174.00            170.00
   8             168.00            165.00
   9             186.00            158.00
  10             172.00            154.00
  11             178.00            169.00
  12             182.00            176.00
  13             186.00            172.00
  14             174.00            174.00
  15             183.00            164.00
  16             192.00            168.00
  17             177.00            169.00
  18             179.00            168.00
  19             176.00            167.00
  20             178.00            163.00
```

```
Descriptive Statistics
Variable      Mean      Std. Dev.      Skew.      Kurt.      Minimum      Maximum      Cas.↓
HMEN          179.95      5.6983        -.020      2.690      168.0        192.0        20
HWOMEN        167.10      6.1976        -.254      2.480      154.0        178.0        20
```

Stem and Leaf Plot for HMEN

Use ↑ and ↓ to scroll. ESC=exit.  
 8 lines TOP of file

```
Total number of observations = 20
   1  Low values discarded
   1  high values discarded

Stem width = 100.00

Count      Stem      Leaves
   18      1 .    77777777778888888888
```

Stem and Leaf Plot for HWOMEN

Use ↑ and ↓ to scroll. ESC=exit.  
8 lines TOP of file

Toal number of observations = 20

1 Low values discarded

0 high values discarded

t

Stem width = 100.00

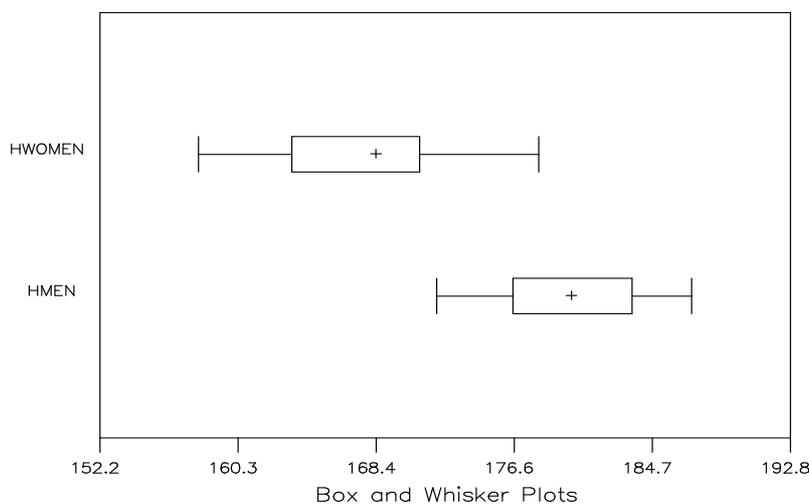
Count	Stem	Leaves
19	1 .	55666666666666777777

## Order Statistics for Variables

Percentile	HMEN	HWOMEN
Min.	168.00	154.00
10th	173.00	158.00
20th	175.00	163.00
25th	176.50	163.50
30th	177.50	164.00
40th	178.50	166.00
Med.	179.50	168.00
60th	182.00	168.50
70th	182.50	169.50
75th	183.50	171.00
80th	185.00	173.00
90th	186.50	175.00
Max.	192.00	178.00

## Partition of range Min to Max

Range of X	HMEN	HWOMEN
Minimum	168.00	154.00
1st.Qrtl	174.00	160.00
Midpoint	180.00	166.00
3rd.Qrtl	186.00	172.00
Maximum	192.00	178.00



Histogram for HMEN computed using 20 observations  
 Obs. out of range: too low= 0, too high= 0

Individual data      Mean= 179.950, std.dev.= 5.698

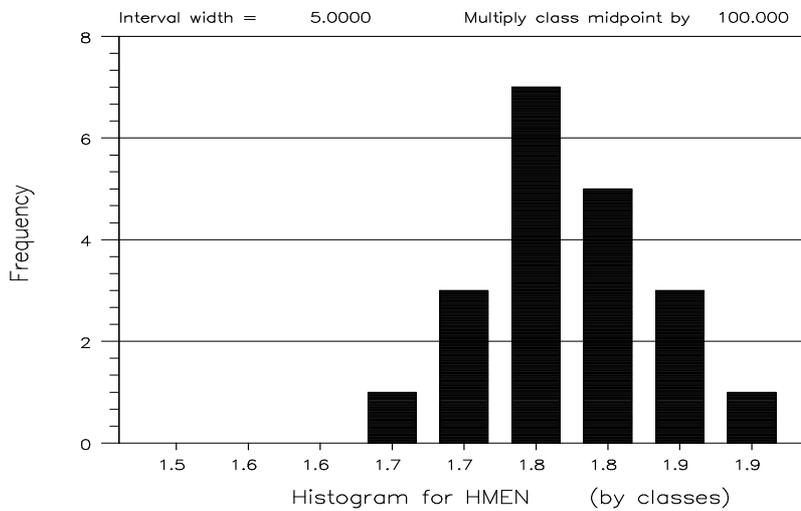
	Lower Limit		Upper Limit		Frequency		Cumulative	
	Lower Limit	Upper Limit	Total	Relative	Total	Relative	Total	Relative
0	150.000	155.000	0	.0000	0	.0000	0	.0000
1	155.000	160.000	0	.0000	0	.0000	0	.0000
2	160.000	165.000	0	.0000	0	.0000	0	.0000
3	165.000	170.000	1	.0500	1	.0500	1	.0500
4	170.000	175.000	3	.1500	4	.2000	4	.2000
5	175.000	180.000	7	.3500	11	.5500	11	.5500
6	180.000	185.000	5	.2500	16	.8000	16	.8000
7	185.000	190.000	3	.1500	19	.9500	19	.9500
8	190.000	195.000	1	.0500	20	1.0000	20	1.0000

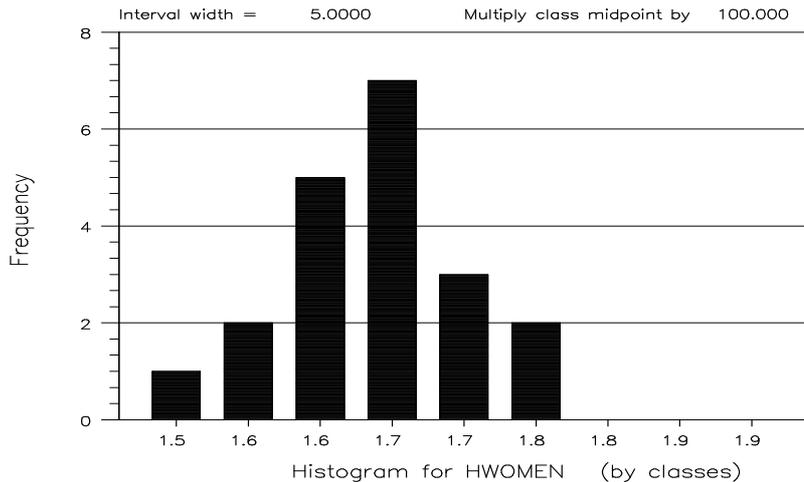
Histogram for HWOMEN computed using 20 observations

Obs. out of range: too low= 0, too high= 0

Individual data      Mean= 167.100, std.dev.= 6.198

	Lower Limit		Upper Limit		Frequency		Cumulative	
	Lower Limit	Upper Limit	Total	Relative	Total	Relative	Total	Relative
0	150.000	155.000	1	.0500	1	.0500	1	.0500
1	155.000	160.000	2	.1000	3	.1500	3	.1500
2	160.000	165.000	5	.2500	8	.4000	8	.4000
3	165.000	170.000	7	.3500	15	.7500	15	.7500
4	170.000	175.000	3	.1500	18	.9000	18	.9000
5	175.000	180.000	2	.1000	20	1.0000	20	1.0000
6	180.000	185.000	0	.0000	20	1.0000	20	1.0000
7	185.000	190.000	0	.0000	20	1.0000	20	1.0000
8	190.000	195.000	0	.0000	20	1.0000	20	1.0000





## 4 Konzentration einer Verteilung

Zur Untersuchung der Frage, ob sich bestimmte Merkmale (Einkommen, Vermögen, Firmenumsätze etc.) bei bestimmten anderen Merkmalen (Personen, Haushalte, Firmentypen etc.) konzentrieren; also Frage nach der Gleich-/Ungleichverteilung.

### 4.1 Konzentration

Die Standardabweichung ist bereits ein Maß für die Konzentration:

$s$  mißt die durchschnittliche Abweichung von der Gleichverteilung des Merkmals  $i$  ( $\bar{x}$  gibt an, welcher Wert sich ergibt, wenn die Merkmalssumme auf alle Einheiten gleich aufgeteilt würde).

Informativer ist die Abweichung von der Gleichverteilung für jede Klasse!

Für Konzentrationsanalysen werden die Merkmale grundsätzlich erst nach ihrer Größe geordnet.

#### Gleich-/Ungleichverteilung über Klassen

Für jede Klasse:

Bilde die **Differenz**  $d_i$  zwischen der **beobachteten Merkmalssumme**  $x_i^* \cdot n_i$  und der **Merkmalssumme bei Gleichverteilung**  $\bar{x} \cdot n_i$ .

$$d_i = x_i^* \cdot n_i - \bar{x} \cdot n_i \quad \text{für } i = 1, 2, \dots, k$$

Anteil an gesamter Merkmalssumme:

$$\tilde{d}_i = \frac{d_i}{\bar{x} \cdot n} = \frac{x_i^* \cdot n_i}{\bar{x} \cdot n} - \frac{n_i}{n} \quad \text{für } i = 1, 2, \dots, k$$

Eine Gleichverteilung liegt dann vor, wenn  $\tilde{d}_i = 0$  für alle Klassen gilt.

**Beispiel:** \_\_\_\_\_

Monatliches Haushaltsnettoeinkommen BRD 2009

$d_i$  = Differenz zwischen

Einkommensanteil der Haushalte der Klasse  $i$  ( $x_i^* \cdot n_i / \bar{x} \cdot n$ ) und

Anteil der Haushalte dieser Klasse  $i$  an allen Haushalten ( $n_i/n$ )

Tab. III.9: Monatliches Haushaltsnettoeinkommen 2009, Konzentration der Verteilung

Einkommens- klasse	$x_i^*$	$n_i$	$\frac{n_i}{n}$	$x_i^* \cdot n_i$	$\frac{x_i^* \cdot n_i}{n \cdot \bar{x}}$	$\tilde{d}_i$	$F(x_i^o)$	$MS(x_i^o)$
unter 500	250	1.400	0,037	350.000	0,005	-0,033	0,037	0,005
500 - unter 1000	750	6.200	0,165	4.650.000	0,062	-0,103	0,203	0,067
1000 - unter 1500	1250	7.500	0,200	9.375.000	0,125	-0,075	0,403	0,192
1500 - unter 2000	1750	7.200	0,192	12.600.000	0,168	-0,024	0,595	0,360
2000 - unter 2500	2250	4.800	0,128	10.800.000	0,144	0,016	0,722	0,504
2500 - unter 3000	2750	3.800	0,101	10.450.000	0,139	0,038	0,824	0,643
3000 - unter 3500	3250	2.400	0,064	7.800.000	0,104	0,040	0,888	0,747
3500 - unter 4000	3750	1.700	0,045	6.375.000	0,085	0,040	0,933	0,832
4000 - unter 4500	4250	860	0,023	3.655.000	0,049	0,026	0,956	0,881
4500 - unter 5000	4750	740	0,020	3.515.000	0,047	0,027	0,976	0,928
5000 - unter 5500	5250	270	0,007	1.417.500	0,019	0,012	0,983	0,947
5500 - unter 7000	6250	640	0,017	4.000.000	0,053	0,036	1,000	1,000
		37.510	1,00	74987500,0	1,000	0,00		
				= $n \cdot \bar{x}$				

Quelle: Sozio-ökonomisches Panel (Welle Z (26), 2009), eigene Berechnungen

Interpretation:

Konzentration (Ungleichheit) dadurch, dass  $\tilde{d}_i$  bis unter 2000 negativ, für die übrigen Klassen positiv ist.

Die beiden letzten Spalten dieser Tabelle sind besonders interessant. Aus

$$\tilde{d}_i = \underbrace{\frac{x_i^* \cdot n_i}{\bar{x} \cdot n}}_{\text{Beitrag zu kumulierter relativer Merkmalssumme}} - \underbrace{\frac{n_i}{n}}_{\text{Verteilungsfunktion}}$$

folgt:

**Kumulierte relative Merkmalssumme:**  $MS(x_i^o) = \frac{\sum_{i=1}^{i^o} x_i^* \cdot n_i}{\bar{x} \cdot n}$

**Verteilungsfunktion:**  $F(x_i^o) = h(x \leq x_i^o) = \sum_{i=1}^{i^o} \frac{n_i}{n}$

Interpretation:

Auf  $F(x_i^o)$  [%] aller Einheiten mit  $x < x_i^o$  entfallen  $MS(x_i^o)$  [%] der gesamten Merkmalssumme.

**Beispiel:**

Monatliches Haushaltsnettoeinkommen 2009, Einkommensklasse  $i=3$

Auf 40,3 % aller Haushalte (mit einem Einkommen unter 1500 EUR) fallen nur 19,2 % des Gesamteinkommens aller Haushalte.

oder:  $i = 11$

Auf 98,3 % aller Haushalte (mit einem Einkommen unter 5500 EUR) fallen 94,7 % des Gesamteinkommens aller Haushalte.

## 4.2 Lorenzkurve und Gini-Koeffizient

Zur zusammenfassenden grafischen und quantitativen Beschreibung der Konzentration einer Verteilung wird die Lorenzkurve und der Gini-Koeffizient verwendet.

### Lorenzkurve

Für die Lorenzkurve überträgt man die Wertepaare  $[F(x_i), MS(x_i)]$  in ein Koordinatensystem.

Lorenzkurve = Streckenzug, der  $[0,0]$  mit allen Wertepaaren  $[F(x_i), MS(x_i)]$  verbindet.

Gleichverteilung:

Gleichverteilung liegt dann vor, wenn  $F(x_i) = MS(x_i)$  für alle  $i$  ist. Es herrscht dann keine Konzentration vor, da alle Punkte auf der Diagonalen eines Quadrates liegen. Die Diagonale verläuft von  $[0,0]$  bei:  $[(F(x_i), MS(x_i))] = [(F(x_i), MS(x_i))]$  bis  $[1,1]$ .

### Voraussetzungen für den Vergleich zweier Verteilungen:

Die Lorenzkurven dürfen sich nicht schneiden!

Es sollten zusätzliche Informationen verwendet werden, damit keine Fehlinterpretationen entstehen.

### Beispiel:

Lorenzkurve für das monatliche Haushaltsnettoeinkommen in der BRD 2009 (Abb. III.11)

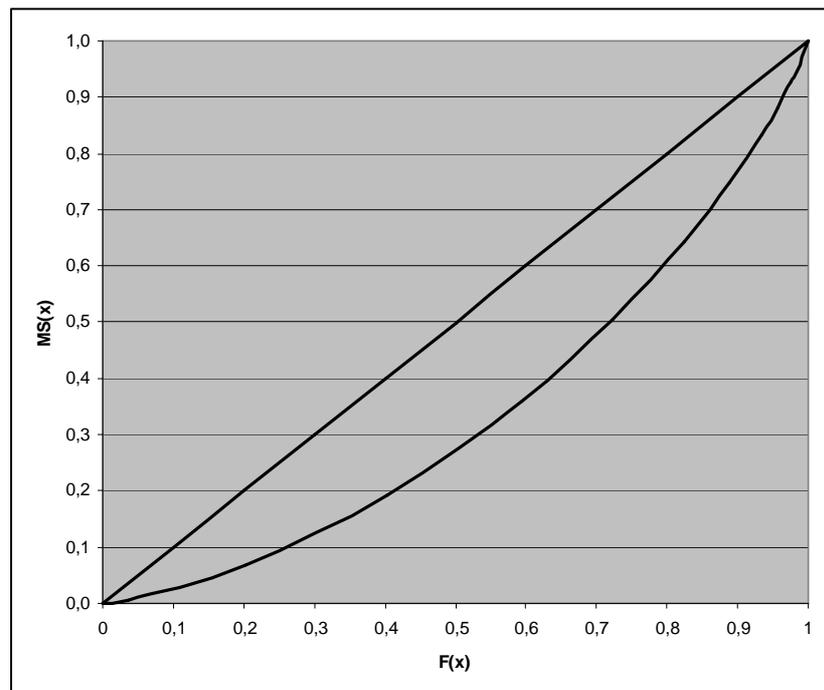


Abb. III.11: Lorenzkurve für das monatliche Haushaltsnettoeinkommen 2009 in der BRD  
Quelle: Sozio-ökonomisches Panel (Welle Z (26), 2009), eigene Berechnungen

### Gini-Koeffizient

Die Fläche zwischen der Gleichverteilungsgeraden und der Lorenzkurve wird als Maß für die Konzentration verwendet.

$$G = \frac{\text{Fläche zwischen Lorenzkurve und Gleichverteilungsgerade}}{\text{Fläche des Dreiecks unter der Gleichverteilungsgeraden}}$$

Je kleiner die Fläche (bzw. G), desto gleichverteilter.

$$G = \sum_{i=1}^k \left\{ \left[ F(x_{i-1}) + F(x_i) \right] \cdot \frac{n_i \cdot x_i^*}{n \cdot \bar{x}} \right\} - 1$$

ungruppiert ( $x_i$  **geordnet!**)

$$G = \frac{2 \cdot \sum_{i=1}^n i \cdot x_i - (n+1) \cdot \sum_{i=1}^n x_i}{n \cdot \sum_{i=1}^n x_i}$$

### Beispiele

a) Gleicher Gini-Koeffizient bei verschiedenen Sachverhalten:

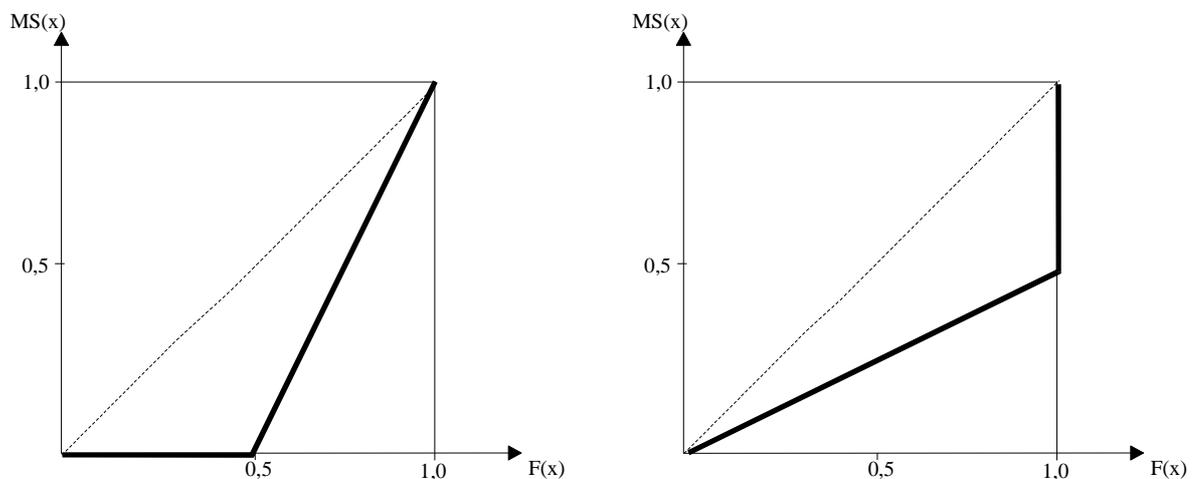


Abb. III.12: **A** **B**  
Gleiche Gini-Koeffizienten bei verschiedenem Sachverhalten

A: Auf die eine Hälfte der Haushalte entfällt ein Haushaltsnettoeinkommen von (z.B.) Null, während die andere Hälfte alle das gleiche Haushaltsnettoeinkommen haben.

B: Die Hälfte des gesamten Einkommens entfällt gleichmäßig auf alle Haushalte mit Ausnahme eines einzigen Haushalts, der die andere Hälfte des Gesamteinkommens bezieht.

b) Gini-Koeffizienten und zusammenfassende deskriptive Informationen für das monatliche Haushaltsnettoeinkommen 1983, 1984, 1992 und 2009 in der BRD

Zur Einkommens- und Verbrauchsstatistik vgl. auch Abschnitt II.6.1.

Tab. III.10: Gini-Koeffizienten und zusammenfassende deskriptive Informationen für das monatliche Haushaltsnettoeinkommen 1983, 1984, 1992 und 2009

	EVS (1983)	SOEP (1984)	SOEP (1991)	SOEP(2009)
Zahl der Beobachtungen	42750	5587	6431	9768
Hochger. Merkmalsumme (Mrd. DM)	75,35	61,88	111,06	83,36
Hochger. Zahl d. Bezugseinheiten	22545383	24599576	34763559	37509999
Hochger. arithmetisches Mittel (DM)	3342	2514	3195	2230
Varianz	4,4648 10 <sup>6</sup>	5,2203 10 <sup>6</sup>	3,738 10 <sup>6</sup>	1,45 10 <sup>6</sup>
Schiefe	1,03 10 <sup>9</sup>	2,927 10 <sup>9</sup>	0,5598 10 <sup>10</sup>	2,03 10 <sup>9</sup>
Exzeß	2,18 10 <sup>13</sup>	2,12 10 <sup>13</sup>	4,4 10 <sup>13</sup>	9,52 10 <sup>12</sup>
Spannweite (DM)	24836	82185	20750	29850
Quartilsabweichung (DM)	1250	800	1050	762,5
Median (DM)	2917	2100	2800	1924
Gini-Koeffizient	0,32851	0,31921	0,30149	0,32
N-tils-Verteilung (in %)				
0% - 25%	9,69	10,12	10,27	10,17
25% - 50%	17,67	18,09	18,12	17,79
50% 75%	26,80	26,08	26,70	26,24
75% 100%	45,02	44,78	44,56	45,80
Ant. d. untersten 20% (in %)	7,08	7,30	7,45	7,93
Ant. der obersten 5% (in %)	14,27	15,00	13,01	14,93
Randgruppenrelation 90/10	8,06	8,22	6,99	6,87

Quelle: (Hansen (1974), S. 18); Die Berechnungen auf der Basis der EVS 1983 wurden freundlicherweise durch Herrn Jürgen Faik an der Professur für Sozialpolitik, Prof. Dr. Hauser, Universität Frankfurt, vorgenommen (vgl. auch Merz und Faik (1992)); Sozio-ökonomisches Panel (Welle I (1), 1984; Welle I (9), 1992; Welle Z (26), 2009), eigene Berechnungen, hochgerechnete Werte

c) Einkommensverteilung im Längsschnitt aus Daten des Sozio-ökonomischen Panels (vgl. Abb. III.13)

Tabelle 1: Regelsätze nach § 22 Bundessozialhilfegesetz und Bedarfsgewichte

Zeitraum	Personen					
	Erwachsene		Kinder			
	HHV <sup>1)</sup>	Andere	Alter 0-6	Alter 7-10	Alter 11-14	Alter 15-21
<b>Mindestsätze in DM</b>						
1.7.1982-30.6.1983	338	270	152	220	254	304
1.7.1983-30.6.1984	345	276	155	224	259	311
1.7.1984-30.6.1985	356	285	160	232	267	321
1.7.1985-30.6.1986	384	307	173	250	288	346
1.7.1986-30.6.1987	394	315	177	256	295	354
<b>Bedarfsgewicht in %</b>	1,0	0,8	0,45	0,65	0,75	0,9

Anmerkung: Rechnerischer Durchschnitt für das Bundesgebiet; Regelsätze ohne Mehrbedarf

Quelle: Nachrichtendienst des Deutschen Vereins für öffentliche und private Fürsorge; verschiedene Jahrgänge

1) HHV = Haushaltsvorstand

Tabelle 2: Kennziffern der Verteilung des Nettoeinkommens von Haushalten und der Nettowohlstandsposition von Haushalten und Personen (NWP errechnet) für die Jahre 1983 bis 1986

	1983	1984	1985	1986
<b>Haushalte</b>				
Arithmetisches Mittel	2919	2991	3019	3070
Zentralwert	2548	2543	2560	2612
Gini-Koeffizient	0,324	0,338	0,334	0,338
Quintilsverteilung in %				
1.Quartil	6,89	6,59	6,90	6,65
2.Quartil	12,33	11,94	12,02	11,93
3.Quartil	17,48	17,15	17,10	17,10
4.Quartil	24,20	24,17	23,99	24,20
5.Quartil	39,08	40,15	39,99	40,12
Anteil der oberen 5%	13,47	14,25	14,43	14,38
<b>Personen</b>				
Arithmetisches Mittel	1434	1493	1516	1558
Zentralwert	1287	1333	1346	1403
Gini-Koeffizient	0,259	0,267	0,268	0,268
Quintilsverteilung in %				
1.Quartil	9,61	9,48	9,38	9,15
2.Quartil	14,15	13,84	13,86	14,02
3.Quartil	17,99	17,90	17,77	17,96
4.Quartil	22,76	22,91	22,80	22,89
5.Quartil	35,49	35,98	36,18	35,97
Anteil der oberen 5%	12,48	12,90	13,13	12,82

Quelle: Berntsen, R. (1991)

Nettowohlstandsposition:

$$NWP = \frac{\sum \text{Einkünfte (HH)}}{\sum \text{Bedarfsgewichte}^*}$$

(\* aus Regelsätzen nach § 22 Bundessozialhilfegesetz)

Tabelle 3: Verteilung der Wohlstandsposition von Personen nach Vielfachen der durchschnittlichen Wohlstandsposition - Vergleich der Jahre 1983 mit 1986 (Wanderungsbilanz)

Wohlstandsposition 1983	Wohlstandsposition 1986								
	unter 0,5 %	0,50 bis 0,75 %	0,75 bis 1,00 %	1,00 bis 1,25 %	1,25 bis 1,50 %	1,50 bis 1,75 %	1,75 bis 2,00 %	2,00 und mehr %	Ge- sam t %
Nettowohlstandsposition unter 0,5 %	39,3	42,6	8,7	5,3	2,1	[0,9]	[0,5]	[0,6]	100
0,50 bis unter 0,75 %	16,1	49,6	24,9	6,1	2,2	(1,0)	[0,1]	[0,1]	100
0,75 bis unter 1,00 %	5,8	22,6	43,1	18,8	7,7	(0,9)	(0,7)	(0,4)	100
1,00 bis unter 1,25 %	3,9	9,4	26,1	36,9	14,1	4,7	2,0	3,0	100
1,25 bis unter 1,50 %	(2,1)	8,1	9,6	25,7	30,4	14,2	7,0	2,9	100
1,50 bis unter 1,75 %	(2,7)	(4,5)	9,1	14,4	23,3	27,3	9,3	9,4	100
1,75 bis unter 2,00 %	[0,4]	(5,0)	(6,0)	9,6	12,4	25,3	24,6	16,6	100
2,00 und mehr %	[1,5]	(2,7)	(4,8)	9,8	10,0	12,1	12,8	46,2	100

Erläuterung: ( ) = Fallzahl unter 30 Personen;  
[ ] = Fallzahl unter 10 Personen

Quelle: Berntsen, R. (1991)

Tabelle 4: Kurzfristige relative Veränderungsklassen der Wohlstandsposition von Personen 1983 bis 1986

Veränderungsklassen	Kurzfristige Veränderungsrate* Anteile in %		
	1984 zu 1983	1985 zu 1984	1986 zu 1985
<b>Nettowohlstandsposition</b>			
Relative Aufstiege			
50 % u.m.	5,7	6,0	5,5
25 bis unter 50 %	8,8	10,0	8,0
10 bis unter 25 %	13,6	14,6	14,5
0 bis unter 10 %	19,5	21,2	24,0
Relative Abstiege			
0 bis unter 10 %	22,4	21,0	23,7
10 bis unter 25 %	19,6	16,7	13,6
25 bis unter 50 %	8,4	7,7	8,1
50 % u.m.	2,0	2,8	2,7

\* Relative Veränderung zwischen zwei Jahren in Veränderungsklassen

Quelle: Berntsen, R. (1991)

Abb. III.13: Einkommensverteilung im Längsschnitt in der BRD 1983 bis 1986 aus Daten des Sozio-ökonomischen Panels

Wenn man einen Wechsel der Wohlstandsposition so definiert, dass mehr als 10 % Auf- oder Abstieg eintreten muß, dann haben 1984 zu 1983 mehr als 50 % ihre Wohlstandsposition verändert.

***Keyconcepts***

*Häufigkeits-/Verteilungsfunktion*

*Modus, Median*

*Arithmetisches Mittel, Geometrisches Mittel, Harmonisches Mittel*

*Spannweite*

*p-Quantile*

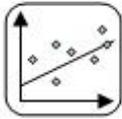
*Mittlere absolute Abweichung*

*Varianz und Standardabweichung*

*Konzept von Momente, Schiefe und Exzeß*

*Konzentration (Lorenzkurve, Gini-Koeffizient)*

## IV Statistische Analyse mehrerer Merkmale



Analyse mehrerer Merkmale, Messung von Zusammenhängen mittels  
Korrelationsrechnung und Regressionsrechnung

Mehrdimensionale Betrachtung gemeinsam auftretender Merkmale zur Analyse der Zusammenhänge zwischen mehreren sozioökonomischen Merkmalen einer statistischen Masse

Sozioökonomische Merkmale einer Person

Arbeitszeit, Einkommen, Alter, Geschlecht, Haushaltsgröße, Haushaltszusammensetzung etc.

### Zwei zentrale Fragestellungen:

- Wie stark ist der Zusammenhang zwischen den Variablen?

#### Beispiele:

Privater Konsum - Volkseinkommen über die Jahre

Gewinn - Umsatz über Betriebe

Benzinverbrauch - Geschwindigkeit über Meßpunkte

→ **Korrelationsrechnung** (nicht gerichtete Analyse)

- Gerichtete Analyse: In welcher funktionalen Weise können die Abhängigkeiten zwischen den Variablen beschrieben werden? → Kausalanalyse

#### Beispiele:

Privater Konsum =  $f(\text{Volkseinkommen})$

Arbeits(zeit)angebot =  $f(\text{Lohnsatz, Alter, Ausbildung,...})$

Gewinn =  $f(\text{Werbung, Marktkonzentration,...})$

→ **Regressionsrechnung**

## 1 Zweidimensionale Häufigkeitsverteilungen und ihre Darstellung

### 1.1 Allgemeine Grundbegriffe und Darstellungsweisen

Für jede statistische Einheit (wie z.B. Person, Haushalt, Betrieb) werden die Merkmalsausprägungen von **zwei Merkmalen x und y** (z.B. Alter x und Geschlecht y einer Person) erhoben und tabellarisch dargestellt: → Zweidimensionale Tabelle

Jeder Kombination von  $(x_i, y_j)$  wird die Anzahl (absolute Häufigkeit)  $n_{ij}$  zugeordnet:

$$n(x_i, y_j) = n_{ij}$$

$x_i$  und  $y_j$  könne dabei nominal, ordinal oder metrisch skaliert sein.

### Allgemeiner Aufbau einer zweidimensionalen Tabelle

	$y_1$	$y_2$	$\cdots$	$y_j$	$\cdots$	$y_m$	Zeilen- summen
$x_1$	$n_{11}$	$n_{12}$	$\cdots$	$n_{1j}$	$\cdots$	$n_{1m}$	$n_{1\cdot}$
$x_2$	$n_{21}$	$n_{22}$	$\cdots$	$n_{2j}$	$\cdots$	$n_{2m}$	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$x_i$	$n_{i1}$	$n_{i2}$	$\cdots$	$n_{ij}$	$\cdots$	$n_{im}$	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\cdots$	$n_{kj}$	$\cdots$	$n_{km}$	$n_{k\cdot}$
Spalten- summen	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot m}$	$n (= n_{\cdot \cdot})$

### Absolute Häufigkeit

$$n(x_i, y_j) = n_{ij}$$

### Relative Häufigkeit

$$h(x_i, y_j) = \frac{n_{ij}}{n}$$

### Beispiel:

Privathaushalte in der BRD (Mai 1987) nach Familienstand (x) und Geschlecht (y) des Haushaltsvorstandes (HHV) (in 1000)

Tab. IV.1: Privathaushalte in der BRD (Mai 1987) nach Familienstand (x) und Geschlecht (y) des Haushaltsvorstandes (HHV) (in 1000)

Familienstand	(x)	Geschlecht des HHV (y)		Zeilensumme von $x_i$ : $\sum_{j=1}^2 n_{ij}$
		männlich ( $y_1$ )	weiblich ( $y_2$ )	
ledig	( $x_1$ )	2.755	2.403	5.158
verheiratet	( $x_2$ )	14.929	680	15.609
verwitwet	( $x_3$ )	694	3.988	4.682
geschieden	( $x_4$ )	744	1.211	1.955
Spaltensumme		19.122	8.282	27.404
	von $y_j$ : $\sum_{i=1}^4 n_{ij}$			

Quelle: Statistisches Jahrbuch 1989, S. 56, Volkszählung 1987

$$n_{32} = 3.988, \quad n_{2.} = 15.609, \quad n_{.1} = 19.122, \quad n = 27.404$$

$n_{23}$  = gibt es nicht!

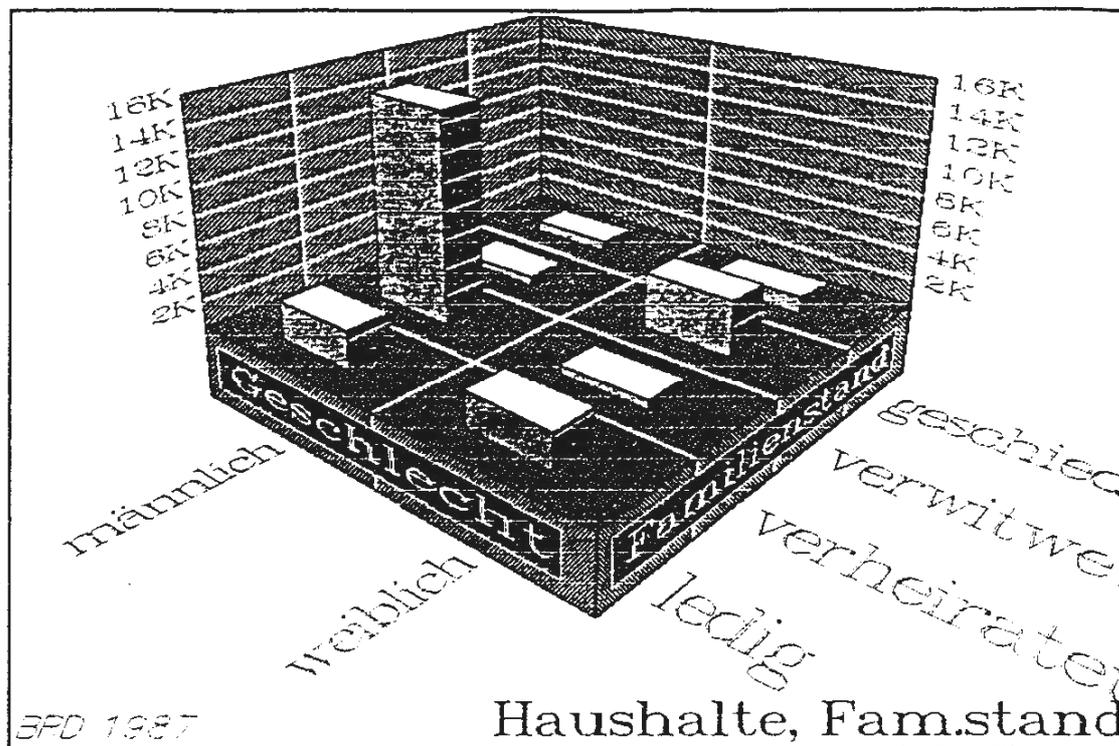


Abb. IV.1: Privathaushalte in der BRD (Mai 1987) nach Familienstand und Geschlecht des Haushaltsvorstands (vgl. Tab. IV.1)

## 1.2 Randverteilungen

Randverteilungen = eindimensionale Verteilungen eines Merkmals, marginale Häufigkeiten

### Randverteilung des 1. Merkmals ( $x$ ) (Zeilensummen)

$$n_{i.} = \sum_{j=1}^m n_{ij} \quad (\text{marginale absolute Häufigkeiten})$$

$$h_{i.} = \frac{n_{i.}}{n} \quad (\text{marginale relative Häufigkeiten})$$

### Randverteilungen des 2. Merkmals ( $y$ ) (Spaltensummen)

$$n_{.j} = \sum_{i=1}^m n_{ij} \quad (\text{marginale absolute Häufigkeiten})$$

$$h_{.j} = \frac{n_{.j}}{n} \quad (\text{marginale relative Häufigkeiten})$$

**Beispiel:**

Privathaushalte BRD 1987 nach Familienstand (x) und Geschlecht (y) (vgl. Tab IV.1)

Familienstand:

$$\begin{array}{ll} n_{1.} = 5158 & h_{1.} = 0,1882 \\ n_{2.} = 15609 & h_{2.} = 0,5696 * \\ n_{3.} = 4682 & h_{3.} = 0,1709 \\ n_{4.} = 1955 & h_{4.} = 0,0713 \\ \hline n = 27404 & h_{..} = 1,0 \end{array}$$

\*56,96% sind verheiratet

Geschlecht:

$$\begin{array}{ll} n_{.1} = 19122 & h_{.1} = 0,6978 \\ n_{.2} = 8282 & h_{.2} = 0,3022 \\ \hline n = 27404 & h_{..} = 1,0 \end{array}$$

### 1.3 Bedingte Verteilungen

Neben des Bezugs auf die Gesamtsumme aller Merkmalsausprägungen n auch Bezüge auf die jeweilige Randverteilung

#### Bedingte Verteilung

Bezug auf die Gesamtheit **einer** Zeile oder Spalte.

Bei gegebener Ausprägung des einen Merkmals werden den Ausprägungen des anderen Merkmals relative Häufigkeiten zugeordnet:

eindimensionale Häufigkeitsverteilung von x bei gegebenem  $y_j$ , wobei  $x_i$  = Wirkung,  $y_j$  = Ursache:

$$h(x_i | y_j) = \frac{n(x_i, y_j)}{n(y_j)} = \frac{n_{ij}}{n_{.j}}$$

eindimensionale Häufigkeitsverteilung von y bei gegebenem  $x_i$ , wobei  $y_j$  = Wirkung,  $x_i$  = Ursache:

$$h(y_j | x_i) = \frac{n(x_i, y_j)}{n(x_i)} = \frac{n_{ij}}{n_{i.}}$$

Die bedingten Verteilungen liefern nur dann mehr Informationen, wenn sie sich unterscheiden. Sind alle bedingten Verteilungen gleich, dann sind sie auch identisch mit der Randverteilung.

### Statistische Unabhängigkeit

Merkmale sind statistisch voneinander unabhängig, wenn alle bedingten Verteilungen **gleich** der entsprechenden Randverteilung sind. Die Verteilung des Merkmals  $x$  ist dann unabhängig von spezieller Ausprägung des Merkmals  $y$ :

$$h(x_i | y_1) = h(x_i | y_2) = \dots = h(x_i | y_m) = h(x_i) = \frac{n_{i.}}{n}$$

Weiter gilt bei statistischer Unabhängigkeit:

$$\frac{n_{ij}}{n_{.j}} = \frac{n_{i.}}{n} \rightarrow n_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

$$\rightarrow \frac{n_{ij}}{n} = \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} \quad \text{oder}$$

$$h(x_i, y_j) = h(x_i) \cdot h(y_j)$$

### Aussagen

1. Ist  $x$  von  $y$  statistisch unabhängig, so ist auch  $y$  von  $x$  statistisch unabhängig (symmetrische Beziehung).
2. Sind  $x$  und  $y$  statistisch unabhängig, so sind die bedingten Häufigkeitsverteilungen gleich der zugehörigen Randverteilung.

### Beispiel:

Wohnbevölkerung der BRD (1987) nach der Beteiligung am Erwerbsleben und dem Geschlecht (in 1.000)

Tab. IV.2: Wohnbevölkerung der BRD (1987) nach der Beteiligung am Erwerbsleben und dem Geschlecht (in 1.000), Häufigkeiten ( $n_{ij}$ )

Beteiligung am Erwerbsleben $x$	Geschlecht $y$		Zeilensumme
	männlich	weiblich	
Erwerbspersonen	17.834	11.160	28.994
Nicht-Erwerbspersonen	11.489	20.594	32.083
Spaltensumme	29.323	31.754	61.077

Quelle: Statistisches Jahrbuch der BRD 1989, S. 89, Volkszählung 1987

$n = 61.077$ ,  $n_{22} = 20.594$  (weibliche Nicht-Erwerbspersonen)

Häufigkeiten  $h(x_i, y_j)$  (in %):

Beteiligung am Erwerbsleben x	Geschlecht y		Zeilensumme
	männlich	weiblich	
Erwerbspersonen	29,2	18,3	47,5
Nicht-Erwerbspersonen	18,8	33,7	52,5
Spaltensumme	48,0	52,0	100,0

$h(x_2, y_1) = \frac{n_{21}}{n} = 18,8 \%$  sind männliche Nicht-Erwerbspersonen.

$h(x_1) = \frac{n_{1.}}{n} = 47,5 \%$  der Wohnbevölkerung sind Erwerbspersonen.

Bedingte Verteilung:

Frage: Wie gliedert sich die Beteiligung der männlichen und weiblichen Personen am Erwerbsleben auf?  $\rightarrow h(x_i | \text{Geschlecht})$

von 100 Personen, die das Geschlecht ... hatten, wa- ren	Geschlecht y		Randverteilung
	männlich	weiblich	
Erwerbspersonen	60,8	35,2	47,5
Nicht-Erwerbspersonen	39,2	64,8	52,5
Spaltensumme	100,0	100,0	100,0

$h(x_2 | y_1) = h(\text{Nicht-Erwerbspersonen} | \text{alle Männer}) = \frac{n_{21}}{n_{1.}} = 39,2 \%$

$h(x_1 | y_2) = h(\text{Erwerbspersonen} | \text{alle Frauen}) = \frac{n_{12}}{n_{2.}} = 35,2 \%$

Frage: Wie gliedert sich der Erwerbsstatus auf das Geschlecht auf?  $\rightarrow (h | y_j | \text{Erwerbsstatus})$

von 100 Personen, die den Erwerbsstatus ... hatten, waren	Geschlecht y		Zeilensumme
	männlich	weiblich	
Erwerbspersonen	61,5	38,5	100,0
Nicht-Erwerbspersonen	35,8	64,2	100,0
Randverteilung	48,0	52,0	100,0

$h(y_1 | x_1) = h(\text{Männer} | \text{alle Erwerbspersonen}) = \frac{n_{11}}{n_{1.}} = 61,5 \%$

$h(y_2 | x_2) = h(\text{Frauen} | \text{alle Nichterwerbspersonen}) = \frac{n_{22}}{n_{2.}} = 64,2 \%$

Wäre die Beteiligung am Erwerbsleben unabhängig vom Geschlecht, so müsste die Randverteilung gleich den prozentualen Besetzungszahlen für die einzelnen Geschlechter sein.

Fiktive Verteilung der Beteiligung am Erwerbsleben bei Unabhängigkeit vom Geschlecht:

von 100 Personen, die das Geschlecht ... hatten, waren	Geschlecht y		Randverteilung
	männlich	weiblich	
Erwerbspersonen	47,5	47,5	47,5
Nicht-Erwerbspersonen	52,5	52,5	52,5
Spaltensumme	100,0	100,0	100,0

Statistisch unabhängig?

$$h(x_i | y_1) = h(x_i | y_2) = h(x_i) = \frac{n_i}{n} \quad (i = 1, 2)$$

$$h(x_1 | y_1) = 60,8 \% \neq \tilde{h}(x_1 | y_1) = 47,5 \% \Rightarrow$$

Die Beteiligung am Erwerbsleben ist **nicht unabhängig** vom Geschlecht.

## 2 Korrelationsrechnung

### 2.1 Zusammenhangsmaße

Zusammenhang zwischen den Merkmalsausprägungen verschiedener Merkmale; aus der Vielzahl der Maßzahlen des Zusammenhangs (Kontingenz, Assoziation, Korrelation) werden folgende drei Konzepte näher betrachtet:

Korrelation zwischen

- nominal skalierten Merkmalen  
Kontingenzanalyse und Kontingenzkoeffizient
- ordinal skalierten Merkmalen  
Rangkorrelation nach Spearman
- metrisch skalierten Merkmalen  
Bravais-Pearson-Korrelationskoeffizient

### 2.2 Korrelation zwischen nominal skalierten Merkmalen: Kontingenzanalyse und Kontingenzkoeffizient

Wie die Analyse im letzten Abschnitt gezeigt hat, gilt bei Abhängigkeit der Merkmale

$$n_{ij} \neq \frac{n_{i \cdot} \cdot n_{\cdot j}}{n} .$$

Damit besteht eine Differenz zwischen beobachteter Häufigkeit  $n_{ij}$  und fiktiver Häufigkeit bei Unabhängigkeit  $\tilde{n}_{ij}$ :

$$\tilde{n}_{ij} = \frac{n_{i \cdot} \cdot n_{\cdot j}}{n}$$

### Quadratische Kontingenz $\chi^2$ (Chi-Quadrat)

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}},$$

wobei

$n_{ij}$  = beobachtete Häufigkeit

$\tilde{n}_{ij}$  = Häufigkeit bei Unabhängigkeit

$\chi^2 = 0 \Rightarrow$  statistische Unabhängigkeit

Nachteil:  $\chi^2$  ist nicht normiert und kann daher unbegrenzt große Werte annehmen. Deshalb verwendet man besser:

### Kontingenzkoeffizient $K^*$

$$K^* = \sqrt{\frac{\chi^2}{n + \chi^2} \cdot \frac{M}{M - 1}},$$

wobei

$M = \min(k, m)$  mit

$k$  = Zeilenanzahl

$m$  = Spaltenanzahl

Wertebereich:  $0 \leq K^* \leq 1$ , damit ist  $K^*$  normiert zwischen 0 und 1.

### Beispiel:

Gliederung der Wohnbevölkerung der BRD 1987 nach Beteiligung am Erwerbsleben und nach Geschlecht:

$$\begin{aligned}\tilde{n}_{11} &= \frac{n_{1.} \cdot n_{.1}}{n} = \frac{28.994 \cdot 29.323}{61.077} = 13.920 \\ \tilde{n}_{12} &= \frac{n_{1.} \cdot n_{.2}}{n} = \frac{28.994 \cdot 31.754}{61.077} = 15.074 \\ \tilde{n}_{21} &= \frac{n_{2.} \cdot n_{.1}}{n} = \frac{32.083 \cdot 29.323}{61.077} = 15.403 \\ \tilde{n}_{22} &= \frac{n_{2.} \cdot n_{.2}}{n} = \frac{32.083 \cdot 31.754}{61.077} = 16.680\end{aligned}$$

Kontingenztabellen

Beobachtete absolute Häufigkeit  $n_{ij}$ :

Beteiligung am Erwerbsleben $n_{ij}$	Geschlecht		Zeilensumme
	männlich	weiblich	
Erwerbspersonen	17.834	11.160	28.994
Nicht-Erwerbspersonen	11.489	20.594	32.083
Spaltensumme	29.323	31.754	61.077

Fiktive absolute Häufigkeit Ausgang bei statistischer Unabhängigkeit  $\tilde{n}_{ij}$ :

Beteiligung am Erwerbsleben $n_{ij}$	Geschlecht		Zeilensumme
	männlich	weiblich	
Erwerbspersonen	13.920	15.074	28.994
Nicht-Erwerbspersonen	15.403	16.680	32.083
Spaltensumme	29.323	31.754	61.077

Für  $k = 2$  und  $m = 2$ :

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}} = 1100,53 + 1016,28 + 994,57 + 918,43 = 4029,81$$

Für  $M = \min(2,2) = 2$ :

$$K^* = \sqrt{\frac{\chi^2}{n + \chi^2} \cdot \frac{M}{M-1}} = \sqrt{\frac{4029,81}{61077 + 4029,81} \cdot \frac{2}{2-1}} = \sqrt{0,123791} = 0,35 \quad (0 \leq K^* \leq 1)$$

Interpretation: Beteiligung am Erwerbsleben ist geschlechtsabhängig (Vergleich mit anderen Jahren, Ländern ist aussagekräftiger).

### 2.3 Korrelation zwischen ordinal-skalierten Merkmalen: Rangkorrelationskoeffizient nach Spearman

Ordinal-skalierte Merkmale, daher Rangnummern (-plätze)

$R_i$ : Rangnummern des 1. Merkmals ( $x$ )

$R'_i$ : Rangnummern des 2. Merkmals ( $y$ )

Jeder statistischen Einheit  $i$  ( $i = 1, \dots, n$ ) werden Rangnummern beider Merkmalsausprägungen  $R_i$  und  $R'_i$  zugeordnet. Gibt es übereinstimmende Beobachtungswerte für mehrere statistische Einheiten, dann wird hier das arithmetische Mittel der diesen Werten zuzuordnenden Nummern als Rangzahl zugeordnet.

### Spearman'scher Rangkorrelationskoeffizient

$$r_{Sp} = 1 - \frac{6 \sum_{i=1}^n (R_i - R'_i)^2}{(n-1)n(n+1)},$$

Normiert auf  $[-1 \leq r_{Sp} \leq +1]$

$r_{Sp} = -1$ : Die Ränge verhalten sich völlig gegenläufig.  
 $(R_i = n+1 - R'_i; i = 1, \dots, n)$ ;  
 negativer Zusammenhang

$r_{Sp} = 0$ : kein Zusammenhang

$r_{Sp} = +1$ : Die Ränge verhalten sich völlig gleichläufig.  
 $(R_i = R'_i; i = 1, \dots, n)$   
 positiver Zusammenhang

#### Beispiel:

Für zehn Angestellte wurden organisatorische Geschicklichkeit ( $x$ ) und Arbeitssorgfalt ( $y$ ) ermittelt.

Rangziffern (-plätze):

Angestellter i	1	2	3	4	5	6	7	8	9	10
x: $R_i$	7	3	9	10	1	5	4	6	2	8
y: $R'_i$	3	9	10	8	7	1	5	4	2	6

$$r_{Sp} = 1 - \frac{6 \sum_{i=1}^{10} (R_i - R'_i)^2}{(10-1) \cdot 10 \cdot (10+1)} = 1 - \frac{6 \cdot 118}{9 \cdot 10 \cdot 11} = 0,28$$

Interpretation: Es liegt eine schwach gleichläufige (positive) Korrelation vor.

## 2.4 Korrelation zwischen metrisch-skalierten Merkmalen: Bravais-Pearson-Korrelationskoeffizient

Metrisch-skalierte Merkmale in ungruppiertes und gruppiertes Form

### Ungruppiertes Datenmaterial:

Gegeben:  $n$  Beobachtungspaare von Merkmalsausprägungen

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$$

**Mittelwerte**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

**Varianzen**

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

**Kovarianz**

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y} = \overline{xy} - \bar{x} \cdot \bar{y}$$

Die Kovarianz stellt eine Beziehung zwischen x und y her (vgl. Abb. IV.2):

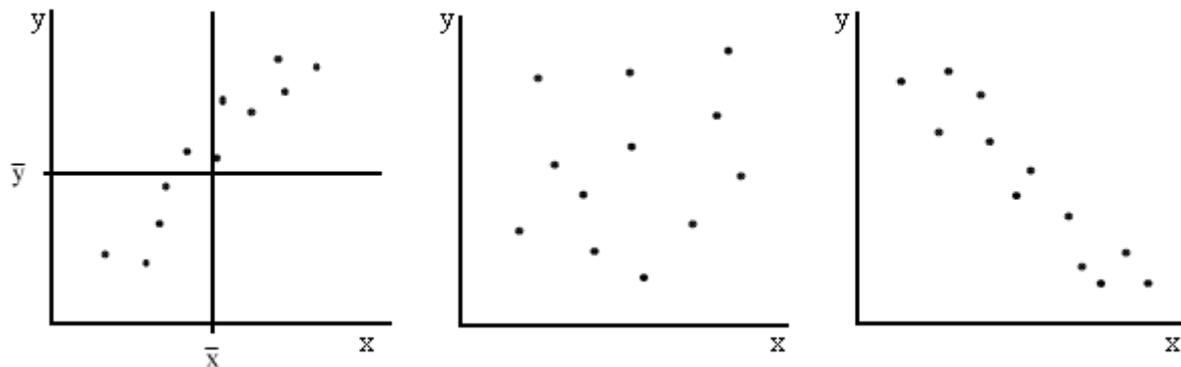


Abb. IV.2: Kovarianzen

Kreuzprodukte:  $(x_i - \bar{x})(y_i - \bar{y})$  positiv,  $\approx 0$  oder negativ

$s_{xy}$  wird umso größer, je stärker die Wertepaare überwiegen, bei denen große x-Werte mit großen y-Werten und kleine x- mit kleinen y-Werten gekoppelt sind.

Die Kovarianz ist ein Maß für die Stärke des Zusammenhangs zwischen zwei Merkmalen x und y. Normiert ergibt sich:

**Bravais-Pearson-Korrelationskoeffizient**

$$r = \frac{s_{xy}}{s_x s_y}$$

(Normierung auf das Produkt der Standardabweichungen)

r liegt im Intervall  $[-1 \leq r \leq +1]$

Interpretation:

$r = -1$	extrem starker negativer Zusammenhang
$r = 0$	keine Korrelation
$r = +1$	extrem starker positiver Zusammenhang

**Beispiel:**

Umsatz und Werbeausgaben eines Industrieunternehmens

Tab. IV.3: Umsatz und Werbeausgaben der Firma IXWHYZET

Jahr	Umsatz y (Mio. EUR)	Werbeausgaben x (Mio. EUR)
1996	17,0	1,4
1997	17,6	1,7
1998	17,5	1,6
1999	18,1	1,8
2000	18,7	2,0
2001	19,1	1,9
2002	19,0	2,0
2003	20,5	2,2
2004	21,8	2,0
2005	21,3	2,1
2006	26,5	2,5
2007	25,8	3,0
2008	26,3	2,8
2009	27,8	3,2
2010	30,0	3,0

Werbeausgaben:  $\bar{x} = 2,21$   $s_x^2 = 0,29$

Umsatz:  $\bar{y} = 21,80$   $s_y^2 = 17,40$

Kovarianz:  $s_{xy} = 2,20$

Korrelationskoeffizient:  $r = \frac{2,20}{0,54 \cdot 4,17} = 0,977$

Es besteht also ein starker positiver Zusammenhang, d.h. hohe (geringe) Werbeausgaben korrelieren mit hohen (geringen) Umsätzen.

**ET: Zusammenhangsanalyse**

ET: Histogram, Scatter

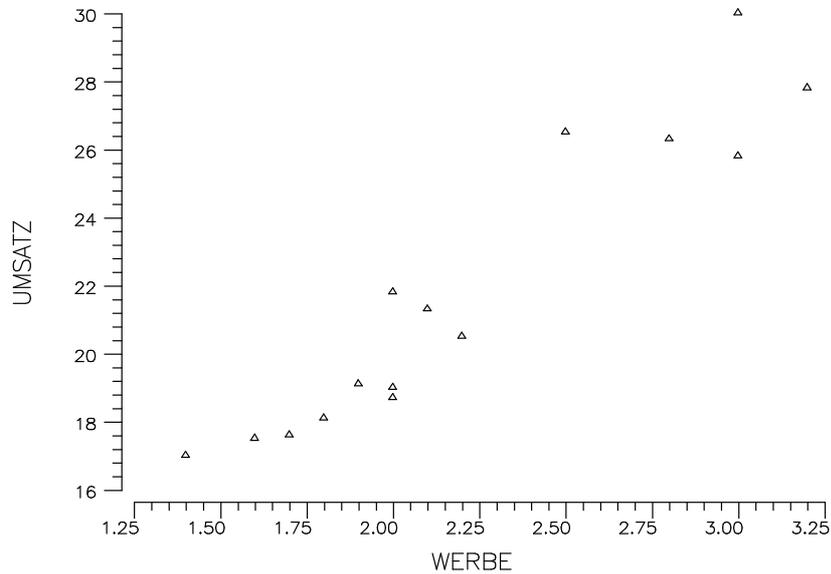


Abb. IV.3:ET: Scatterplot der Umsätze und Werbeausgaben der Firma IXWHYZET

ET: Histogram, Crosstab

UMSATZ	WERBE					Total	Chi-squared[ 0]= .0000, P= .00000
	0	1	2	3	4		
0	3	4	0	0	0	7	
1	0	2	1	0	0	3	
2	0	0	0	0	0	0	
3	0	0	0	2	1	3	
4	0	0	0	0	2	2	
Total	3	6	1	2	3	15	

Classes	UMSATZ /Out of range= 0	WERBE /Out of range= 0
0	16.9900 to 19.6000	1.3900 to 1.7600
1	19.6000 to 22.2000	1.7600 to 2.1200
2	22.2000 to 24.8000	2.1200 to 2.4800
3	24.8000 to 27.4000	2.4800 to 2.8400
4	27.4000 to 30.0100	2.8400 to 3.2100

**Gruppiertes Datenmaterial:**

Die Daten werden in Klassen eingeteilt und die Klassenmitten  $x_i^*$  und  $y_j^*$  ermittelt.

**Mittelwerte**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i^* \cdot n_i.$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^m y_j^* \cdot n_{.j}$$

**Varianzen**

$$s_x^2 = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot n_i = \frac{1}{n} \sum_{i=1}^k (x_i^*)^2 \cdot n_i - \bar{x}^2$$

$$s_y^2 = \frac{1}{n} \sum_{j=1}^m (y_j^* - \bar{y})^2 \cdot n_{.j} = \frac{1}{n} \sum_{j=1}^m (y_j^*)^2 \cdot n_{.j} - \bar{y}^2$$

**Kovarianz**

$$s_{xy} = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})(y_i^* - \bar{y}) \cdot n_i = \frac{1}{n} \sum_{i=1}^k x_i^* y_i^* \cdot n_i - \bar{x}\bar{y}$$

**Korrelationskoeffizient**

$$r = \frac{s_{xy}}{s_x \cdot s_y} \quad -1 \leq r \leq 1$$

**Beispiel:**

Aufwendungen für Forschung und Entwicklung FuE (R&D = RESEARCH AND DEVELOPMENT) von Unternehmen (Tabelle IV.4)

Tab. IV.4: Aufwendungen für Forschung und Entwicklung (FuE) von Unternehmen

Umsatz x	Aufwendungen für Forschung und Entwicklung (FuE) y					Zeilensumme
	5 ( $y_1^*$ )	15 ( $y_2^*$ )	25 ( $y_3^*$ )	35 ( $y_4^*$ )	45 ( $y_5^*$ )	
100 ( $x_1^*$ )	2	3	1	-	-	6 = $n_1$
300 ( $x_2^*$ )	2	6	3	1	-	12 = $n_2$
500 ( $x_3^*$ )	1	4	5	4	-	14 = $n_3$
700 ( $x_4^*$ )	-	2	4	3	2	11 = $n_4$
900 ( $x_5^*$ )	-	-	1	2	4	7 = $n_5$
Spaltensumme	5 $n_{.1}$	15 $n_{.2}$	14 $n_{.3}$	10 $n_{.4}$	6 $n_{.5}$	50 = n

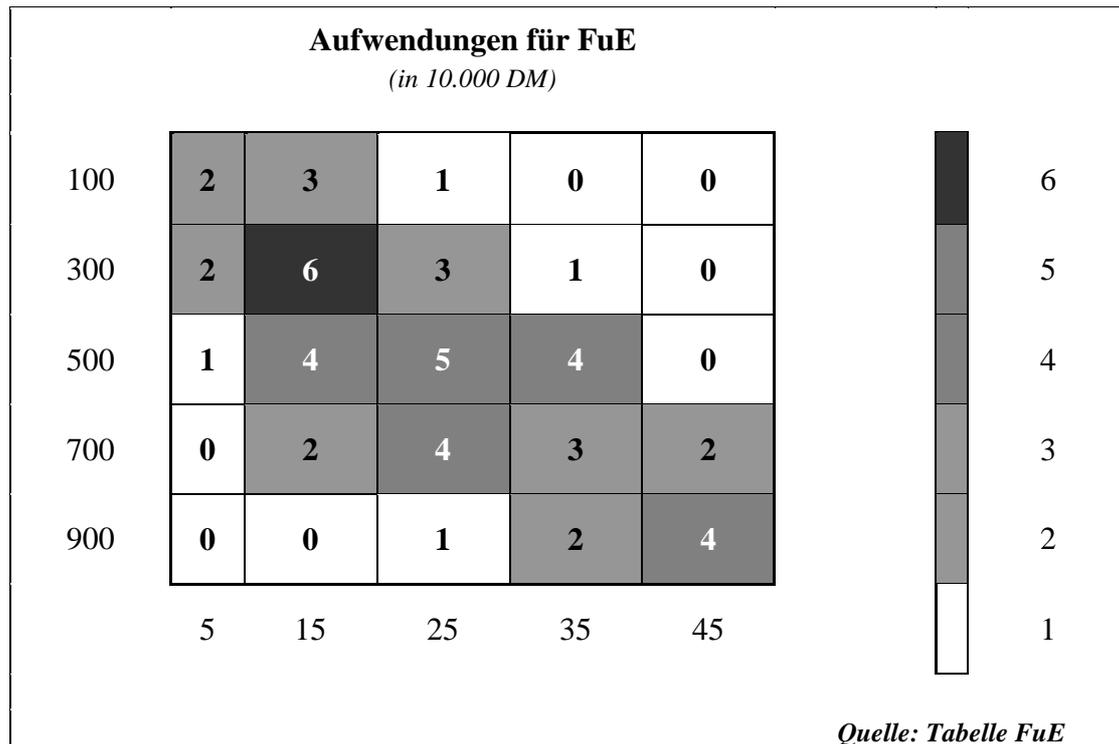


Abb. IV.4: Aufwendungen für FuE: Spektraldarstellung

$$\bar{x} = \frac{1}{50} (100 \cdot 6 + 300 \cdot 12 + 500 \cdot 14 + 700 \cdot 11 + 900 \cdot 7) = 504$$

$$\bar{y} = \frac{1}{50} (5 \cdot 5 + 15 \cdot 15 + 25 \cdot 14 + 35 \cdot 10 + 45 \cdot 6) = 24,4$$

$$s_x^2 = \frac{1}{50} (100^2 \cdot 6 + 300^2 \cdot 12 + 500^2 \cdot 14 + 700^2 \cdot 11 + 900^2 \cdot 7) - (504^2) = 59984$$

$$s_x = 244,916$$

$$s_y^2 = \frac{1}{50} (5^2 \cdot 5 + 15^2 \cdot 15 + 25^2 \cdot 14 + 35^2 \cdot 10 + 45^2 \cdot 6) - (24,4^2) = 137,64$$

$$s_y = 11,732$$

$$s_{xy} = 1/50 (100 \cdot 5 \cdot 2 + 100 \cdot 15 \cdot 3 + 100 \cdot 25 \cdot 1 + 300 \cdot 5 \cdot 2 + 300 \cdot 15 \cdot 6 + 300 \cdot 25 \cdot 3 + 300 \cdot 35 \cdot 1 + 500 \cdot 5 \cdot 1 + 500 \cdot 15 \cdot 4 + 500 \cdot 25 \cdot 5 + 500 \cdot 35 \cdot 4 + 700 \cdot 15 \cdot 2 + 700 \cdot 25 \cdot 4 + 700 \cdot 35 \cdot 3 + 700 \cdot 45 \cdot 2 + 900 \cdot 25 \cdot 1 + 900 \cdot 35 \cdot 2 + 900 \cdot 45 \cdot 4) - 504 \cdot 24,4 = 1922,4$$

$$r = \frac{1922,4}{244,916 \cdot 11,732} = 0,669$$

Hier liegt also eine positive Korrelation vor.

### 3 Regressionsrechnung

#### Korrelation

Aussagen über die Stärke des Zusammenhangs zwischen den Variablen; keine Unterscheidung zwischen abhängigen (endogenen) und unabhängigen (exogenen) Variablen:

Interdependenzanalyse

#### Regression

Aussagen über funktionale Form der Abhängigkeit zwischen den Variablen; **gerichtete, 'kausale' Analyse** zwischen abhängigen und unabhängigen Variablen

Dependenzanalyse

$$\underbrace{y}_{\text{(lhs) 'left hand side'}} = \underbrace{f(x)}_{\text{(rhs) 'right hand side'}}$$

$y$  ist die abhängige Variable (Regressand, erklärte/ endogene Variable).

$x$  ist die unabhängige Variable (eine oder mehrere Variablen), auch Regressor genannt

#### Längsschnitt, Makro

$$\text{Konsum}_t = f(\text{Volkseinkommen}_t) \quad t = 1, \dots, T$$

Für jedes Jahr  $t$  ein Beobachtungspaar

#### Querschnitt, Mikro

$$\text{Konsum}_i = f(\text{Einkommen}_i) \quad i = 1, \dots, n$$

Für jede Person  $i$  ein Beobachtungspaar

#### Merke

Jeweiliger Beobachtungsträger (Jahr oder Person) ist (in aller Regel) ungeordnet in der Punktwolke; es kommt nur auf  $(x_i, y_i)$  an, (vgl. Abb. IV.5).

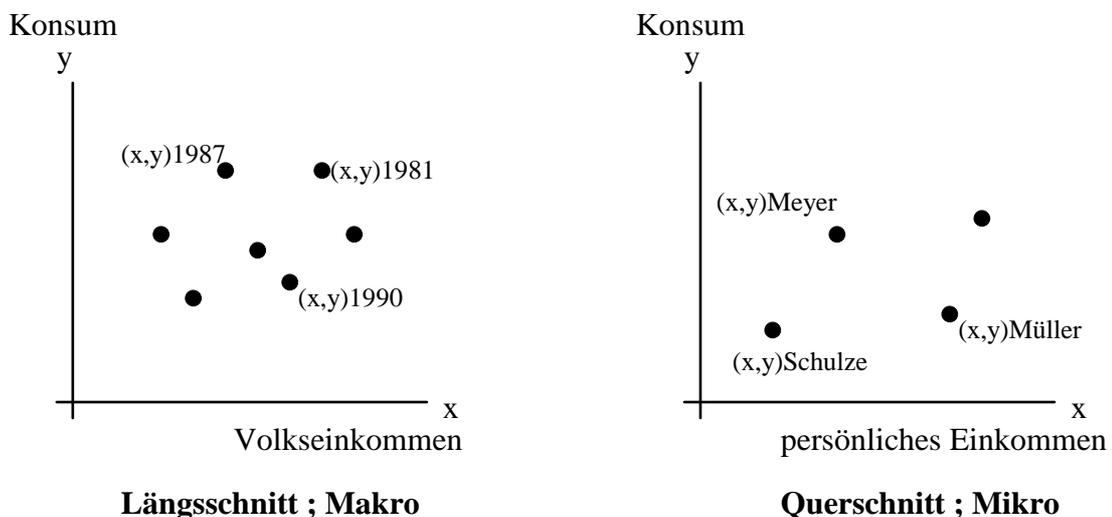


Abb. IV.5: Scatterdiagramm von Konsum und Einkommen im Längs- und Querschnitt

**Gesucht**

Beziehung zwischen  $x$  und  $y$  als mittlere Tendenz, Ausgleichsgerade

**3.1 Lineare Regression**

Einfachregression:  $y = f(x)$

Mehrfachregression:  $y = f(x_1, x_2, \dots, x_k)$

**Einfachregression:**

Geg.:  $n$  Beobachtungspaare  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

Man legt dann ein Streudiagramm mit der Punktwolke ('scatter diagram') und der Ausgleichsgerade an:

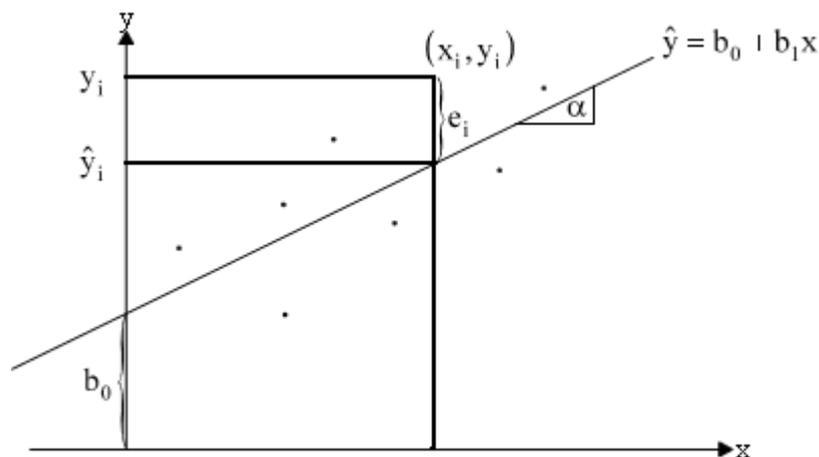


Abb. IV.6: Streudiagramm ('scatterplot') von  $n$  Beobachtungspaaren

Gesucht: Ausgleichsline des mittleren Zusammenhangs zwischen den Punkten  $(x_i, y_i) (i = 1, \dots, n)$

Regressionsansatz:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Stochastischer Ansatz:  $\beta_0, \beta_1$  zu schätzende Koeffizienten der Grundgesamtheit,  $\varepsilon$  Fehlerterm (siehe Merz (1994e), Skriptum Statistik II).

Im weiteren: deterministischer Ansatz allein als Ausgleichsgeradenproblem

$$\hat{y} = b_0 + b_1 x \quad b_0, b_1 \text{ gesuchte Regressionkoeffizienten, die die Lage der Ausgleichsgerade bestimmen}$$

$$y = \hat{y} + e = b_0 + b_1 x + e,$$

wobei  $b_0, b_1$  die (bspw. nach der Methode der kleinsten Quadrate (MKQ, Ordinary Least Squares = OLS) geschätzten Regressionskoeffizienten und  $e_i = y_i - \hat{y}_i$  die Abweichung zwischen beobachtetem Wert und geschätztem Wert des Merkmalsträger  $i$  ist.

Die Abweichungen  $e_i$  werden als Residuen bezeichnet.

$$e_i = y_i - \hat{y}_i$$

Man bestimmt die gesuchten Regressionskoeffizienten  $b_0$  und  $b_1$  mit der

**Methode der kleinsten Quadrate (MKQ, Gauß).** Häufig findet man auch die Bezeichnung OLS (Ordinary Least Squares).

Die Schätzung erfolgt über die **Minimierung der Summe der quadrierten Abweichungen:**

$$Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (b_0 + b_1 x_i))^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = \min!$$

Quadrierungseffekt:

- positive und negative Abweichungen zählen (heben sich nicht auf)
- größere Abweichungen erhalten höheres Gewicht

$Q(b_0, b_1)$  ist bei gegebenen  $(x_i, y_i)$  also zu minimieren.

**Minimierungsproblem ist ein Extremwertproblem**

Die optimalen Schätzungen für  $b_0$  und  $b_1$  sind dort, wo die ersten Ableitungen Null sind.

**Partielle Ableitungen**

$$\frac{\partial Q}{\partial b_0} = \sum_{i=1}^n [2(y_i - b_0 - b_1 x_i)^{2-1} \cdot (-1)] = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) \quad \text{Ableitung nach der Kettenregel}$$

$$\frac{\partial Q}{\partial b_1} = \sum_{i=1}^n [2(y_i - b_0 - b_1 x_i) \cdot (-x_i)] = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) \cdot x_i$$

**Nullsetzen**

$$\frac{\partial Q}{\partial b_0} = 0 \quad \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial Q}{\partial b_1} = 0 \quad \sum_{i=1}^n x_i \cdot (y_i - b_0 - b_1 x_i) = 0$$

Normalgleichungen

$$\sum_{i=1}^n b_0 + \sum_{i=1}^n b_1 x_i = \sum_{i=1}^n y_i \quad n b_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n b_0 x_i + \sum_{i=1}^n b_1 x_i^2 = \sum_{i=1}^n y_i x_i \quad b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i$$

Aufgelöst nach  $b_0$  und  $b_1$  ergibt sich:

$$b_0 = \frac{1}{n} \sum y_i - b_1 \cdot \frac{1}{n} \sum x_i = \bar{y} - b_1 \cdot \bar{x}$$

$$b_1 = \frac{\frac{1}{n} \sum y_i x_i - \bar{y} \cdot \bar{x}}{\frac{1}{n} \sum x_i^2 - \bar{x}^2} = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x}$$

$$b_0 = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - \sum x \sum x} = \frac{\bar{y} \frac{1}{n} \sum x^2 - \bar{x} \frac{1}{n} \sum xy}{\frac{1}{n} \sum x^2 - \bar{x}^2}$$

Damit sind  $b_0$  und  $b_1$  die geschätzten Regressionskoeffizienten im Optimum (Minimum der Fehlerquadratsumme).

### Beispiel:

Eine zu einem bestimmten Zeitpunkt durchgeführte Umfrage (Querschnitt) liefere folgende Mikrodaten mit monatlichem Nettoeinkommen ( $y_i$ ) in 100 EUR und Alter der Personen ( $x_i$ ) in Jahren ( $n = 6$ ).

Tab. IV.5: Mikrodaten zu Nettoeinkommen und Alter der Personen

i	Daten		Arbeitstabelle		
	$y_i$	$x_i$	$x_i^2$	$x_i \cdot y_i$	$y_i^2$
1	6	21	441	126	36
2	15	28	784	420	225
3	5	20	400	100	25
4	31	44	1936	1364	961
5	22	36	1296	792	484
6	11	25	625	275	121
$\Sigma$	90	174	5482	3077	1852

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 29 \text{ Jahre} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 15 \quad (\hat{=} 1500 \text{ € mtl})$$

$$b_1 = \frac{\frac{1}{n} \sum y_i \cdot x_i - \bar{y} \cdot \bar{x}}{\frac{1}{n} \sum x_i^2 - \bar{x}^2} = \frac{\frac{1}{6} \cdot 3077 - 15 \cdot 29}{\frac{1}{6} \cdot 5482 - 29^2} = \frac{77,833}{72,666} = 1,071$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 15 - 1,071 \cdot 29 = -16,06$$

$$\text{Regressionsgerade: } \hat{y}_i = -16,06 + 1,071 \cdot x_i$$

Interpretation:

Bei einer Zunahme des Alters  $x$  um 1 Jahr steigt das Einkommen im Durchschnitt um 107,10 EUR (im Wertebereich 20-44 Jahre).

### ET Histogram, Scatter, Regression (F7, F8)

Grafische Darstellung der Regressionsgeraden (Abb.IV.6)

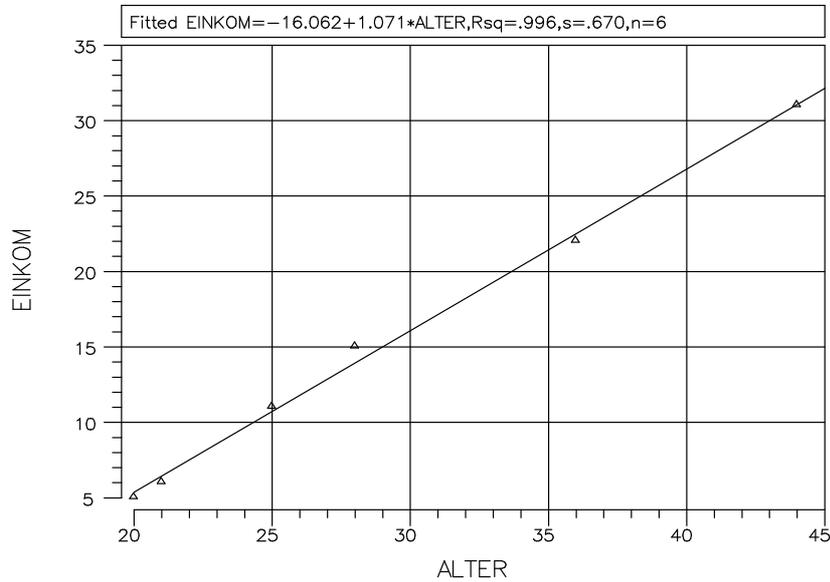


Abb. IV.7: Scatterplot mit Regressionsgerade der Umfragedaten zu Nettoeinkommen und Alter der Personen

Eine Zerlegung der Streuung führt zur Beurteilung der Anpassungsgüte der Regression.

### Streuungszerlegung

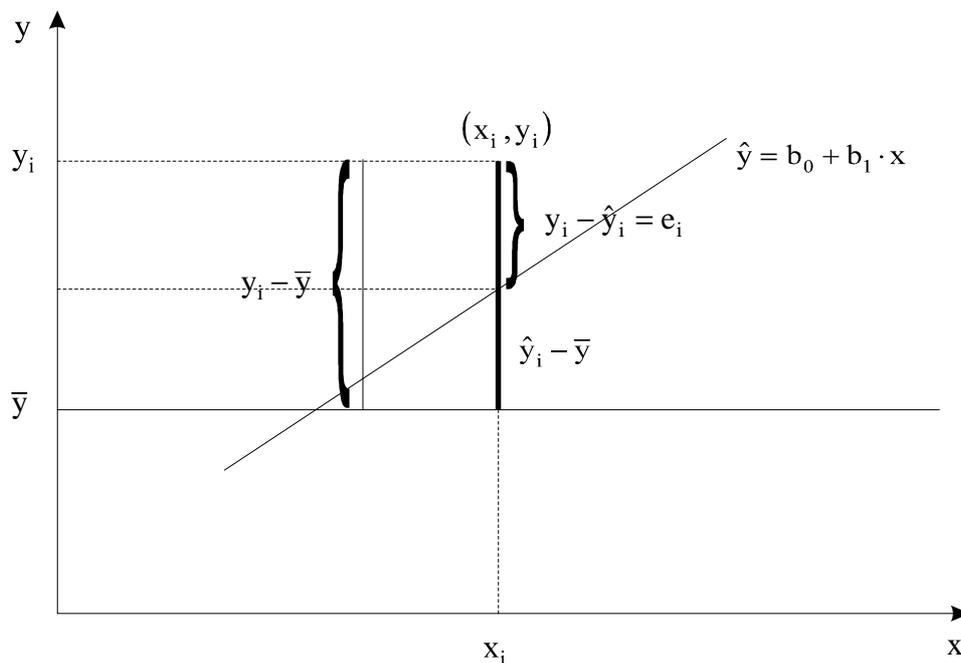


Abb. IV.8: Regressionsgerade und Streuungszersetzung

$$\text{Abweichung} \quad (y_i - \bar{y}) = \underbrace{(\hat{y}_i - \bar{y})}_{\text{zu erklärende}} + \underbrace{(y_i - \hat{y}_i)}_{\text{nicht erklärte}} \quad (i = 1, \dots, n)$$

$$\text{Es gilt} \quad \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Gesamtstreuung = Erklärte + Nicht erklärte Streuung (der  $e_i$ )

Daraus wird ein Maß für die Anpassungsgüte ('goodness of fit') entwickelt:

### Bestimmtheitsmaß

$$B = \frac{\text{erklärte Varianz}}{\text{Gesamt Varianz}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{s_{\hat{y}}^2}{s_y^2}$$

Anteil der durch die Regression erklärten Varianz der y-Werte

Es gilt:  $0 \leq B \leq 1$  (je näher B an 1 liegt, desto besser)

Für die lineare Regression gilt:  $B = r^2$  (= Quadrat des Korrelationskoeffizienten)

### Beispiel:

Umfragedaten mit Nettoeinkommen und Alter der Personen  
Einkommen = f (Alter)

$$B = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

Zähler:

$$\hat{y}_i = -16,06 + 1,071 \cdot x_i$$

$$\sum_i (\hat{y}_i - \bar{y})^2 = 500,2$$

Nenner:

$$\begin{aligned} \sum_i (y_i - \bar{y})^2 &= n \cdot s_y^2 = n \left( \frac{1}{n} \sum_i y_i^2 - \bar{y}^2 \right) \\ &= \sum_i y_i^2 - n \cdot \bar{y}^2 = 1852 - 6 \cdot 15^2 \\ &= 502 \end{aligned}$$

$$B = \frac{500,2}{502} = 0,996$$

Interpretation: 99,6 % der Streuung des Einkommens der sechs Personen können durch die lineare Regression durch das Alter erklärt werden.

Die Korrelation zwischen  $y$  und  $x$  beträgt:  $r = \sqrt{B} = 0,998$ , es existiert also ein starker Zusammenhang.

### Multiple (mehrfache) lineare Regression:

Stochastischer Ansatz:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K + \varepsilon$$

Deterministischer Ansatz:

$$\hat{y} = b_0 + b_1 x_1 + \dots + b_K x_K$$

$$y = \hat{y} + e = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_K x_K + e,$$

wobei  $b_0, b_1, \dots, b_K$  die (bspw.) mit OLS geschätzten Koeffizienten sind.

Minimieren der Fehlerquadratsumme wie bei der einfachen linearen Regression  
Anwendung von MKQ/OLS:

$$Q = \sum_{i=1}^n (y_i - (b_0 + b_1 x_{i1} + \dots + b_K x_{iK}))^2 = \min! \Rightarrow \frac{\partial Q}{\partial b_0} = 0, \frac{\partial Q}{\partial b_1} = 0, \dots, \frac{\partial Q}{\partial b_K} = 0$$

Aus den nullgesetzten 1. partiellen Ableitungen der Fehlerquadratsumme werden die Regressionskoeffizienten  $b_k$  ( $k = 1, \dots, K$ ) bestimmt

$$X' X b = X' y$$

Lösung des  $(k+1)$ -Gleichungssystems bei existierender Inverse  $(X' X)_{(K+1) \times (K+1)}^{-1}$  ergibt

### Vektor der OLS-geschätzten Regressionskoeffizienten

$$b^{OLS} = b = (X' X)^{-1} X' y$$

bei vereinfachter Matrixschreibweise mit:

$$b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{pmatrix} \quad \text{Koeffizientenvektor}$$

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nK} \end{pmatrix} \quad \text{Datenmatrix}$$

$n \times (K+1)$

$$\text{Designmatrix } (X'X) = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} & \dots & \sum x_{iK} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & & \\ \sum x_{i2} & & \sum x_{i2}^2 & & \\ \vdots & & & & \\ \sum x_{iK} & \sum x_{iK}x_{i1} & \dots & & \sum x_{iK}^2 \end{pmatrix}$$

$(K+1) \times (K+1)$

$$\text{Kreuzprodukte } (X'y) = \begin{pmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \vdots \\ \sum x_{iK}y_i \end{pmatrix}$$

$(K+1) \times 1$

**Berechnung einer Inversen  $2 \times 2$  Matrix** (vgl. Merz (1987))

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{Inverse: } A^{-1} = \frac{1}{|A|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

wobei Determinante  $|A| = a_{11}a_{22} - a_{12}a_{21}$

$$A = (X'X) = \begin{pmatrix} 6 & 160 \\ 160 & 4398 \end{pmatrix}$$

$$A^{-1} = \frac{1}{6 \cdot 4398 - 160 \cdot 160} \begin{pmatrix} 4398 & -160 \\ -160 & 6 \end{pmatrix} = \begin{pmatrix} 5,5812 & -0,2030 \\ -0,2030 & 0,7614 \cdot 10^{-2} \end{pmatrix}$$

**Beispiel:** \_\_\_\_\_

Einkommenserklärung durch Alter und Geschlecht der Personen (Querschnitt) aus einer fiktiven Stichprobe mit  $n = 6$ :

Tab. IV.6: Querschnittsdaten zur Einkommenserklärung

$$\Rightarrow X = \begin{pmatrix} 1 & 22 & 1 \\ 1 & 24 & 1 \\ 1 & 28 & 1 \\ 1 & 27 & 0 \\ 1 & 23 & 1 \\ 1 & 36 & 0 \end{pmatrix}$$

$(6 \times 3)$

i	income y	$x_0 = 1$	age $x_1$	sex $x_2$
1	1200	1	22	1
2	1700	1	24	1
3	3500	1	28	1
4	4200	1	27	0
5	1600	1	23	1
6	5200	1	36	0

Man hat also eine Stichprobe vom Umfang  $n = 6$  und  $K = 2$  erklärende Variablen, nämlich age und sex. sex ist eine 'Dummy Variable', die nur die Ausprägungen 0 oder 1 annehmen kann: 0 sei Mann, 1 sei Frau.

Designmatrix

$$(X'X) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 22 & 24 & 28 & 27 & 23 & 36 \\ 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 22 & 1 \\ 1 & 24 & 1 \\ 1 & 28 & 1 \\ 1 & 27 & 0 \\ 1 & 23 & 1 \\ 1 & 36 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 160 & 4 \\ 160 & 4398 & 97 \\ 4 & 97 & 4 \end{pmatrix}$$

(3x6)                      (6x3)                      (3x3)

Erläuterungen zu der (symmetrischen) Designmatrix:

$$6 = \sum_i x_{i0}^2 = \sum_i (x_{i0} = 1)^2 = n$$

$$160 = \sum_i 1 \cdot x_{i1} = \sum_i x_{i1} = \text{Summe 'age'}$$

$$97 = \sum_i x_{i1} x_{i2} = \text{Summe der Kreuzprodukte (age} \cdot \text{sex)}$$

$$4 = \sum_i x_{i2}^2 = \text{Summe 'sex}^2'$$

**Kreuzprodukte:**

$$(X'y) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 22 & 24 & 28 & 27 & 23 & 36 \\ 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1200 \\ 1700 \\ 3500 \\ 4200 \\ 1600 \\ 5200 \end{pmatrix} = \begin{pmatrix} 17400 \\ 502600 \\ 8000 \end{pmatrix} = \begin{pmatrix} \sum_i \text{income} \\ \sum_i \text{age} \cdot \text{income} \\ \sum_i \text{sex} \cdot \text{income} \end{pmatrix}$$

(3,6)                      (6,1)                      (3,1)

## Vektor der OLS-geschätzten Regressionskoeffizienten

$$b = (X'X)^{-1} X'y \quad (X'X) \xrightarrow{\text{Gauß}} (X'X)^{-1} \quad \text{symmetrisch!}$$

$$b = \begin{pmatrix} 16,7000 & -0,514286 & -4,22857 \\ -0,514286 & 0,163265E-01 & 0,118367 \\ -4,22857 & 0,118367 & 1,60816 \end{pmatrix}_{(3,3)} \begin{pmatrix} 17400 \\ 502600 \\ 8000 \end{pmatrix}_{(3,1)}$$

$$= \begin{pmatrix} -1728,57 \\ 204,082 \\ -1220,41 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \quad \hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

$$\text{inc\^o}me = -1728,57 + 204,08 \cdot \text{age} - 1220,41 \cdot \text{sex}$$

### Interpretation

- Im Durchschnitt bringt ein weiteres Altersjahr 204,08 EUR.
- Frauen verdienen im Durchschnitt 1220,41 EUR weniger als Männer.

### ET-Programm

```
? -----
? ET income: ols by matrix algebra
? -----
?
? read data
?   y : income
?   x0: one
?   x1: age
?   x2: sex (0=male,1=female)
?
read; file=income.dat;nvar=3;nobs=6;names=1$
list; income, age, sex$
?
? create X'X (=XSX) und X'y (XSy)
? -----
namelist; X=one,age$
matrix; XSX=xdot(one,age)
        ; XSX=xdot(X)$
matrix; XSy=xdot(X,income)$
?
? create invers of (X'X)
? -----
matrix; XSXINV=ginv(XSX)$
?
? compute ols
? -----
matrix; bols=XSXINV | XSy
        ; bet=xlsq(X,income)$
?
? regression by et
? -----
regres; dep=income; ind=one,age$
```

Ergebnisse: ET Multiple Regression:  $\text{income} = b_0 + b_1 \cdot \text{age}$ 

DATA LISTING (Current sample)

Observation	INCOME	AGE	SEX
1	1200.0	22.000	1.0000
2	1700.0	24.000	1.0000
3	3500.0	28.000	1.0000
4	4200.0	27.000	.00000
5	1600.0	23.000	1.0000
6	5200.0	36.000	.00000

1. Matrix ->  $\text{XSX}=\text{XDOT}(\text{ONE},\text{AGE})$ 

<<<< XSX		>>>> COLUMN	
1		2	
ROW	1	6.00000	160.000
ROW	2	160.000	4398.00

2. Matrix ->  $\text{XSX}=\text{XDOT}(\text{X})$ 

<<<< XSX		>>>> COLUMN	
1		2	
ROW	1	6.00000	160.000
ROW	2	160.000	4398.00

1. Matrix ->  $\text{XSY}=\text{XDOT}(\text{X},\text{INCOME})$ 

<<<< XSY		>>>> COLUMN	
1			
ROW	1	17400.0	
ROW	2	502600.	

1. Matrix ->  $\text{XSXINV}=\text{GINV}(\text{XSX})$ 

<<<< XSXINV		>>>> COLUMN	
1		2	
ROW	1	5.58122	-.203046
ROW	2	-.203046	.761421E-02

1. Matrix ->  $\text{BOLS}=\text{XSXINV}|\text{XSY}$ 

<<<< BOLS		>>>> COLUMN	
1			
ROW	1	-4937.56	
ROW	2	293.909	

2. Matrix ->  $\text{BET}=\text{XLSQ}(\text{X},\text{INCOME})$ 

<<<< BET		>>>> COLUMN	
1			
ROW	1	-4937.56	
ROW	2	293.909	

```

=====
Ordinary Least Squares
Dependent Variable      INCOME      Number of Observations  6
Mean of Dep. Variable  2900.0000  Std. Dev. of Dep. Var.  1634.625339
Std. Error of Regr.    709.7758   Sum of Squared Residuals .201513E+07
R - squared             .84917     Adjusted R - squared    .81146
F( 1, 4)               22.5194    Prob. Value for F      .00900
=====
Variable Coefficient Std. Error t-ratio Prob|t|>x Mean of X Std.Dev.of X
-----
Constant -4937.56  1677.   -2.945  .04220
AGE       293.909  61.93   4.745  .00900  26.66667  5.12510

```

Ergebnisse: ET Multiple Regression:  $\text{income} = b_0 + b_1 \cdot \text{age} + b_2 \cdot \text{sex}$

DATA LISTING (Current sample)

Observation	INCOME	AGE	SEX
1	1200.0	22.000	1.0000
2	1700.0	24.000	1.0000
3	3500.0	28.000	1.0000
4	4200.0	27.000	.00000
5	1600.0	23.000	1.0000
6	5200.0	36.000	.00000

1. Matrix ->  $\text{XSX} = \text{XDOT}(\text{ONE}, \text{AGE}, \text{SEX})$

```

<<<< XSX      >>>> COLUMN
          1          2          3
ROW  1  6.00000  160.000  4.00000
ROW  2  160.000  4398.00  97.0000
ROW  3  4.00000  97.0000  4.00000

```

2. Matrix ->  $\text{XSX} = \text{XDOT}(X)$

```

<<<< XSX      >>>> COLUMN
          1          2          3
ROW  1  6.00000  160.000  4.00000
ROW  2  160.000  4398.00  97.0000
ROW  3  4.00000  97.0000  4.00000

```

1. Matrix ->  $\text{XSY} = \text{XDOT}(X, \text{INCOME})$

```

<<<< XSY      >>>> COLUMN
          1
ROW  1  17400.0
ROW  2  502600.
ROW  3  8000.00

```

1. Matrix ->  $\text{XSXINV} = \text{GINV}(\text{XSX})$

```

<<<< XSEXINV >>>> COLUMN
      1          2          3
ROW  1  16.7000  -0.514286  -4.22857
ROW  2  -0.514286  0.163265E-01  0.118367
ROW  3  -4.22857  0.118367  1.60816

```

1. Matrix -> BOLS=XSEXINV|XSY

```

<<<< BOLS >>>> COLUMN
      1
ROW  1 -1728.57
ROW  2  204.081
ROW  3 -1220.41

```

2. Matrix -> BET=XLSQ(X, INCOME)

```

<<<< BET >>>> COLUMN
      1
ROW  1 -1728.57
ROW  2  204.082
ROW  3 -1220.41

```

```

=====
Ordinary Least Squares
Dependent Variable      INCOME      Number of Observations      6
Mean of Dep. Variable  2900.0000  Std. Dev. of Dep. Var.  1634.625339
Std. Error of Regr.    602.4892   Sum of Squared Residuals  .108898E+07
R - squared            .91849     Adjusted R - squared      .86415
F( 2, 3)              16.9025   Prob. Value for F        .02327
=====
Variable  Coefficient  Std. Error  t-ratio  Prob|t|>x  Mean of X  Std.Dev.of X
-----
Constant  -1728.57     2462.      -.702    .53320
AGE       204.082     76.98     2.651   .07693    26.66667   5.12510
SEX      -1220.41     764.0     -1.597   .20848    .66667     .51640

```

### 3.2 Nichtlineare Regression

Liegt ein nichtlinearer Funktionstyp vor, dann transformiert man diesen in eine lineare Form. Danach kann wieder die MKQ/OLS-Methode zur Anwendung kommen.

- **Hyperbel:**  $y = b_0 + \frac{b_1}{x}$

Lineare Transformation:  $y = b_0 + b_1 z$  mit  $z = \frac{1}{x}$

- **Parabel:**  $y = a \cdot x^b$  (Bsp.: Preisabsatzfunktion mit  $x$  = Preis und  $b$  = Preiselastizität)

Lineare Transformation:  $\ln y = \ln a + b \cdot \ln x$  bzw.

$y^* = b_0 + b_1 \cdot z$  mit  $z = \ln x$  und  $y^* = \ln y$

da  $c^y = x$ ,  $y = \log_c x$  ist  $a = \exp(b_0 = \ln a)$

- **e-Funktion:**  $y = a \cdot e^{bx}$

Da allgemein  $c^y = x$ ,  $y = \log_c x$ , ist  $a = \exp(b_0 = \ln a)$

Lineare Transformation:  $\ln y = \ln a + b \cdot x \cdot \underbrace{\frac{\ln e}{1}}$  bzw.

$$y^* = b_0 + b \cdot x$$

---

**Beispiele:**

ET: Transformation der Werte z.B. mit

CREATE;  $Z = \frac{1}{x}$ ;  $z \ln = \log(x)$

---

## 4 Zur Analyse drei- und mehrdimensionaler Häufigkeitsverteilungen

Berücksichtigung von mehr als zwei Variablen

**Interdependenzanalyse: multivariate Verfahren**

- Varianzanalyse
- Faktor-, Clusteranalyse (Backhaus et al. (1989))

**Dependenzanalyse: ökonometrische Verfahren**

- Multiple Regression
- Regression mit beschränkt abhängigen Variablen (Limited Dependent Variables, LDV)  
Discrete choice  $y = (0, 1)$  z.B. 'to work or not to work'
- Simultane Systeme
- etc. (siehe z.B. Greene (1991), Goldberger (1991))

Meine Vorlesungen: Im Rahmen der empirischen Wirtschaftsforschung:

- Regressionsanalyse - Einführung in die Ökonometrie
- Mikroökonometrie - Diskrete Entscheidungsmodelle
- Panelökonometrie

***Keyconcepts****Mehrdimensionale Verteilung**Randverteilungen**Bedingte Verteilungen**Kontingenzanalyse und Kontingenzkoeffizient**Rangkorrelationskoeffizient nach Spearman**Bravais-Pearson-Korrelationskoeffizient**Lineare Regression**Matrizenrechnung**Nichtlineare Regression*

## V Verhältnis- und Indexzahlen



Analyse von Relationen und Entwicklungen mit Verhältnis- und Indexzahlen, Bewertung von Warenkörben durch Preis-, Mengen- und Umsatzindizes

Für Vergleichszwecke und Charakterisierung von bestimmten Sachverhalten benötigt man Maßzahlen, deren Konstruktion und Interpretation von der Problemstellung abhängt.

### 1 Verhältniszahlen

Verhältniszahlen werden gebildet als Quotient zweier statistischer Größen, zwischen denen eine sinnvolle Beziehung vermutet werden kann als:

- Gliederungszahl
- Beziehungszahl
- Meßziffer

#### 1.1 Gliederungszahlen

Unter einer Gliederungszahl versteht man das **Verhältnis einer Teilgröße zu einer Gesamtgröße**; sie wird auch als Quoten-, Anteil- oder Prozentzahl (z.B. relative Häufigkeiten) bezeichnet.

**Beispiele:** \_\_\_\_\_

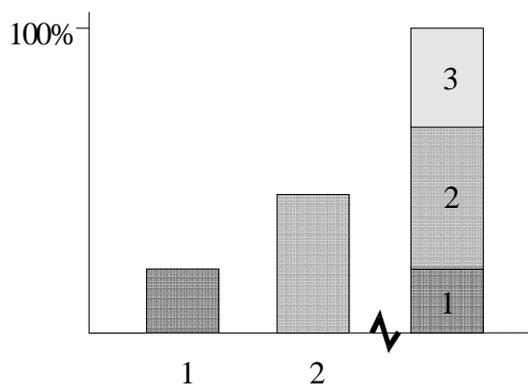
Aufgliederung

- eines Betriebsvermögens nach Herkunft finanzieller Mittel: z.B. 24 % Eigenkapital, 76 % Fremdkapital
- von Konsumausgaben (Budgetanteile):  

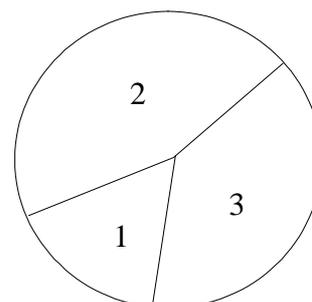
$$BA - \text{Nahrungs- und Genußmittel} = \frac{\text{Nahrungs- und Genußmittelausgaben}}{\text{Alle Ausgaben}} \cdot 100$$
- von Schülern nach Schularten
- von Erwerbstätigen nach Wirtschaftszweigen

Grafische Darstellung: als unterschiedliche Balken- oder Kreisdiagramme (Abb. V.1)

**Balkendiagramm**



**Kreisdiagramm**



1= 15%  
 2= 45%  
 3= 40%  
 -----  
 100%

Abb. V.1: Grafische Darstellung von Gliederungszahlen

## 1.2 Beziehungszahlen

Bei einer Beziehungszahl werden verschiedenartige Größen in sachlich sinnvoller Beziehung zueinander in Beziehung gesetzt als:

- Verursachungszahlen
- Entsprechungszahlen

### Verursachungszahlen

Bewegungsmasse wird auf Bestandsmasse bezogen.

#### Beispiele:

---

$$\text{Allgemeine Geburtenziffer} = \frac{\text{Zahl der Lebendgeborenen im Jahr } i}{\text{Mittlerer Bestand der Bevölkerung } (i)}$$

$$\text{'Konkursziffer'} = \frac{\text{Anzahl der Konkurse in Periode } t}{\text{Anzahl der Unternehmen in Periode } t}$$


---

### Entsprechungszahlen

Allgemein sinnvolle Größen werden in Beziehung gesetzt.

#### Beispiele:

---

$$\text{Kapitalintensität} = \frac{\text{Anlageinvestitionen}}{\text{Anzahl der Beschäftigten}}$$

$$\text{Arbeitsproduktivität} = \frac{\text{Reales Bruttoinlandsprodukt}}{\text{Anzahl der Erwerbstätigen}}$$

$$\text{Bevölkerungsdichte} = \frac{\text{Wohnbevölkerung}}{\text{Fläche in km}^2}$$

Arithmetisches Mittel, Variationskoeffizient

---

## 1.3 Meßziffern

Meßziffern sind Quotienten aus (Zeit-) Reihenwerten eines (metrisch skalierten) Merkmals:

$$\frac{x_j}{x_i} \quad (x_i \neq 0).$$

Damit gibt die Meßziffer  $x_j/x_i$  an, in welchem Verhältnis sich die Merkmalsausprägungen in einer **Berichtsperiode j** gegenüber einer **Basisperiode i** geändert hat. Meßziffern sind wichtig für Vergleichszwecke und Entwicklungen.



## b) Meßziffernreihe der Privathaushalte in der BRD

Tab. V.2: Meßziffernreihe der Privathaushalte in der BRD

Volkszählung	Privathaushalte (in 1000)					
	insgesamt	Meßziffer	Kettenmeßziffer	Einpersonen	Meßziffer	Kettenmeßziffer
13.9.1950	16650	100,0	1,00	3229	100,0	1,00
6.6.1961	19399	116,5	1,17	4126	127,8	1,28
27.5.1970	21991	132,1	1,13	5527	171,2	1,34
25.5.1987	26218	157,5	1,19	8767	271,5	1,59

Quelle: Der Bundesminister für Arbeit und Sozialordnung, Statistisches Taschenbuch (1989), Tab. 2.1

## 2 Indexzahlen

Bisher haben wir nur einzelne Meßziffernreihen betrachtet. Jetzt wollen wir mehrere, sachlich zusammengehörige Reihen betrachten. Der Verlauf all dieser Reihen soll durch eine globale Meßziffer angegeben werden: **Indexzahl**.

### Beispiel:

Sind die Lebenshaltungskosten eines Berichtsjahres gegenüber dem Basisjahr gestiegen?

Lebenshaltung: Vielzahl von Einzelprodukten, die mit Preisen bewertet werden

Interessant: Welchen Einfluß haben Preis- bzw. Mengenänderungen?

Berechnung von

- Preisindizes
- Mengenindizes
- Umsatz- (Wert-) indizes

### 2.1 Preisindizes: Laspeyres und Paasche

Preisindizes haben die Aufgabe, die Preisentwicklung der Gesamtheit von Gütern zwischen Berichts- und Basisjahr in einer einzigen Zahl anzugeben.

### Beispiel:

Tab. V.3: Preis- und Mengenentwicklung eines Gutes

Gut	Preise		Mengen	
	Basiszeit 0	Berichtszeit t	Basiszeit 0	Berichtszeit t
1	4	6	5	4
2	6	8	10	15
3	10	12	18	16

Bezeichnungen:

$p_0^i$  = Preis des Gutes i zur Basiszeit

$p_t^i$  = Preis des Gutes i zur Berichtszeit t

$q_0^i$  = Menge des Gutes i zur Basiszeit

$q_t^i$  = Menge des Gutes i zur Berichtszeit

### Preismeßziffer

$$I_{0,t}(p^i) = \frac{p_t^i}{p_0^i} \cdot 100$$

### Beispiel:

---

Preismeßziffern eines Gutes

$$I_{0,t}(p^1) = \frac{6}{4} \cdot 100 = 150$$

$$I_{0,t}(p^2) = \frac{8}{6} \cdot 100 = 133$$

$$I_{0,t}(p^3) = \frac{12}{10} \cdot 100 = 120$$


---

### Konstruktion einer einheitlichen Indexzahl für die Preisentwicklung

Bildung des ungewogenen arithmetischen Mittels aus Preismeßziffern

$$I_{0,t}(p) = \frac{1}{n} \sum_{i=1}^n I_{0,t}(p^i)$$

### Beispiel:

---

Preismeßziffern eines Gutes

$$I_{0,t}(p) = \frac{I_{0,t}(p^1) + I_{0,t}(p^2) + I_{0,t}(p^3)}{3} = 134,33$$


---

Die Indexzahl bleibt aber unbefriedigend, da unterschiedliche Mengen nicht berücksichtigt werden. Also müssen die Güter im Index gewichtet werden:

### Preisindex nach Laspeyres

$$I_{0,t}^L(p) = \frac{\sum_{i=1}^n p_t^i \cdot q_0^i}{\sum_{i=1}^n p_0^i \cdot q_0^i} \cdot 100$$

Der Warenkorb  $(q_0^1, q_0^2, q_0^3, \dots)$  wird dabei zur **Basisperiode** konstant gehalten ('NULL = Laspeyres und Basis Null').

### Preisindex nach Paasche

$$I_{0,t}^P(p) = \frac{\sum_{i=1}^n p_t^i \cdot q_t^i}{\sum_{i=1}^n p_0^i \cdot q_t^i} \cdot 100$$

Der Warenkorb der **Berichtsperiode** bleibt konstant.

### Beispiel:

$$I_{0,t}^L = \frac{6 \cdot 5 + 8 \cdot 10 + 12 \cdot 18}{4 \cdot 5 + 6 \cdot 10 + 10 \cdot 18} \cdot 100 = \frac{326}{260} \cdot 100 = 125,4 \quad \text{Mengen der Basiszeit 0}$$

$$I_{0,t}^P = \frac{6 \cdot 4 + 8 \cdot 15 + 12 \cdot 16}{4 \cdot 4 + 6 \cdot 15 + 10 \cdot 16} \cdot 100 = \frac{336}{266} \cdot 100 = 126,3 \quad \text{Mengen der Berichtszeit t}$$

### Anwendungsbeispiele:

Preisindizes für die Lebenshaltung:

Die Entwicklung der für die privaten Haushalte bedeutsamen Preise soll wiedergeben werden. Hierfür eignet sich der Index von Laspeyres. In 'Wirtschaft und Statistik' veröffentlicht das Statistische Bundesamt mehrere Preisindizes für die Lebenshaltung nach Laspeyres, die sich bezüglich des verwendeten Warenkorbes unterscheiden:

- a) - Preisindex für die Lebenshaltung aller privaten Haushalte (Warenkorb aus 800 Waren- und Dienstleistungen)

Tab. V.4: Preisindex für die Lebenshaltung aller privaten Haushalte (2005=100)

	2006	2007	2008	2009	2010
Preisindex für die Lebenshaltung aller priv. Haushalte (2005=100)	101,6	103,9	106,6	107,0	108,2
Veränderung (%)		+2,3	+2,6	+0,4	+1,1

Quelle: Statistisches Bundesamt, Wiesbaden (2011) GESIS-Online Datenbank

- b) - von 4-Personen-Arbeitnehmerhaushalten mit mittlerem Einkommen des alleinverdienenden Haushaltsvorstands  
 c) - von 4-Personen-Haushalten von Angestellten und Beamten mit höherem Einkommen  
 d) - von 2-Personen-Haushalten von Renten- und Sozialhilfeempfängern  
 e) - Preisindex für die einfache Lebenshaltung eines Kindes

Jedem dieser Indizes liegt ein spezieller Warenkorb zugrunde.

**Aussagekraft und Probleme:**

- Qualitätsänderungen der Waren und Dienstleistungen sind kaum zu erfassen;
- Warenkorb oft schnell überholt;
- Übereinstimmung des individuellen Warenkorbs mit dem Warenkorb des Indexhaushaltes ist entscheidend;
- Ermittlung von Preisreihen in örtlicher Differenzierung (verschiedene Gemeindegrößenklassen); Preisunterschiede.

**Weitere Indizes:**

- Index der Einzelhandelspreise: Verkaufspreise (incl. MwSt.), Laspeyres (beinhaltet auch Waren, die nicht von privaten Haushalten gekauft werden (z.B. Maschinen, Büroeinrichtungen));
- Index der Großhandelsverkaufspreise: VK-Preise (ohne MwSt.), Inlandsabsatz, Laspeyres;
- Indizes zum Kaufkraftvergleich: Interregionaler Preisvergleich zwischen Ländern mit verschiedenen Währungen; (Purchasing Power Priceindices (PPP) wie Geary-Khamis, Eltetö-Köves-Szuk oder Iklé Preisindizes)

**Beispiel:**

Kaufkraftvergleich zwischen den USA und D  
Ein deutscher Standardwarenkorb koste

$$\text{in den USA} \quad \sum p_{US} \cdot q_D = 720 \$$$

$$\text{in D} \quad \sum p_D \cdot q_D = 1800 \text{ EUR}$$

Als Verbrauchergeldparität, also die Kaufkraftparität, ergibt sich:

$$VG_{D,US} = \frac{720 \$}{1800 \text{ EUR}} = 0,40 \$ / \text{EUR}$$

$$d.h.: \quad 1 \$ \cong 2,50 \text{ EUR} \quad \text{bzw.} \quad 1 \text{ EUR} \cong 0,40 \$$$

Ein amerikanischer Standardwarenkorb koste

$$\text{in D} \quad \sum p_D \cdot q_{US} = 2065 \text{ EUR}$$

$$\text{in den USA} \quad \sum p_{US} \cdot q_{US} = 620 \$$$

Als Verbrauchergeldparität erhält man jetzt:

$$VG_{US,D} = \frac{2065 \text{ EUR}}{620 \$} = 3,33 \text{ EUR} / \$$$

$$d.h.: \quad 1 \$ \cong 3,33 \text{ EUR} \quad \text{bzw.} \quad 1 \text{ EUR} \cong 0,30 \$$$

Will man aus beiden Größen den Mittelwert bilden, dann muß man das geometrische Mittel errechnen:

$$1\$ \hat{=} 2,89 \text{ EUR} \quad \text{bzw.} \quad 1\text{EUR} \hat{=} 0,35\$$$

- Deflationierung:  
realer Verbrauch der Periode t bewertet mit konstanten Preisen der Basisperiode (Division durch den Preisindex von Paasche)
- Nominallohn- und Reallohnindizes:  
durchschnittliche Veränderung des Lohnniveaus; das Statistische Bundesamt berechnet:

- Index der tariflichen Stundenlöhne

$$I_{0,t} = \frac{\sum_i \frac{L_t^i}{L_0^i} \cdot L_0^i \cdot Z_0^i}{\sum_i L_0^i \cdot Z_0^i} \cdot 100$$

mit

$L_0^i$  = Tariflohn je Beschäftigten der Beschäftigtengruppe i zur Basiszeit

$Z_0^i$  = Anzahl der Beschäftigten der Beschäftigungsgruppe i zur Basiszeit

$L_t^i$  = Tariflohn je Beschäftigten der Beschäftigtengruppe i zur Berichtszeit

- Index der tariflichen Wochenlöhne  
 $I_{0,t}$  multipliziert mit Index der tariflichen Wochenarbeitszeit der Arbeiter
- Reallohnindex

$$\text{Reallohnindex} = \frac{\text{Nominallohnindex}}{\text{Preisindex für die Lebenshaltung}} \cdot 100$$

## 2.2 Mengenindizes

Bei Preisindizes werden die Mengen jeweils konstant gehalten, bei **Mengenindizes** werden die **Preise konstant** gehalten, um Aussagen über die durchschnittliche Mengenänderung zu erhalten.

Mengenindex nach Laspeyres:

$$I_{0,t}^L(q) = \frac{\sum_{i=1}^n p_0^i \cdot q_t^i}{\sum_{i=1}^n p_0^i \cdot q_0^i} \cdot 100$$

Die Preise der **Basisperiode** werden konstant gehalten.

Mengenindex nach Paasche:

$$I_{0,t}^P(q) = \frac{\sum_{i=1}^n p_t^i \cdot q_t^i}{\sum_{i=1}^n p_t^i \cdot q_0^i} \cdot 100$$

Die Preise der **Berichtsperiode** werden konstant gehalten.

**Beispiel:** \_\_\_\_\_

Index der industriellen Nettoproduktion

Unter Ausschaltung der Preisveränderungen wird die **Entwicklung der Nettoproduktionsleistung** der einzelnen Industriezweige und der Gesamtindustrie gemessen werden.

Mengenindex nach Laspeyres:

$$I_{0,t}^L(q) = \frac{\sum_{i=1}^n \frac{q_t^i}{q_0^i} \cdot q_0^i \cdot p_0^i}{\sum_{i=1}^n q_0^i \cdot p_0^i} \cdot 100$$

Dabei sind die Nettoproduktionswerte  $(q_0^i \cdot p_0^i)$  des Industriezweiges die Ausgangsmaße; man schreibt dann mit Hilfe der Mengenmaßziffern  $q_t^i/q_0^i$  fort, siehe auch Kapitel II.3.3.

### 2.3 Umsatzindizes (Wertindizes)

Wenn sowohl mengen- als auch preismäßige Änderungen voll erfaßt werden sollen, verwendet man Umsatz- oder Wertindizes:

$$I_{0,t}(p, q) = \frac{\sum_{i=1}^n p_t^i \cdot q_t^i}{\sum_{i=1}^n p_0^i \cdot q_0^i} \cdot 100$$

## 3 Umbasierung, Verknüpfung und Preisbereinigung von Indizes

### Umbasierung

Unter der Umbasierung versteht man das Umstellen eines Indexes von einer alten auf eine neue Basis:

$$I_{t_1,t} = \frac{I_{0,t}}{I_{0,t_1}} \cdot 100$$

mit t: Zeitindex  
 0: alte Basisperiode  
 t<sub>1</sub>: neue Basisperiode

### Beispiel:

Umbasierung des Index der Erzeugerpreise industrieller Produkte in der Bundesrepublik Deutschland von 1970 auf 1974:

Tab. V.5: Umbasierung des Index der Erzeugerpreise industrieller Produkte in der Bundesrepublik Deutschland von 1970 auf 1974:

Jahr t	Index I <sub>70,t</sub> (%)	Index I <sub>74,t</sub> (%)
1970	<b>100</b>	81
1971	106	85
1972	109	<b>88</b>
1973	114	92
1974	124	<b>100</b>
1975	133	107
1976	138	111
1977	142	115

Quelle: Hujer (1990), S. 116 f.

Umbasierung über:  $I_{74,t} = \frac{I_{70,t}}{I_{70,74}} \cdot 100$        $\left( \text{aus Dreisatz: } \frac{x}{100} = \frac{I_{70,t}}{I_{70,74}} \right)$

z.B. t = 1972:

$$I_{74,72} = \frac{I_{70,72}}{I_{70,74}} \cdot 100 = \frac{109}{124} \cdot 100 = 88 \% \quad \left( \frac{x}{100} = \frac{109}{124} \right)$$

### Verknüpfung

Um eine Preisentwicklung über eine größere Zeitspanne zu verfolgen, muß ein alter Index mit dem neuen Index verknüpft werden.

### Beispiel:

In der BRD wurde im Jahr 1970 der Preisindex für die Lebenshaltung auf eine neue Basis umgestellt.

Tab. V.6: Verknüpfung der Preisindizes für die Lebenshaltung

Jahr t	Alt	Neu	(a)	(b)
	$I_{62,t}$	$I_{70,t}$		
1962	100	.	100	81
1963	103	.		83
1964	104	.		84
1965	109	.		88
1966	113	.		91
1967	115	.		93
1968	116	.		94
1969	120	.		97
1970	124	100	124	100
1971	.	105	130	
1972	.	111	138	
1973	.	119	148	
1974	.	127	157	
1975	.	135	167	
1976	.	141	175	
1977	.	146	181	146

Quelle: Hujer (1990), S. 117

a) Fortführung des alten Index:

Annahme: Der neue Index verhalte sich in der Vergangenheit **proportional** zum alten Index. Über einen Dreisatz erhält man:

z.B.

$$\frac{I_{62,77}^*}{I_{62,70}} = \frac{I_{70,77}}{100} \quad \Rightarrow \quad I_{62,77}^* = \frac{I_{62,70} \cdot I_{70,77}}{100} = \frac{124 \cdot 146}{100} = 181 \quad \left( \text{über: } \frac{x}{124} = \frac{146}{100} \right)$$

b) Rückrechnung des neuen Index:

$$\frac{I_{70,63}^*}{100} = \frac{I_{62,63}}{I_{62,70}} \quad \Rightarrow \quad I_{70,63}^* = \frac{I_{62,63} \cdot 100}{I_{62,70}} = \frac{103 \cdot 100}{124} = 83 \quad \left( \text{über: } \frac{x}{100} = \frac{103}{124} \right)$$

### Allgemeine Formeln für die Verknüpfung

a) Fortführung des alten Index:

$$I_{0,t}^* = \frac{I_{0,t_1} \cdot I_{t_1,t}}{100}$$

mit  $I_{0,t}^*$ : verknüpfter Index zur alten Basis

0: alte Basisperiode

$t_1$ : Verknüpfungsperiode

b) Rückrechnung des neuen Index:

$$I_{t_1,t}^* = \frac{I_{0,t} \cdot 100}{I_{0,t_1}}$$

mit  $I_{t_1,t}^*$ : verknüpfter Index zur neuen Basis

0: alte Basisperiode

$t_1$ : Verknüpfungsperiode

### Preisbereinigung

Für die Analyse **realer** Entwicklungen ist es nötig, Preisveränderungen möglichst auszuschalten. Dies geschieht über die **Preisbereinigung** (Deflationierung).

Grundidee:

Die Preismeßziffer eines Beobachtungszeitraumes  $t$ , nämlich  $\sum_i p_t^i \cdot q_t^i$ , soll für die Mengen  $(q_t^1, \dots, q_t^n)$  in eine fiktive Größe  $\sum_i p_0^i \cdot q_t^i$  umgerechnet werden. Die Mengen in  $t$  sollen also mit den Preisen der Basisperiode 0 bewertet werden.

Diese preisbereinigte Größe  $\sum_i p_0^i \cdot q_t^i$  soll dann mit der tatsächlichen Größe  $\sum_i p_0^i \cdot q_0^i$  der Basisperiode 0 verglichen werden.

Die **reale Größe** erhält man, in dem die nominale Größe  $\sum_i p_t^i \cdot q_t^i$  durch den Paasche-Preisindex dividiert wird:

$$\frac{\sum_i p_t^i \cdot q_t^i}{I_{0,t}^p(p)} = \frac{\sum_i p_t^i \cdot q_t^i}{\frac{\sum_i p_t^i \cdot q_t^i}{\sum_i p_0^i \cdot q_t^i}} = \sum_i p_0^i \cdot q_t^i \quad (\text{preisbereinigte, reale Größe})$$

Die preisbereinigte Größe  $\sum_i p_0^i \cdot q_t^i$  gibt an, wie groß die Ausgaben in der Beobachtungsperiode  $t$  tatsächlich gewesen wären, wenn sich zwischen Beobachtungs- und Basisperiode nur die Mengen, nicht aber die Preise geändert hätten.

Nun kann auch ein Vergleich zwischen  $\sum_i p_t^i \cdot q_t^i$  mit  $\sum_i p_0^i \cdot q_t^i$  stattfinden.

Sind keine Paasche-Preisindizes in der Statistik verfügbar, dann werden approximativ Laspeyres-Preisindizes verwendet.

Daneben gibt es **reale Indizes**, wie z.B.

$$\text{Realer Lohn - und Gehaltsindex} = \frac{\text{Nominaler Lohn - + Gehaltsindex}}{\text{Preisindex für die Lebenshaltung}} \cdot 100$$

## Aktuelle Preisinformationen

Eine Darstellung unterschiedlicher Indizes ist auf der Homepage des Statistischen Bundesamtes abrufbar (<http://www.destatis.de>)



Abb. V.2: Verbraucherpreisindex

Quelle: Statistisches Bundesamt, Online im Internet (<http://www.destatis.de>)

### Keyconcepts

*Gliederungs- und Beziehungszahlen*

*Meßziffern und –reihe*

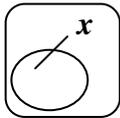
*Preisindizes: Laspeyres und Paasche*

*Mengenindizes: Laspeyres und Paasche*

*Umsatzindizes*

*Umbasierung*

## VI Voll- und Teilerhebungen



*Stichprobenverfahren, ADM-Stichprobe, Repräsentativität durch Hochrechnung von Stichproben*

### 1 Allgemeine Bemerkungen und Beispiele aus der amtlichen und nichtamtlichen Statistik

#### Vollerhebung

Alle Einheiten einer Grundgesamtheit, z.B.

- Volkszählung BRD 1970, 1987
- Alle Mitarbeiter eines Betriebes, einer Gewerkschaft etc.

#### Teilerhebung

Teilmenge einer Grundgesamtheit wegen Kosten- und Zeitersparnis, detailliertere Fragen, z.B.

- Mikrozensus BRD
- Sozio-ökonomisches Panel des Sfb3/DIW:  
6000 Haushalte mit ca. 12000 Personen, jährlich wiederholt befragt seit 1984
- Sfb3-Nebenerwerbstätigkeitsumfrage (Sfb3):  
8000 Personen, 4 saisonale 'Einschaltungen'
- ALLBUS, Konsumdaten der GfK etc. (Infos: Zentralarchiv Köln)

### 2 Auswahlverfahren

Die Ergebnisse einer Teilerhebung sollen für Rückschlüsse auf die Grundgesamtheit verwendet werden. Daher muß die Teilerhebung (Stichprobe) repräsentativ sein!

Zu klären ist:

1. Wie sollen die Stichprobeneinheiten ausgewählt werden?
2. Wieviele Einheiten (Stichprobenumfang)?
3. Wie genau stellen die Stichprobenergebnisse den 'wahren' Sachverhalt dar?

→ Stichprobentheorie

Ausgehend vom **Untersuchungsziel** (z.B. Einkommen der Haushalte, Personen) wird eines der folgenden **Auswahlverfahren** gewählt:

- Zufallsauswahl
- bewußte Auswahl

#### Zufallsauswahl:

Jede Einheit der Grundgesamtheit hat die gleiche Chance, in die Stichprobe zu gelangen. Der 'Fehler', der durch die Beschränkung der Zufallsauswahl entsteht, ist berechenbar, (Wahrscheinlichkeitsrechnung).

**Bewusste Auswahl:**

Auswahl nach gewissen Vorkenntnissen aus der Grundgesamtheit:

z.B. Quotenverfahren: VZ  $\Rightarrow$  Geschlechteranteil  $\Rightarrow$  Quote für neue Stichprobe

Im Weiteren wird nur die Zufallsauswahl betrachtet.

### 3 Stichprobenerhebungen

#### 3.1 Einfaches Stichprobenverfahren

Eine einfache Stichprobe ist eine reine Zufallsauswahl: jede Einheit der Grundgesamtheit hat die gleiche Chance.

Zufallszahlen: Spezielle Tabellen, Random Number Generators

**Beispiel:**

Adressenverzeichnis einer Bevölkerung (1,...,N):

- Zufallszahlen  $\rightarrow$  Adressennummern (1,...,n), n = Stichprobenumfang
- oder z.B. jeder 6. im Adressenverzeichnis kommt in die Stichprobe. Achtung: wegen möglicher Sortierungen besteht die Gefahr der Korrelation der Anordnung mit dem Untersuchungsmerkmal.

#### 3.2 Geschichtete Stichprobenverfahren

Einfache Stichprobenverfahren sind eher für homogene Grundgesamtheiten (GG) geeignet.

Jetzt betrachten wir inhomogene GG, die homogene Teilgesamtheiten (= Schichten) besitzen.

Aus einer geschichteten Stichprobe lassen sich dann bessere Ergebnisse erzielen.

Es besteht eine berechenbare Chance nach Festlegung des Auswahlatzes in die Stichprobe zu gelangen:

- Schichtaufteilung 
$$N = \sum_{i=1}^I N_i$$
- Elemente pro Schicht 
$$n = \sum_{i=1}^I n_i \quad n_i \leq N_i$$

Es gibt nun verschiedene Möglichkeiten der Aufteilung:

- proportionale Aufteilung:  $p_i = N_i/N \quad \rightarrow \quad n_i = n \cdot p_i = n \cdot N_i/N$
- optimale Aufteilung durch Minimierung der Varianz der Stichprobenergebnisse

### 3.3 Klumpenstichprobe

Die Grundgesamtheit wird aufgeteilt in  $m$  - nicht notwendig umfanggleiche - Klumpen. Daraus erfolgt eine zufällige Auswahl von Klumpen, von denen dann alle Einheiten dieser Klumpen in die Stichprobe übernommen werden.

#### Vorteile:

- wesentliche Vereinfachung der Erhebung
- Zeit- und Kostenersparnis

Die Klumpen sollten bezüglich ihres Untersuchungsmerkmals möglichst inhomogen sein.

Eine Klumpenauswahl erfolgt dann, wenn nur für die Klumpen und nicht für die Untersuchungseinheiten eine Auswahlgrundlage vorliegt.

### 3.4 Mehrstufige Stichprobenverfahren

Es werden mehrere Zufallsauswahlen hintereinander durchgeführt. Der Umfang der Teilgesamtheiten verringert sich dann von Auswahl zu Auswahl.

Anwendung: z.B. in der Qualitätskontrolle:

1. Stufe: Auswahl aus einer Menge von Kisten
2. Stufe: Artikel aus ausgewählten Kisten

## 4 Das ADM - Stichprobenverfahren: Ein Beispiel aus der privaten Umfrageforschung

Das ADM-Mastersample ist Grundlage vieler Umfragen privater Forschungsinstitute (z.B. Infratest, Marplan, Emnid...).

#### Grundgesamtheit

Personen:

Die deutsche Bevölkerung in Privathaushalten am Ort der Hauptwohnung in der Bundesrepublik im Alter von 14 und mehr Jahren

Haushalte:

Die Privathaushalte am Ort der Hauptwohnung mit deutscher Bezugsperson

#### Stichprobenanlage

Stichprobenband: Länder, Regierungsbezirke, Kreise

Gemeindegrößenklassen nach Boustedt: alphabetisches Listing der Gemeinden, Nummernsystem der Wahlbezirke, Zusammenfassung → 50.000 originäre, synthetische Wahlbezirke

#### 1. Schichtung

→ 1001 Zellen (besetzt von 3280) nach 328 Kreisen, 10 Boustedt

Bedeutungsgewicht der 50.000 Einheiten (entspricht der Anzahl der Haushalte pro Einheit)

Sampling points mit Optimierungsverfahren

→ Verteilung der sampling points auf die Schichtungszellen

## 2. Drei Ziehungsstufen

Erste Stufe:

systematische Zufallsauswahl

→ Sample points je Schichtungszelle

Zweite Stufe:

Gleiche Anzahl von Adressen für jeden sample point ('Türschild')

Dritte Stufe:

Vor Ort: Pro Haushalt Auflistung der zu befragenden Personen (vom Interviewer). Auswahl der Person nach Permutation ('Schwedenschlüssel')

## 5 Hochrechnung von Teilerhebungen

Die Gewichtung der Auswahleinheiten erfolgt so, dass die gewichtete Merkmalssumme gleich der gewünschten 'Randverteilung' ist (GG-Wert).

Eine Hochrechnung ist notwendig, da die Stichprobe durch die konkrete Ziehung ('sampling error' etc.) Verzerrungen aufweist.

Hochrechnung als 'static aging' (Umgewichtung, 'grossing-up') ist zentral in statischen Mikrosimulationsmodellen

### Möglichkeiten:

- 'economic aging'-Inflationierung (ein Faktor für alle Einheiten der Stichprobe)
- demografische Hochrechnung als simultane Hochrechnung: Einheitenabhängiger Faktor soll simultan bestimmte Restriktionen erfüllen. Dafür ist die **Lösung eines Optimierungsproblems unter Nebenbedingungen** nötig:

$$Z = Z(p, q) = \min! \quad s.t. \quad Sp = r$$

$q_n$  = alte Hochrechnungsfaktoren (HRF), auch  $q_j = 1$  für  $j = 1, \dots, n$  Mikroeinheiten

$p_n$  = neue HRF

$S_{(m,n)}$  = Sample Matrix

$r_m$  = Restriktionen

$$S_{mn} = \begin{bmatrix} S_{11} & \dots & S_{1j} & \dots & S_{1n} \\ & & \dots & & \\ S_{i1} & & S_{ij} & & S_{in} \\ \hline S_{i+1,1} & & S_{i+1,j} & & S_{i+1,n} \\ & & \dots & & \\ S_{k1} & & S_{kj} & & S_{kn} \\ \hline S_{k+1,1} & & S_{k+1,j} & & S_{k+1,n} \\ & & \dots & & \\ S_{m1} & & S_{mj} & & S_{mn} \end{bmatrix}$$

z.B. Minimum Information Loss (MIL)-Prinzip (Merz (1994b))

$$Z = \sum_{j=1}^n p_j \cdot \log \left( \frac{p_j}{q_j} \right) = \min! \quad \text{s.t.} \quad S \cdot p = r$$

Besitzt die wünschenswerte Positivitäts-Eigenschaft:  $p_j > 0$ ; d. h., jede Mikroeinheit  $j$  bleibt der weiteren Analyse erhalten. Schnelle Prozedur mit 'globaler exponentieller Approximation' → ADJUST-Programmpaket (Merz (1993a))

ADJUST - Demo: <http://www2.leuphana.de/ffb/adjust>

### **Keyconcepts**

*Zufallsauswahl*

*Bewusste Auswahl*

*Einfaches Stichprobenverfahren*

*Geschichtetes Stichprobenverfahren*

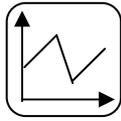
*Klumpenstichprobe*

*Hochrechnung*

*MIL-Prinzip*



## VII Zeitreihen- und Querschnittsanalyse



Zeitreihenzerlegung in unterschiedliche Komponenten: Trend, Zyklus und Saison;  
Spektralanalyse (Zeitreihenanalyse im Frequenzbereich, Querschnitts- und  
Panelanalyse)

### 1 Charakteristika und Komponenten einer Zeitreihe

#### Zeitreihe

Hier ökonomische Größen, die über einen längeren Zeitraum (Umsatz, Volkseinkommen von z.B. 1990 bis 2010) betrachtet werden.

#### Analyse

Untersuchung auf 'Gesetzmäßigkeiten'

- Aussagen über die Entwicklung in der Zeit (Erklärung)
- Aussagen über zukünftige Zeitpunkte (Extrapolation, Prognose)

'Gesetzmäßigkeiten' ermitteln durch Regressions- und Korrelationsrechnung:

$$y_t = f(x_t) \text{ oder } y_t = f(x_{1t}, x_{2t}, x_{3t}, \dots) \quad (t = 1, \dots, T)$$

→ Multivariate Zeitreihenanalyse

Konzentration auf das Zeitverhalten mit  $y_t = f(t)$   $t = \text{Zeitwerte z.B. 1987, 1988, ...}$  führt zur univariablen Zeitreihenanalyse.

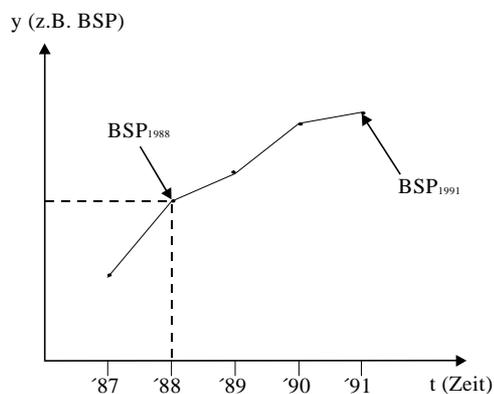


Abb. VII.1: BSP im Zeitverlauf

Eine Zeitreihenanalyse zerlegt die Zeitreihe in verschiedene Komponenten.

- **Trendkomponente  $T_t$ :** langfristige Entwicklung
- **Zyklische Komponente  $Z_t$ :** mittelfristige, sich periodisch wiederholende (insbesondere konjunkturelle Schwankungen) überlagern den Trend

- **Saisonale Komponente  $S_t$ :** jahreszeitliche Schwankungen (Jahreszeiten, Ferientermine, Steuertermine, etc.)
- **Zufallskomponente  $U_t$ :** Restschwankungen (mit der Annahme Mittelwert = Null)

### Mögliche Verknüpfungen der Komponenten:

- Additiv:  $y_t = T_t + Z_t + S_t + U_t \quad t = 1, \dots, T (= n)$
- Multiplikativ:  $y_t = T_t \cdot Z_t \cdot S_t \cdot U_t$
- Logarithmisch:  $\log y_t = \log T_t + \log Z_t + \log S_t + \log U_t$

### Beispiel:

Zeitreihenzerlegung

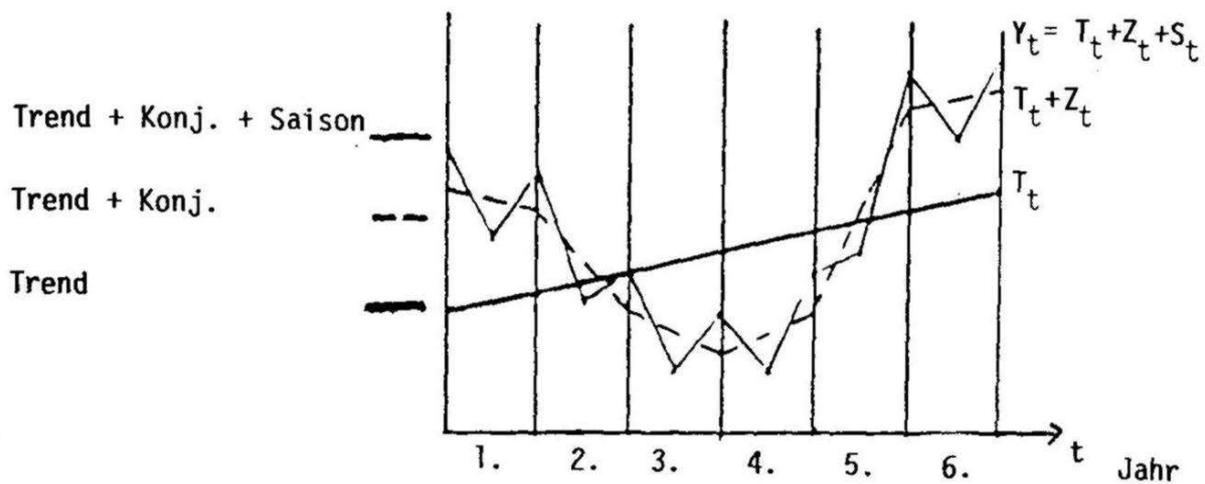


Abb. VII.2: Zeitreihenzerlegung

## 2 Bestimmung der Trend- und Zykluskomponente

Für die Ermittlung der **Trendkomponente** wird eine möglichst lange Zeitreihe (mit mehreren Konjunkturzyklen) benötigt.

Für die Ermittlung der **zyklischen Komponente** ist ein vollständiger Konjunkturzyklus notwendig.

Methoden für die Bestimmung der Trend- und Zykluskomponente: Methode der kleinsten Quadrate, gleitende Durchschnitte, exponentielle Glättung

### 2.1 Methode der kleinsten Quadrate

Eine Methode zur Trendermittlung ist die Methode der kleinsten Quadrate (MKQ = OLS, Ordinary Least Squares).

Nach der Wahl des Funktionstyps muß hier u.U. eine **Transformation** von einer nichtlinearen in eine lineare Funktion vorgenommen werden.

- a)  $\hat{y} = T_t = b_0 + b_1 \cdot t + b_2 \cdot t^2$  Polynom  
 b)  $\hat{y} = T_t = b_0 (b_1)^t$  Exponentialansatz  
 c)  $\hat{y}^2 = T_t^2 = b_0 + b_1 \cdot t$  Wurzelfunktion

und weitere Funktionstypen; Transformation nichtlinearer zu linearen Ansätzen, dann OLS.

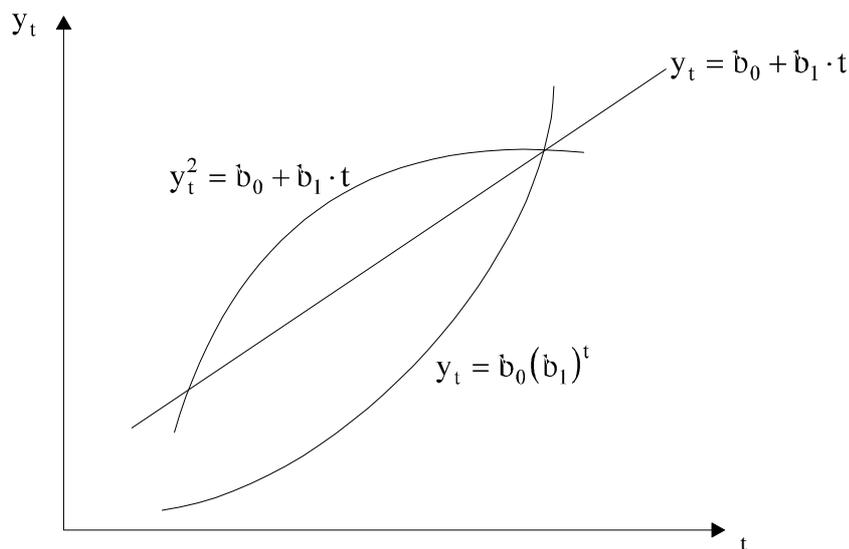


Abb. VII.3: Nichtlineare und lineare Funktionstypen

Nach entsprechender Transformation lautet der lineare Ansatz allgemein:  $\hat{y}_t = b_0 + b_1 \cdot t$

Da  $t$  dem  $x$  aus der OLS-Ableitung entspricht (vgl. Abschnitt IV.3) können die Ergebnisse (mit  $x \rightarrow t$ ) direkt entnommen werden.

$$b_0 = \bar{y} - b_1 \cdot \bar{t}$$

$$b_1 = \frac{\frac{1}{n} \sum_i y_i \cdot t_i - \bar{y} \cdot \bar{t}}{\frac{1}{n} \sum_i t_i^2 - \bar{t}^2} = \frac{s_{ty}}{s_t^2}$$

Die Zeitkomponente  $t_i$  wird in eine Zeitkomponente  $t'_i$  transformiert, für die gilt:  $\sum_{i=1}^n t'_i = 0$ .

Es erfolgt also eine Bereinigung um den Mittelwert  $\bar{t}$ .

Bei der Bereinigung muß man beachten:

n ungerade:  $t'_i = t_i - \bar{t}$

n gerade:  $t'_i = 2(t_i - \bar{t})$

$$\hat{y}_i = b'_0 + b'_1 \cdot t'_i$$

Wenn  $\sum_{i=1}^n t'_i = 0$ , gilt folgende Vereinfachung (analog auch für  $y_i = f(x_i)$ ):

$$b'_0 = \bar{y}$$

$$b'_1 = \frac{\sum_{i=1}^n y_i \cdot t'_i}{\sum_{i=1}^n t_i^2}$$

Rücktransformation:

$$\begin{aligned} \text{n ungerade: } \quad b_0 &= b'_0 - b'_1 \cdot \bar{t} \\ & \quad b_1 = b'_1 \\ \text{n gerade: } \quad b_0 &= b'_0 - 2 \cdot b'_1 \cdot \bar{t} \\ & \quad b_1 = 2 \cdot b'_1 \end{aligned}$$

### Beispiel:

Volkseinkommen und privater Verbrauch für die BRD 2002-2010 in Mrd. EUR

Tab. VII.1: Volkseinkommen und privater Verbrauch für die BRD 2002-2010 in Mrd. EUR

Jahr	$t'_i$	Volkseinkommen $y_i$	Privater Verbrauch $x_i$
2002	-4	1576,0	1227,78
2003	-3	1599,0	1247,23
2004	-2	1672,0	1265,29
2005	-1	1694,0	1288,76
2006	0	1779,0	1321,22
2007	1	1837,0	1341,99
2008	2	1871,0	1375,65
2009	3	1791,0	1371,63
2010	4	1903,0	1403,87

Quelle: Statistisches Bundesamt, Wiesbaden (2011) GENESIS-Online Datenbank

OLS-Resultate für das Volkseinkommen als  $f(t)$

$$- \hat{y}_i = -79328,94 + 40,41 \cdot t_i \quad B = 0,88 \quad (t = 1, 2, \dots) \quad a^y = x \quad y = \log_a x$$

$$- \hat{y}_i = 1746,88 + 40,41 \cdot t'_i \quad B = 0,88$$

$$- y_i = b_0 (b_1)^{t'_i} \quad \text{Exponentialfunktion}$$

$$\ln y_i = \ln b_0 + \ln b_1 \cdot t'_i$$

$$y_i^* = b_0^* + b_1^* \cdot t'_i$$

$$\hat{y}_i^* = \underbrace{7,4634}_{b_0^*=1743,06} + \underbrace{0,0233}_{b_1^*=1,0235} \cdot t'_i \quad B = 0,88$$

$$b_0^* = \ln b_0 = 7,4634 \quad \Rightarrow \quad b_0 = \exp(7,4634)$$

$$b_0 = 1743,06$$

also:

$$\hat{y}_i = 1743,06 \cdot (1,0235)^i$$

$$\text{Für 2008 ergibt sich z. B.: } y_{2008} = 1825,94 \quad \hat{y}_{2008} = 1827,7$$

Zum Vergleich das Ergebnis der Regression  $y_i = b_0 + b_1 x_i$ :

$$\hat{y}_i = -662,02 + 1,8305 \cdot \text{Privater Verbrauch}_i \quad B = 0,938$$

Eine Trendextrapolation liefert dann die Prognosewerte. Man setzt dazu die 'erklärende' Variable ein.

## 2.2 Methode der gleitenden Durchschnitte

Zur Schätzung der **glatten Komponente** ( $g_t$ ) muß man den Trend und den Zyklus schätzen:

$$G_t = T_t + Z_t$$

Daraus wird die zyklische Komponente berechnet:  $Z_t = G_t - T_t$ .

### Methode der gleitenden Durchschnitte:

Für eine Zeitreihe  $y_t$  bilde man aus  $y_t$  sowie aus den  $k$  vorangehenden und  $k$  nachfolgenden Zeitreihenwerten jeweils das arithmetische Mittel ( $\rightarrow$  eine um  $2k$  Glieder verkürzte Zeitreihe).

Anders formuliert: Man bilde aus jeweils  $2k+1$  (ungerade Anzahl) zeitlich aufeinanderfolgenden Beobachtungswerten das arithmetische Mittel. Damit ergibt sich ein  $(2k+1)$ -gliedriger gleitender Durchschnitt:

$$y'_t = \frac{1}{2k+1} \sum_{\tau=t-k}^{t+k} y_\tau \quad \text{mit} \quad t = k+1, \dots, n-k$$

### Beispiel:

3-gliedrige Durchschnitte beim Volkseinkommen:  $(2k+1)=3 \Rightarrow k=1$

$$y'_{2010} = \frac{1}{3} \sum_{\tau=2010-1}^{2010+1} y_\tau = \frac{1}{3} (y_{2009} + y_{2010} + y_{2011}) = \frac{1}{3} (530,4 + 588,1 + 645,3) = 587,93 \text{ usw.}$$

### Gleitende Durchschnitte ungerader Ordnung

$$y'_t = \frac{1}{2k+1} \sum_{\tau=t-k}^{t+k} y_\tau$$

### Gleitende Durchschnitte gerader Ordnung

$$y'_t = \frac{1}{2k} \left( \frac{1}{2} y_{t-k} + \sum_{\tau=t-(k-1)}^{t+(k-1)} y_\tau + \frac{1}{2} y_{t+k} \right)$$

#### Beispiel:

---

4-gliedrige gleitende Durchschnitte beim Volkseinkommen

$$y'_{2008} = \frac{1}{4} (0,5 \cdot y_{2006} + y_{2007} + y_{2008} + y_{2009} + 0,5 \cdot y_{2010})$$

$$y'_{2009} = \frac{1}{4} (0,5 \cdot y_{2007} + y_{2008} + y_{2009} + y_{2010} + 0,5 \cdot y_{2011})$$


---

#### Wichtiger Anwendungsfall:

Gleitende 12-Monats-Durchschnitte

$$y'_t = \frac{1}{12} \left( 0,5 y_{t-6} + \sum_{\tau=t-5}^{t+5} y_\tau + 0,5 y_{t+6} \right)$$

## 2.3 Exponentielle Glättung

Bei der Methode der gleitenden Durchschnitte gehen die Werte jeweils mit gleichen Gewichten ein.

Bei der **Methode der exponentiellen Glättung** (R. G. Brown (1963)) ('exponential smoothing') werden abnehmende Gewichte der Vergangenheitswerte eingesetzt.

### a) Exponentielle Glättung 1. Ordnung

Annahme: kein Trend, keine starken periodischen Schwankungen

$$y'_t = a y_t + (1-a) y'_{t-1}, \quad y_t = \text{beobachteter Wert}$$

$y'_t$  = geglätteter Wert

$a$  = Glättungsfaktor ( $0 \leq a \leq 1$ )

$$a = 0 \quad y'_t = y'_{t-1}$$

$$a = 1 \quad y'_t = y_t$$

Damit ist  $y'_t$  ein gewogenes arithmetisches Mittel aus  $y_t$  und  $y'_{t-1}$ .

Praxis: Es wird ein Glättungsfaktor  $0,01 \leq a \leq 0,3$  gewählt.

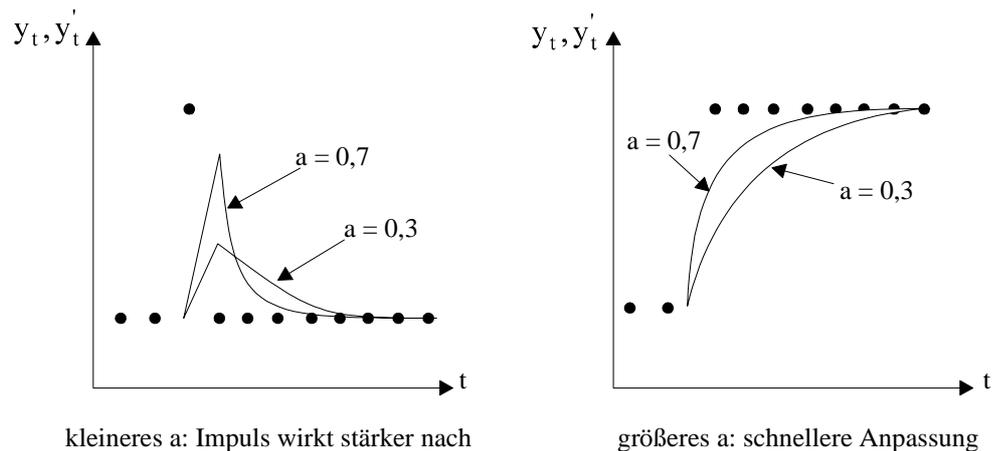


Abb. VII.4: Exponentielle Glättung mit unterschiedlichen Gewichten  
Quelle: Schwarze (1990), S. 230

### Prognose:

$$\hat{y}_{t+1} = y'_t = a \cdot y_t + (1-a) y'_{t-1} = a \cdot y_t + (1-a) \hat{y}_t$$

Prognose als gewogenes Mittel aus **allen** Vergangenheitswerten:

$$y'_t = a y_t + (1-a) y'_{t-1} = a y_t + (1-a) (a y_{t-1} + (1-a) y'_{t-2})$$

$$y'_{t-1} = a y_{t-1} + (1-a) y'_{t-2}$$

⋮

$$y'_t = a y_t + a(1-a) y_{t-1} + a(1-a)^2 y_{t-2} + \dots + (1-a)^{t-1} y_1$$

Durch 'unendlich'-males Einsetzen (für große t oder a nahe 1 kann der Startwert  $y'_1$  durch  $y_1$  ersetzt werden) erhält man:

$$y'_t = \sum_{\tau=0}^{t-1} a \cdot (1-a)^\tau \cdot y_{t-\tau}$$

### b) Exponentielle Glättung 2. Ordnung (Trendmodell)

Bisher: Die Niveaushiftung oder Trendentwicklung wird durch die exponentielle Glättung 1. Ordnung nur zeitlich verzögert angegeben.

Zur Auflösung dieser systematischen 'Verzerrung': Bei der exponentiellen Glättung 2. Ordnung wird die exponentielle Glättung noch einmal auf die schon geglätteten Werte angewendet.

$$y''_t = a \cdot y'_t + (1-a) \cdot y''_{t-1} \quad a = \text{Glättungsparameter}$$

### Ansatz der exponentiellen Glättung:

$$\hat{y}_{t+r} = \bar{\alpha}_t + \bar{\beta}_t \cdot r \quad MKQ: \hat{y}_t = \alpha + \beta \cdot t \quad (t=1, \dots, T)$$

Modell mit 2 Gleichungen:

$$y_t' = a \cdot y_t + (1-a) \cdot y_{t-1}'$$

$$y_t'' = a \cdot y_t' + (1-a) \cdot y_{t-1}''$$

Ziel: Berechnung von  $\bar{\alpha}_t$  und  $\bar{\beta}_t$

Notwendig: Mittleres Alter der einbezogenen Werte:  $\bar{t} = ?$

$y_t$  hat das Alter = 0 mit Gewicht  $a$

$y_{t-1}$  hat das Alter = 1 mit Gewicht  $a(1-a)$

...

$$\bar{t} = 0 \cdot a + 1 \cdot a(1-a) + 2 \cdot a(1-a)^2 + \dots$$

$$= a \sum_{i=0}^{\infty} i(1-a)^i$$

Dies ist eine geometrische Reihe. Man erhält als Ergebnis:

$$\bar{t} = \frac{1-a}{a}$$

Der Wert  $y_t'$  hat das Alter =  $\frac{(1-a)}{a}$  bezogen auf  $t - \left(\frac{1-a}{a}\right)$

Der Wert  $y_t''$  hat das Alter =  $\frac{(1-a)}{a}$  bezogen auf  $t - 2\left(\frac{1-a}{a}\right)$

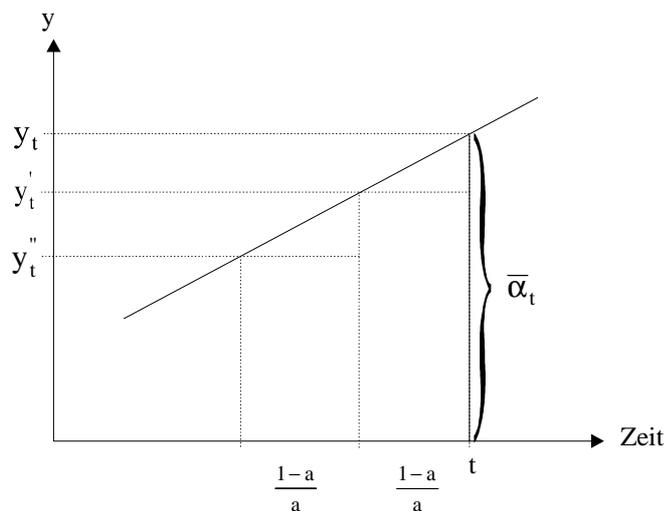


Abb. VII.5: Ansatz der exponentiellen Glättung im Trendmodell ( $\hat{y}_{t+0} = \bar{\alpha}_t + \bar{\beta}_t \cdot 0$ )

$$\text{Steigung: } \bar{\beta}_t = \frac{y_t' - y_t''}{\frac{1-a}{a}} = \frac{y_t' - y_t''}{\bar{t}} = \frac{y_t' - y_t''}{1-a} \cdot a$$

Man kann nun zeigen, dass der zweifach exponentiell geglättete Wert  $y_t''$  hinter dem einfach geglätteten Wert  $y_t'$  um den gleichen Betrag zurückbleibt, wie  $y_t$  und  $y_t'$ :

$$\bar{\alpha}_t = y_t' + (y_t' - y_t'') = 2 \cdot y_t' - y_t''$$

**Prognose:**

$$\hat{y}_{t+r} = \bar{\alpha}_t + \bar{\beta}_t \cdot r$$

$$\hat{y}_{t+1} = \bar{\alpha}_t + \bar{\beta}_t \cdot 1 \quad \text{bei z.B. } r = 1$$

**Beispiel:**

t	$y_t$	$a \cdot y_t$	$(1-a)y_{t-1}'$	$y_t'$	$a \cdot y_t'$	$(1-a)y_{t-1}''$	$y_t''$
0				2478			2478
1	2478	496	1982	2478	496	1982	2478
2	2073	415	1982	2397	479	1982	2462
3	2048	410	1918	2327	465	1969	2435
4	2267	453	1862	2315	463	1948	2411

t = 2

$$y_t' = a \cdot y_t + (1-a) \cdot y_{t-1}' \quad a = 0,2$$

$$= 0,2 \cdot 2073 + 0,8 \cdot 2478 = 2397$$

$$y_t'' = a \cdot y_t' + (1-a) \cdot y_{t-1}''$$

$$= 0,2 \cdot 2397 + 0,8 \cdot 2478 = 2462$$

Berechnung von  $\bar{\alpha}_t$  und  $\bar{\beta}_t$ :

$$\bar{\beta}_t = \frac{y_t' - y_t''}{1-a} \cdot a \quad \text{z.B. } t = 3 \quad \bar{\beta}_3 = \frac{2327 - 2435}{0,8} \cdot 0,2 = -27$$

$$\bar{\alpha}_t = 2 \cdot y_t' - y_t'' \quad \bar{\alpha}_3 = 2 \cdot 2327 - 2435 = 2219$$

**Prognose:**

$$\hat{y}_{t+r} = \bar{\alpha}_t + \bar{\beta}_t \cdot r$$

$$\hat{y}_{t+1} = 2219 - 27 \cdot 1 = 2192 \quad \text{bei z.B. } r = 1$$

Zur Wahl des Glättungsfaktors (Gewichtungsfaktors) a:

a wird vorgegeben. Prinzipiell gilt: Der neueste Beobachtungswert wird umso stärker berücksichtigt, je größer a ist. Stärkere Glättung erfolgt bei kleinerem a.

### 3 Bestimmung der Saisonkomponente

Unterscheidung zwischen Saisonbereinigung bei

- konstanter oder
- variabler Saisonfigur

konstante Saisonfigur: unabhängig von der Periode (Jahr)  
variable Saisonfigur: abhängig vom jeweiligen Jahr

#### 3.1 Saisonbereinigung bei konstanter Saisonfigur

Die Schätzung der glatten Komponente  $G_t$  erfolgt durch die Ermittlung von  $y'_t$  (gleitender Durchschnitt).

Aus dem additiven Modell folgt:

$$y_t = \underbrace{T_t + Z_t}_{G_t = y'_t} + S_t + U_t$$

$$\Rightarrow y_t - y'_t = S_t + U_t \quad \text{Saison + Zufall}$$

#### Berechnung der saisonbereinigten Zeitreihe

1. Für die Zeitreihe  $y_t$  wird der gleitende Durchschnitt  $y'_t$  berechnet (glatte Komponente  $y'_t = G_t = T_t + Z_t$ )
2. Bildung der um die glatte Komponente bereinigten Zeitreihe  $y_t^* = y_t - y'_t$
3. Mittelung der Werte aus 2. ergibt die Saisonindexziffer  $\bar{S}_j$ :

$$\bar{S}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} (y_{i,j} - y'_{i,j})$$

für  $y_{i,j}$  bzw.  $y'_{i,j}$        $i$ =Jahr und  $j$ =Monat

$m_j$  = Anzahl der Jahre für die der Wert des  $j$ -ten Monats berechnet wird

4. Normierte Saisonindexziffer  $\hat{S}_j$ :

Für die konstante Saisonperiode müßte die Summe der  $\bar{S}_j$  über eine Periodenlänge Null ergeben:

$$\tilde{S}_1 + \tilde{S}_2 + \dots + \tilde{S}_{12} = 0 \quad (\text{Monatswerte}).$$

Da Restschwankungen bleiben, muß die Normierung so vorgenommen werden, dass die normierte Summe Null wird:

$$\hat{S}_j = \tilde{S}_j - \frac{1}{\lambda} \sum_{j=1}^{\lambda} \tilde{S}_j \quad \text{wobei } \lambda = \text{Ordnung der gleitenden } \emptyset \ (\emptyset = \text{Durchschnitt})$$

z.B. 12 - Monats- $\emptyset$        $\lambda = 12$

    Quartals  $\emptyset$            $\lambda = 4$

5. Die saisonbereinigte Zeitreihe ergibt sich als Differenz:  $y_{i,j} - \hat{S}_j$

**Beispiel:**

Tab. VII.2: Umsatzzahlen 2009, 2010, 2011 [in Mio EUR]

Monat	2009		2010		2011	
	$y_t$	12- $\emptyset$	$y_t$	12- $\emptyset$	$y_t$	12- $\emptyset$
Jan.	21	-	26	40	34	46
Feb.	29	-	34	41	41	46
März	38	-	40	42	46	45
April	32	-	36	42	37	43
Mai	31	-	24	42	35	42
Juni	24	-	28	42	37	41
Juli	34	39	43	43	42	-
Aug.	54	39	69	44	61	-
Sept.	63	39	63	44	47	-
Okt.	52	40	60	44	41	-
Nov.	46	40	46	45	35	-
Dez.	38	39	42	46	37	-

1. Berechnung des gleitenden Durchschnittes:

$$\begin{aligned} y'_t &= y'_{2011,2} = \frac{1}{12} \left( 0,5 \cdot y_{t-6} + \sum_{\tau=t-5}^{t+5} y_{\tau} + 0,5 \cdot y_{t+6} \right) \\ &= \frac{1}{12} (0,5 \cdot 69 + 63 + 60 + 46 + 42 + 34 + 41 + 46 + 37 + 35 + 37 + 42 + 0,5 \cdot 61) = 45,67 = 46 \end{aligned}$$

$m_j$	Monat j	Jan	Feb	Mrz	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	
2.	1	$y_t - y'_t$	-14	-7	-2	-6	-18	-14	-5	15	24	12	6	-1
	2	$y_t - y'_t$	-12	-5	1	-6	-7	-4	0	25	19	16	1	-4
3.	$\tilde{S}_j$	$= \frac{1}{2} \sum_{i=1}^2 (y_{i,j} - y'_{i,j})$	-13	-6	-0,5	-6	-12,5	-9	-2,5	20	21,5	14	3,5	-2,5
4.	$\hat{S}_j$		-13,6	-6,6	-1,1	-6,6	-13,1	-9,6	-3,1	19,4	20,9	13,4	2,9	-3,1

$$\text{Saisonindexziffern: } \tilde{S}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} (y_{i,j} - y'_{i,j})$$

$$\hat{S}_j = \tilde{S}_j - \frac{1}{\lambda} \sum_{j=1}^{\lambda} \tilde{S}_j = \tilde{S}_j - \underbrace{\frac{1}{12} \sum_{j=1}^{12} \tilde{S}_j}_{\text{Korrektur}} = \tilde{S}_j - \underbrace{0,583}_{\approx 0,6}$$

### 5. Saisonbereinigte Zeitreihe y minus/Saison

Tab. VII.3: Saisonbereinigte Zeitreihe

Monat	2009	2010	2011
Jan.	34,6	39,6	47,6
Feb.	35,6	40,6	47,6
März	39,1	41,1	47,1
April	38,6	42,6	43,6
Mai	44,1	37,1	48,1
Juni	33,6	37,6	46,6
Juli	37,1	46,1	45,1
Aug.	34,6	49,6	41,6
Sept.	42,1	42,1	26,1
Okt.	38,6	46,6	27,6
Nov.	43,1	43,1	32,1
Dez.	41,1	45,1	40,1

```

? ZELTREIHENANALYSE UMSATZENTWICKLUNG
? -----
?
? read: file=zr.dat;nvar=4;nobs=36;names=l$
? list: time, y, glatt, saison$
? -----
?
? regression by et
? -----
?
? regres, dep=y; ind=one,time$
? -----
?
? compute time serie components
? -----
?
? create; trend = yfit
? ; zyklus = y - yfit - zyklus - saison
? ; tz = trend + zyklus
? ; tzs = tz + saison
? ; glattber = y - glatt
? ; saiber = y - saison
? ; trendber = y - yfit
? ; zyklber = y - zyklus$
?
? list; y, trend, zyklus, saison, u$
? list; y, trend, tz, glatt, tzs$
? list; y, trendber, zyklber, glattber, saiber$
?
? ET ERgebnisprotokoll
? -----

```

Observation	Y	TREND	TZ	GLATT	TZS
1	21.000	35.721	37.732	19.400	-4.4000
2	29.000	36.008	38.020	19.400	3.1000
3	38.000	36.295	38.1307	13.400	-1.4000
4	32.000	36.583	38.1595	3.1000	3.1000
5	31.000	36.870	38.882	-3.1000	2.1000
6	24.000	37.158	39.169	-13.600	-4.0000
7	34.000	37.445	39.457	-6.6000	-4.0000
8	54.000	37.732	39.457	-6.6000	-4.0000
9	63.000	38.020	2.2559	-1.1000	-9.0000
10	52.000	38.307	40.032	1.9685	.6000
11	46.000	38.595	40.319	1.6811	.6000
12	38.000	38.882	40.606	-9.4000	-4.9000
13	26.000	39.169	40.894	1.3937	-4.4000
14	34.000	39.457	41.181	3.1000	3.1000
15	40.000	39.744	41.468	19.400	5.6000
16	36.000	40.032	41.756	20.900	-1.9000
17	24.000	40.319	42.043	13.400	2.6000
18	28.000	40.606	42.331	2.9568	-1.9000
19	43.000	40.894	42.618	3.6694	-9.0000
20	69.000	41.181	42.905	-13.600	1.6000
21	63.000	41.468	43.193	-6.6000	1.6000
22	60.000	41.756	43.480	-1.1000	2.1000
23	46.000	42.043	43.768	-4.8018	.6000
24	42.000	42.331	44.055	-13.100	6.1000
25	34.000	42.618	44.342	-9.6000	5.6000
26	41.000	42.905	44.630	-3.1000	
27	46.000	43.193	44.917	19.400	
28	37.000	43.480	45.205	20.900	
29	35.000	43.768	45.492	13.400	
30	37.000	44.055	45.779	2.9000	
31	42.000	44.342		-3.1000	
32	61.000	44.630		19.400	
33	47.000	44.917		20.900	
34	42.000	45.205		13.400	
35	35.000	45.492		2.9000	
36	37.000	45.779		-3.1000	

DATA LISTING (current sample)

Observation	Y	TREND	TZ	GLATT	TZS
1	21.000	35.721	37.732	19.400	39.900
2	29.000	36.008	38.020	19.400	39.000
3	38.000	36.295	38.1307	13.400	39.000
4	32.000	36.583	38.1595	3.1000	40.000
5	31.000	36.870	38.882	40.000	40.000
6	24.000	37.158	39.169	40.000	39.000
7	34.000	37.445	39.457	41.000	39.000
8	54.000	37.732	39.457	41.000	39.000
9	63.000	38.020	2.2559	42.000	39.000
10	52.000	38.307	40.032	42.000	39.000
11	46.000	38.595	40.319	42.000	40.000
12	38.000	38.882	40.606	42.000	40.000
13	26.000	39.169	40.894	42.000	39.000
14	34.000	39.457	41.181	43.000	39.900
15	40.000	39.744	41.468	44.000	39.400
16	36.000	40.032	41.756	44.000	40.900
17	24.000	40.319	42.043	44.000	42.000
18	28.000	40.606	42.331	44.000	42.000
19	43.000	40.894	42.618	44.000	43.000
20	69.000	41.181	42.905	44.000	43.000
21	63.000	41.468	43.193	44.000	43.000
22	60.000	41.756	43.480	44.000	44.000
23	46.000	42.043	43.768	44.000	44.000
24	42.000	42.331	44.055	44.000	45.000
25	34.000	42.618	44.342	44.000	45.000
26	41.000	42.905	44.630	44.000	46.000
27	46.000	43.193	44.917	44.000	46.000
28	37.000	43.480	45.205	44.000	46.000
29	35.000	43.768	45.492	44.000	46.000
30	37.000	44.055	45.779	44.000	46.000
31	42.000	44.342		44.000	46.000
32	61.000	44.630		44.000	46.000
33	47.000	44.917		44.000	46.000
34	42.000	45.205		44.000	46.000
35	35.000	45.492		44.000	46.000
36	37.000	45.779		44.000	46.000

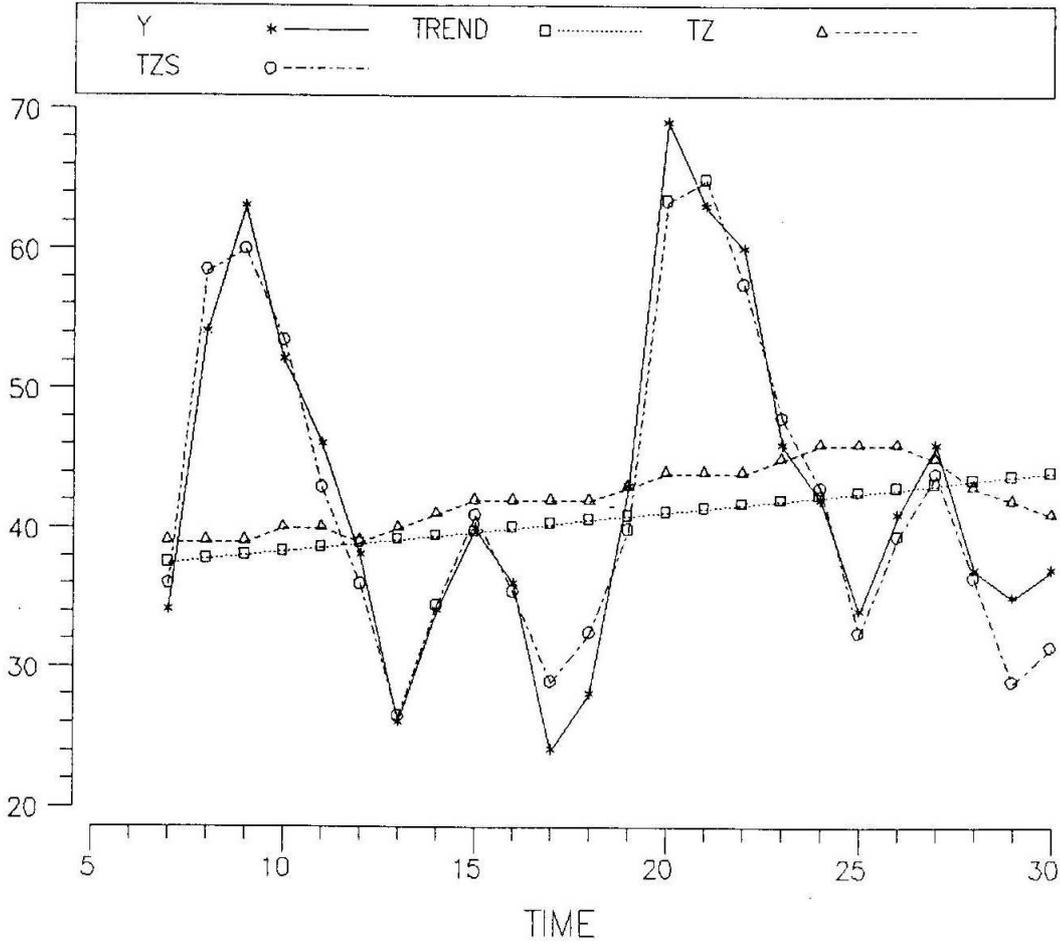
Ordinary Least Squares	Y	Number of Observations	36			
Mean of Dep. Variable	40.7500	Std. Dev. of Dep. Var.	11.864413			
Std. Error of Regr.	11.6390	Sum of Squared Residuals	4605.88			
R - squared	.06513	Adjusted R - squared	.03763			
F ( 1, 34)	2.3686	Prob. Value for F	.13305			
Variable Coefficient	Std. Error	t-ratio	Prob(t)>x	Mean of X	Std.Dev. of X	
Constant	35.4333	3.962	8.943	.00000	18.50000	10.53565
TIME	.287387	.1867	1.539	.13305		

```

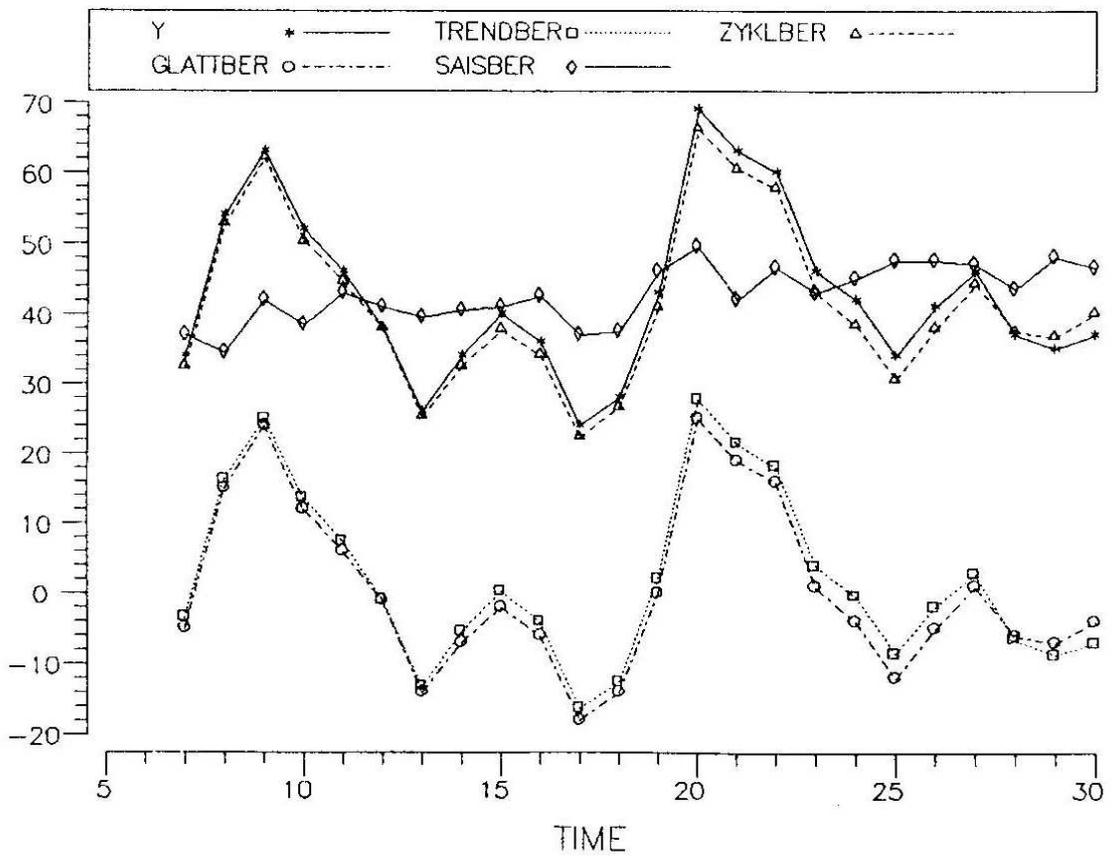
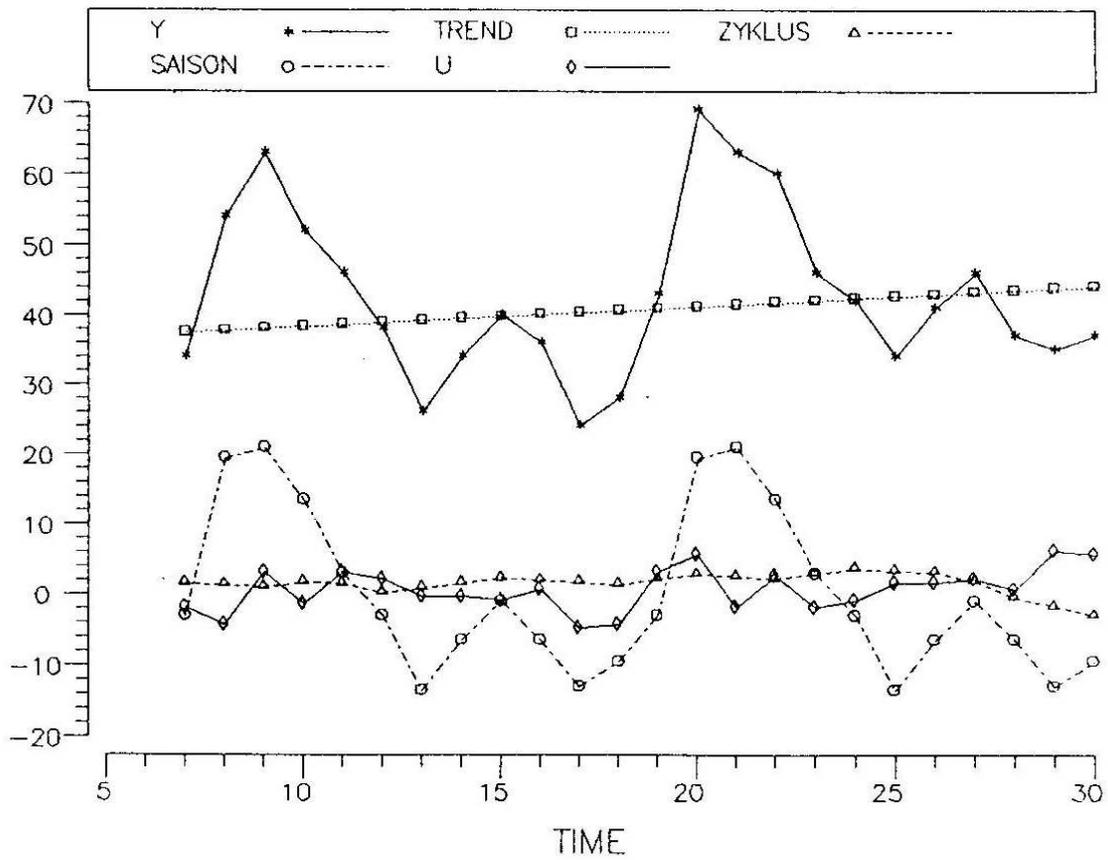
1. Command: -> TREND=YFIT
2. Command: -> ZYKLUS=GLATT-YFIT
3. Command: -> U=Y-YFIT-ZYKLUS-SAISON
4. Command: -> TZ=TREND+ZYKLUS
5. Command: -> TZS=TZ-SAISON
6. Command: -> GLATTBER=Y-GLATT
7. Command: -> SAIBER=Y-SAISON
8. Command: -> TRENDBER=Y-YFIT
9. Command: -> ZYKLBER=Y-ZYKLUS

```

Observation	Y	TREND	TZ	GLATT	TZS
1	21.000	35.721	37.732	19.400	39.900
2	29.000	36.008	38.020	19.400	39.000
3	38.000	36.295	38.1307	13.400	39.000
4	32.000	36.583	38.1595	3.1000	40.000
5	31.000	36.870	38.882	40.000	40.000
6	24.000	37.158	39.169	40.000	39.000
7	34.000	37.445	39.457	41.000	39.000
8	54.000	37.732	39.457	41.000	39.000
9	63.000	38.020	2.2559	42.000	39.000
10	52.000	38.307	40.032	42.000	39.000
11	46.000	38.595	40.319	42.000	40.000
12	38.000	38.882	40.606	42.000	40.000
13	26.000	39.169	40.894	42.000	39.000
14	34.000	39.457	41.181	43.000	39.900
15	40.000	39.744	41.468	44.000	39.400
16	36.000	40.032	41.756	44.000	40.900
17	24.000	40.319	42.043	44.000	42.000
18	28.000	40.606	42.331	44.000	42.000
19	43.000	40.894	42.618	44.000	43.000
20	69.000	41.181	42.905	44.000	43.000
21	63.000	41.468	43.193	44.000	43.000
22	60.000	41.756	43.480	44.000	44.000
23	46.000	42.043	43.768	44.000	44.000
24	42.000	42.331	44.055	44.000	45.000
25	34.000	42.618	44.342	44.000	45.000
26	41.000	42.905	44.630	44.000	46.000
27	46.000	43.193	44.917	44.000	46.000
28	37.000	43.480	45.205	44.000	46.000
29	35.000	43.768	45.492	44.000	46.000
30	37.000	44.055	45.779	44.000	46.000
31	42.000	44.342		44.000	46.000
32	61.000	44.630		44.000	46.000
33	47.000	44.917		44.000	46.000
34	42.000	45.205		44.000	46.000
35	35.000	45.492		44.000	46.000
36	37.000	45.779		44.000	46.000



Observation	Y	TREND	ZYKLEBER	GLATTBER	SAISBER
26	41.000	42.905			39.400
27	46.000	43.193			45.000
28	37.000	43.480			43.900
29	35.000	43.768			36.400
30	37.000	44.055			42.000
31	42.000	44.342			28.900
32	61.000	44.630			31.400
33	47.000	44.917			
34	42.000	45.205			
35	35.000	45.492			
36	37.000	45.779			
1	21.000	-14.721			34.600
2	29.000	-7.0081			35.600
3	38.000	1.7045			39.100
4	32.000	-4.5829			38.600
5	31.000	-5.8703			44.100
6	24.000	-13.158			33.600
7	34.000	-3.4450	32.445	-5.0000	37.100
8	54.000	16.268	52.732	15.000	34.600
9	63.000	24.980	62.020	24.000	42.100
10	52.000	13.693	50.307	12.000	38.600
11	46.000	7.4054	44.595	6.0000	43.100
12	38.000	-8.8198	37.882	-1.0000	41.100
13	26.000	-13.163	25.169	-14.000	39.600
14	34.000	-5.4568	32.457	-7.0000	40.600
15	40.000	-25.886	37.734	-2.0000	41.100
16	36.000	-4.0315	34.032	-6.0000	42.600
17	24.000	-16.319	22.319	-18.000	37.100
18	28.000	-12.606	26.606	-14.000	46.100
19	43.000	2.1963	40.894	-0.0000	49.600
20	69.000	27.819	66.181	25.000	42.100
21	63.000	21.532	60.458	16.000	46.600
22	60.000	18.244	57.756	1.0000	43.100
23	46.000	3.9568	43.043	-1.0000	45.100
24	42.000	-3.3063	38.331	-12.000	47.600
25	34.000	-8.6180	30.618	-5.0000	47.100
26	41.000	-1.9054	37.905	-1.0000	48.100
27	46.000	2.8072	44.193	-6.0000	46.600
28	37.000	-6.4802	37.480	-7.0000	45.100
29	35.000	-8.7676	36.768	-4.0000	41.600
30	37.000	-7.0550	40.055		26.100
31	42.000	-2.3423			23.600
32	61.000	16.370			32.100
33	47.000	2.0829			40.100
34	42.000	-3.2045			
35	35.000	-10.492			
36	37.000	-8.7793			



### 3.2 Saisonbereinigung bei variabler Saisonfigur

Wenn ein Zeitreihendiagramm (Scatterplot) darauf hinweist, dass die Saisonschwingung mit der glatten Komponente wächst, dann sollte ein **Modell mit variabler Saisonfigur** verwendet werden.

Voraussetzung: Die Saisonkomponente für den Monat (Quartal)  $j$  beträgt ein **Vielfaches der glatten Komponente**, also  $S_{i,j} = \alpha_j \cdot G_{i,j}$

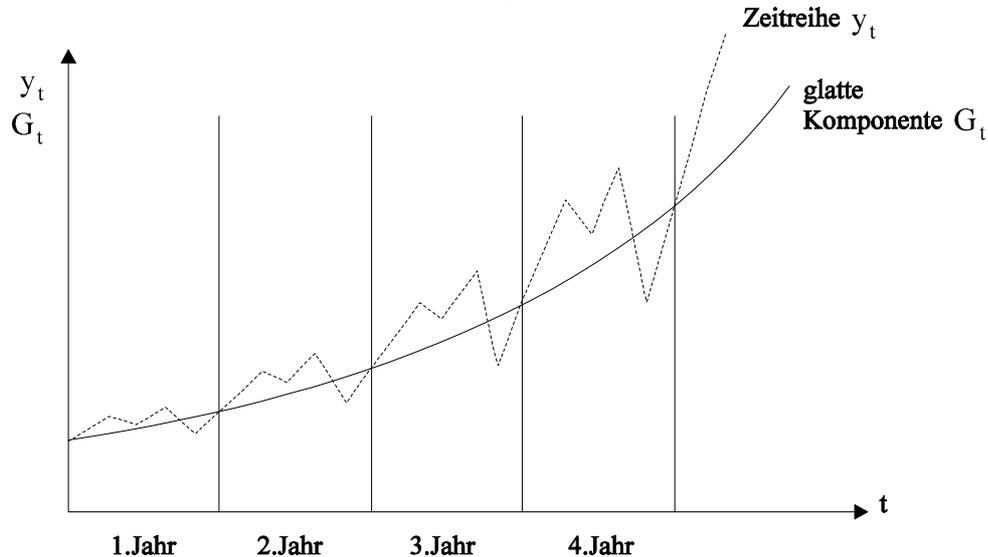


Abb. VII.6: Zeitreihe mit variabler Saisonfigur

Mit obiger Saisonkomponente folgt das Zeitreihenmodell

$$y_{i,j} = G_{i,j} + S_{i,j} + U_{i,j} = G_{i,j} + \alpha_j \cdot G_{i,j} + U_{i,j} = I_j \cdot G_{i,j} + U_{i,j}$$

mit  $I_j = (1 + \alpha_j)$  = Saisonindexziffer für den Monat (Quartal)  $j$

#### Berechnung der saisonbereinigten Zeitreihe

1. Gleitende Durchschnitte  $\bar{y}_t$  (glatte Komponente) bilden
2. Bildung der um die glatte Komponente bereinigten Zeitreihe:

$$\frac{y_{i,j}}{\bar{y}_{i,j}} = \frac{I_j \cdot G_{i,j} + U_{i,j}}{G_{i,j}} = I_j + \frac{U_{i,j}}{G_{i,j}}$$

3. Saisonindexziffern: Mittelung dieser Quotienten über alle Jahre  $i$

$$\tilde{I}_j = \frac{1}{m_j} \sum_{i=1}^{m_i} \frac{y_{i,j}}{\bar{y}_{i,j}} \quad \begin{array}{l} i = \text{Jahr} \\ j = \text{Monat} \end{array}$$

#### 4. Normierte Saisonindexziffern $\hat{I}_j$ :

$$\hat{I}_j = \frac{\tilde{I}_j}{\frac{1}{\lambda} \sum_{j=1}^{\lambda} \tilde{I}_j} = \tilde{I}_j \cdot \frac{\lambda}{\sum_{j=1}^{\lambda} \tilde{I}_j} \quad \lambda = 12, \text{ wegen 12 - Monats - Durchschnitt}$$

#### 5. Saisonbereinigte Zeitreihe als Quotient

$$\frac{y_{i,j}}{\hat{I}_j}$$

Diese Saisonbereinigungsverfahren sind lediglich ein Einstieg in die Problematik und für die manuelle Saisonbereinigung gedacht.

Institutionen verwenden heute Verfahren, die komplexer sind, z.B.:

- Census II-Verfahren (Variante X-11)  
seit 1970 von der Deutschen Bundesbank und US Bureau of the Census verwendet;  
Lit.: Monatsbericht der Deutschen Bundesbank 1970
- Berliner Verfahren (spektralanalytischer Ansatz);  
Lit.: Nullau, B. (1969)

## 4 Spektralanalyse: Zeitreihenanalyse im Frequenzbereich

### 4.1 Spektralanalyse

Eine Spektralanalyse ist eine Zeitreihenanalyse auf der Basis stochastischer Modelle. Die Zeitreihenwerte werden hier als Realisationen eines stochastischen Prozesses betrachtet. Die Zeitreihenwerte werden im Frequenzbereich aus überlagerten Schwingungen analysiert.  
Ziel: Dominante Schwingungen bestimmter Frequenz herausarbeiten.

Die Spektralanalyse verwendet dabei so genannte schwach stationäre Prozesse  $\{X_t\}$ .

Analyse eines stochastischen Prozesses

- im Zeitbereich (t)
- im Frequenzbereich ( $\omega$ )  
(vgl. z.B. Merz (1980b), Prognosegüte und Spektraleigenschaften ökonomischer Modelle)

#### Zeitbereich

Autokorrelation

$$\rho_\tau = \frac{\text{Cov}(X_t, X_{t+\tau})}{\text{Var}(X_t)} = \frac{\gamma_\tau}{\gamma_0}$$

Überlagern sich mehrere Schwingungen, wird es schwierig, aus dem Korrelogramm die einzelnen Schwingungen zu isolieren, da die Autokorrelationskoeffizienten gewogene Mittel aller Schwingungskomponenten sind.

**Frequenzbereich:**

Fourier-Transformation der Autokovarianzfolge ergibt die

**Spektraldichtefunktion**

$$\gamma_\tau = \int_{-\pi}^{\pi} \cos \omega\tau dF(\omega)$$

$\omega = 2\pi\nu$  in Bogenmaß gemessene Kreisfrequenz

$T = \frac{1}{\nu}$  Zeitdauer einer Schwingung

Die einzelnen Schwingungen sind nun isoliert und ihre jeweilige Periodendauer bestimmbar.

 **$F(\omega)$  Spektralverteilungsfunktion**

$$\sigma^2 = \gamma_0 = 2 \int_0^{\pi} dF(\omega) = 2 \int_0^{\pi} f(\omega) d\omega = F(\pi) \quad \text{Varianz}$$

$f(\omega)$  entspricht dem Spektrum,  $F(\omega)$  kumuliert die Dichten und kann als kumulierter Varianzbeitrag (obere Grenze ist  $T^2$ ) interpretiert werden.

**Beispiel:**

Zur Veranschaulichung betrachten wir als Beispiel einen Prozeß und seine Autokovarianzfolge, die sich jeweils aus der Überlagerung endlich vieler Schwingungen ergeben (vgl. Merz (1980b), S. 50-51).

$$X_t = \sum_{j=1}^m a_j \cdot \cos(\omega_j \cdot t + \theta_j) \quad 0 \leq \omega_j \leq \pi,$$

mit  $\theta_j$  als einer über  $[-\pi, \pi]$  gleichverteilten Zufallsvariablen. Die entsprechende Autokovarianzfolge (vgl. Granger und Newbold (1977), Kap. 2.1) ist

$$\gamma_\tau = \frac{1}{2} \sum_{j=1}^m a_j^2 \cos(\omega_j \tau).$$

Die zu  $\gamma_\tau$  zugehörige normierte Spektralverteilungsfunktion lautet

$$F(\omega) = \begin{cases} \frac{1}{2} & 0 \leq \omega \leq \omega_1 \\ \frac{1}{2} \left( \gamma_0 + \frac{1}{2} \sum_{j=1}^k a_j^2 \right) / \gamma_0 & \omega_k \leq \omega < \omega_{k+1}, \quad k=1, \dots, m-1 \\ 1 & \omega_m \leq \omega \leq \pi \end{cases}$$

mit der Varianz des Prozesses  $\gamma_0 = \frac{1}{2} \sum_{j=1}^m a_j^2$ .

Für  $m=3$  Schwingungskomponenten sind eine Realisation des zyklischen Prozesses, das Korrelogramm (Autokorrelationen) und die Spektralverteilungsfunktion mit zugehöriger Dichtefunktion in Abb. VII.7 grafisch dargestellt.

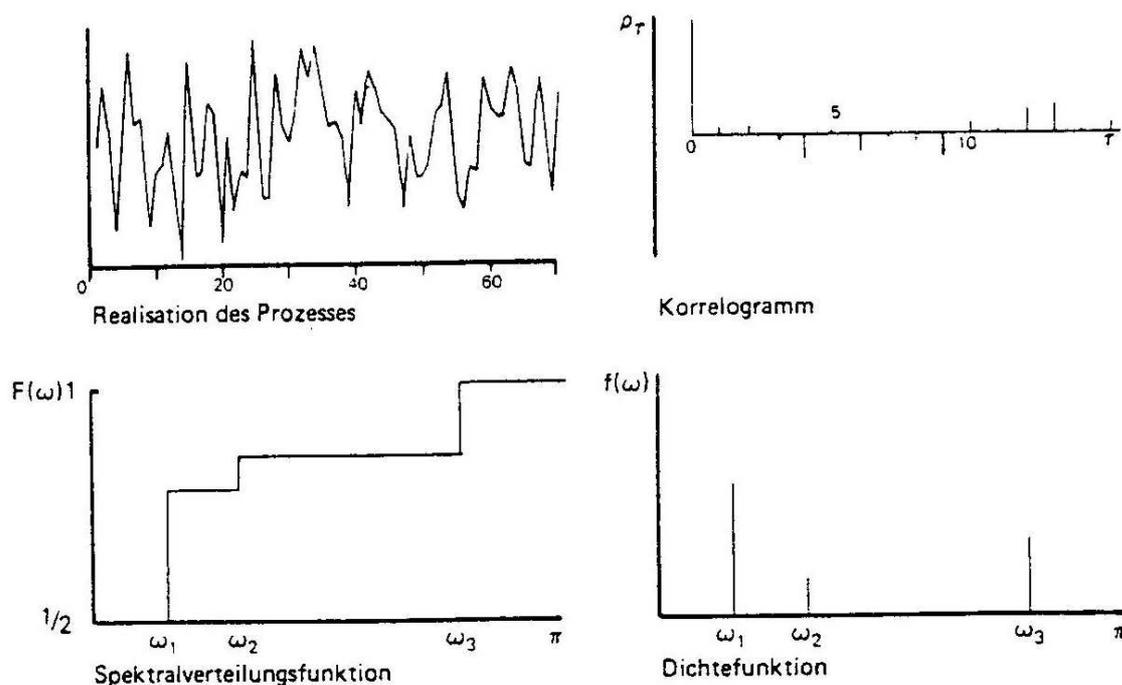


Abb. VII.7: Beispiel eines stochastischen Prozesses mit 3 Schwingungskomponenten:

$$X_t = \sum_{j=1}^3 a_j \cos(\omega_j t + \theta_j); \quad a_1 = 4, a_2 = 2, a_3 = 3, \omega_1 = 0,5, \omega_2 = 1,0, \omega_3 = 2,5$$

Die Zeitreihe als Überlagerung von drei Schwingungen mit den Kreisfrequenzen  $\omega_1, \omega_2, \omega_3$  wird nur deutlich im Frequenzbereich (Spektraldichtefunktion); das Autokorrelogramm als Analyseinstrument im Zeitbereich lässt keine bestimmten ausgeprägten Muster erkennen.

## 4.2 Kreuzspektralanalyse

Die Kreuzspektralanalyse beschreibt **den Zusammenhang zwischen zwei stationär korrelierten stochastischen Prozessen** im Frequenzbereich (Merz (1980b), S. 57 ff).

### Kreuzkorrelationsfolgen

$$\rho_{ij,\tau} = \frac{\gamma_{ij,\tau}}{(\gamma_{ii,0} \cdot \gamma_{jj,0})^{1/2}} = \frac{\text{Cov}(X_{i,t}, X_{j,t+\tau})}{[\text{Var } X_i \cdot \text{Var } X_j]^{1/2}}$$

### Spektraldichtefunktionen

$$F(\omega) = \begin{pmatrix} f_{11}(\omega) & f_{12}(\omega) \\ f_{21}(\omega) & f_{22}(\omega) \end{pmatrix}$$

- **Phasenwinkel**  
zeitliche Verschiebung zyklischer Schwankungen einer bestimmten Frequenz für zwei Prozesse ('lead-lag'-Struktur);
- **Kohärenz**  
Intensität des Zusammenhangs zwischen den Schwingungskomponenten gleicher Frequenz für zwei Prozesse;
- **Gain**  
gibt an, wie die Bewegungskomponenten des einen Prozesses verändert werden müssen, um die entsprechenden Bewegungskomponenten des anderen Prozesses zu erhalten.

### Beispiel: \_\_\_\_\_

Auto- und Kreuzbeziehungen von Moving-Average-Prozessen zweiter Ordnung (Merz (1980b), S. 62-63)

*Prognosegüte und Spektraleigenschaften ökonomischer Modelle*

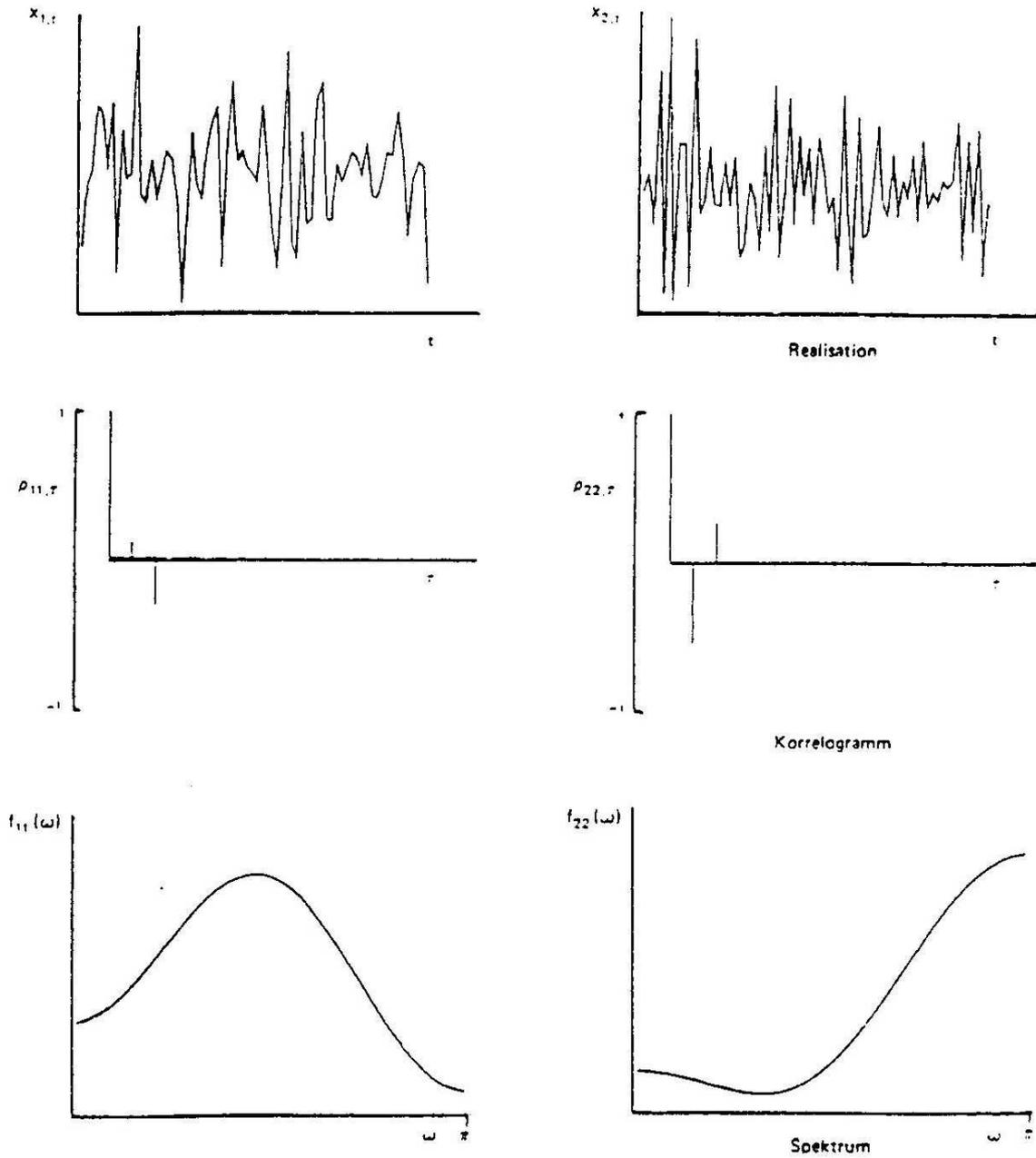


Abb. VII.8: Auto- und Kreuzbeziehungen von  $X_{1,t} = \varepsilon_t + 0,9\varepsilon_{t-1} - 0,7\varepsilon_{t-2}$

Untersuchung der Spektraleigenschaften ökonomischer Modelle

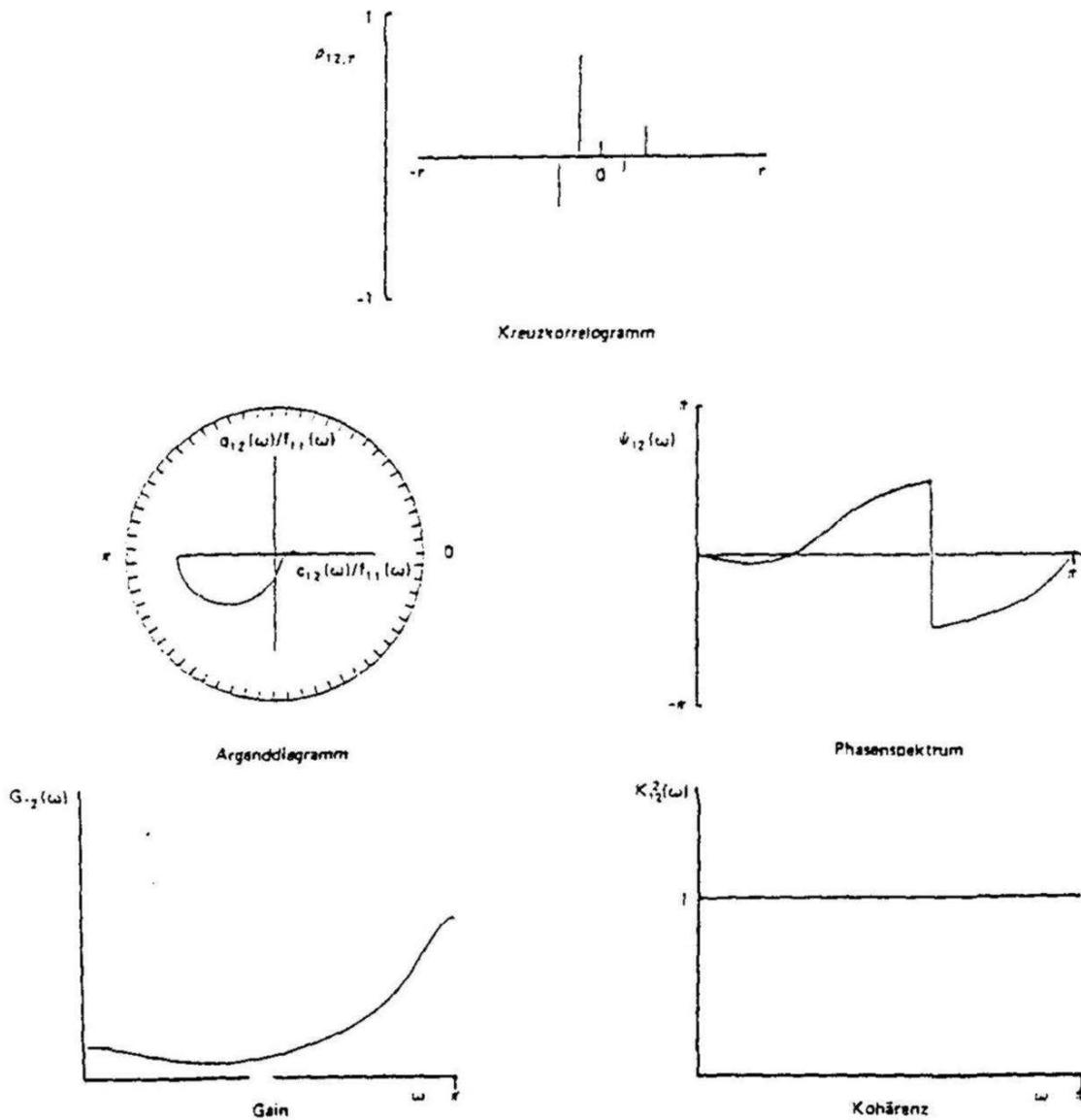


Abb. VII.9: Moving-Average-Prozesse zweiter Ordnung und

$$X_{2,t} = \varepsilon_t - 0,6\varepsilon_{t-1} + 0,4\varepsilon_{t-2}$$

Weitere Informationen zur Zeitreihenanalyse (univariate Ansätze, d.h. nur die Zeitreihe selbst betrachtend): Schlittgen und Streitberg (1987) oder Hansen (1993).

## 5 Querschnittsanalyse

### 5.1 Charakteristika und Beispiele

Während in der Zeitreihenanalyse die zu untersuchenden Merkmalsausprägungen eines Merkmals als Beobachtungen aus verschiedenen Zeitpunkten stammen (Merkmalsträger = Zeitpunkte), stammen die Beobachtungen für Querschnittsanalysen (cross section analyses) aus allen Einheiten (z.B. Personen) eines bestimmten Zeitpunktes.

**Typische Datengrundlage:** Umfragedaten zu einem Zeitpunkt

**Ziel:** Qualifizierung sozio-ökonomischer Einflußfaktoren

**Beispiele:** \_\_\_\_\_

- Kanzlerfrage: 'Wenn Sie am nächsten Sonntag wählen würden, wen würden Sie als Bundeskanzler wählen?' (Liste)
- Sfb3 - Nebenerwerbstätigkeitsumfrage
- Regressionsanalysen wie

$\text{Ausgaben}_i = f(\text{Einkommen}_i, \text{sozio-ökonomische Merkmale}_i (\text{Alter, Geschlecht, Haushaltsgröße etc.}))$

$\text{Arbeitsangebot (Std.)}_i = f(\text{Lohnsatz, sozio-ökonomische Merkmale, regionale Arbeitsmarktlage etc.})$

---

### 5.2 Analysemethoden

Prinzipiell sind alle bisher verwendeten Analysemethoden anwendbar.

Da im allgemeinen die Beobachtungen aus Zufallstichproben stammen, gilt folgende Ausnahme:

Fragen, die die Beziehung zwischen den statistischen Einheiten im Sinne eines Zeitbezugs thematisieren

**Ausnahmebeispiel:** \_\_\_\_\_

Autokorrelation, generell zeitpunktebezogene Methoden

---

## 6 Panelanalyse

### 6.1 Panel- und Längsschnittsanalyse

Paneldaten haben sowohl eine Querschnitts- als auch Längsschnittsdimension, da die statistischen Einheiten zu mehreren Zeitpunkten wiederholt befragt werden.

**Längsschnittsanalyse:** Zeitentwicklung einer statistischen Einheit

**Panelanalysen:** Zeitentwicklung vieler statistischer Einheiten. Die gleichen statistischen Einheiten werden wiederholt befragt.

**Ziel:** Quantifizierung sozio-ökonomischer Einflußfaktoren in zeitlicher Entwicklung.

**Beispiele:**

---

- Sozio-ökonomisches Panel des Sfb3/DIW: Alle Personen über 16 Jahre in ca. 6000 Haushalten werden seit 1984 jährlich wieder befragt.
  - Panel Study of Income Dynamics PSID der University of Michigan seit 1968/69 (5000 families)
  - NIFA-Betriebspanel des Sfb 187  
Betriebe des Maschinenbaus werden jährlich wiederholt befragt.
  - Niedersachsen-Firmenpanel Universität Hannover, Universität Lüneburg (J. Wagner)
- 

### 6.2 Analysemethoden

In jüngerer Zeit wurden in der Ökonometrie verschiedene Methoden entwickelt, die u.a. über eine Aufteilung der Fehlerterme Zeit- und Querschnittseffekteinflüsse quantifizieren (z.B. Quantifizierung der Heterogenität zwischen den Untersuchungseinheiten (z.B. Personen)).

- Fehlerkomponentenmodelle
- Fixed Effect-Modelle und Fixed Coefficient-Modelle
- Random Effect-Modelle

Ausführliche Behandlung in meiner Vorlesung: Panel-Ökonometrie.

Weitere Hinweise: Hsiao (1986), Mátyás und Sevestre (1992)

#### **Keyconcepts**

*Zeitreihenzerlegung*

*Trendkomponente*

*Zyklische Komponente*

*Saisonale Komponente*

*Methode der gleitenden Durchschnitte*

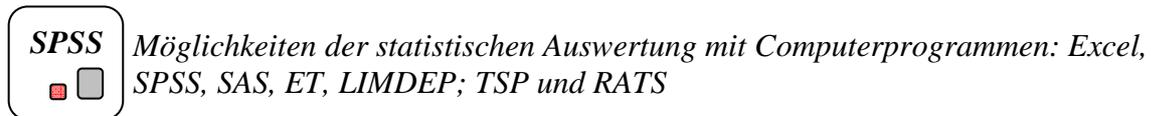
*Methode der exponentiellen Glättung 1. und 2. Ordnung*

*Spektralanalyse*

*Längsschnitt, Querschnitt*

*Panel*

## VIII Computerprogramme zur statistischen Auswertung und Analyse



Mittlerweile gibt es vor allem für den PC-Sektor viele statistische Programmpakete und Auswertungsmöglichkeiten in allgemeineren Paketen (wie z.B. Tabellenkalkulation).

### 1 Statistikauswertungen mit Tabellenkalkulationsprogrammen

Alle neueren Tabellenkalkulationsprogramme wie Lotus 1-2-3, Symphony oder EXCEL haben eine besondere Statistikfunktion, die unter Windows leicht zu bedienen ist.

Die Daten sind leicht in Tabellenform einzubringen und werden dann mit allerdings meist wenigen deskriptiven Maßen analysiert.

#### Beispiel:

---

Microsoft EXCEL

- Dateneingabe über Tabellenhandling
  - Hauptmenü: FORMEL
  - FORMEL: FUNKTION EINFÜGEN: dann
  - STATISTIK  
(Häufigkeiten, Max, Min, Median, Mittelwert, Range etc.)
- 

### 2 SPSS und SAS

SPSS und SAS sind wohl die verbreitetsten allgemeinen und mächtigen Statistik-Softwarepakete. SYSTAT ist ein weiteres weniger umfangreiches Paket (Student-SYSTAT).

SPSS/PC+

PC-Version, Menü-orientiert

SPSS/Windows: FILE, TRANSFORM, STATISTICS, GRAPHS, UTILITIES

Tab. VIII.1: SPSS/PC+ command summary

<b>Function</b>	<b>Commands</b>
<i>Operation commands</i>	
Provide assistance	SHOW, DISPLAY
Specify options for operations and output	SET
Submit SPSS/PC+ commands from a file	INCLUDE
Edit a file	REVIEW
Access DOS or other facilities	DOS, EXECUTE
<i>Data definition and manipulation commands</i>	
Read data	DATA LIST, BEGIN DATA, END DATA, IMPORT GET, TRANSLATE
Transform data	RECODE, COMPUTE, IF, COUNT
Define missing data	MISSING VALUE
Select and weight cases	SELECT IF, PROCESS IF, N, SAMPLE, WEIGHT
Provide labels and formats	TITLE, SUBTITLE, *, VARIABLE LABELS, VALUE LABELS, FORMAT
<i>Procedure commands</i>	
Data display	LIST, PLOT, REPORT, TABLES*
Descriptive statistics	DESCRIPTIVES, FREQUENCIES
Categorical statistics	CROSSTABS, HILOGLINEAR*
Group comparisons	T-TEST, ONEWAY, MEANS, ANOVA
Multivariate statistics	CORRELATION, REGRESSION, CLUSTER*, QUICK CLUSTER*, FACTOR*, DISCRIMINANT*, MANOVA*, RELIABILITY*
Nonparametric statistics	NPART TESTS
Time series analysis*	ACF, AREG, ARIMA, CASEPLOT, CCF, CURVEFIT, EXSMOOTH, FIT, NPLOT, PACF, RMV, SPECTRA
Utilities	WRITE, AGGREGATE, SORT CASES, JOIN, EXPORT, SAVE
Graphics	GRAPH*, MAP*, FASTGRAF*

\* Available only in SPSS/PC+ options

---

**Beispiel:**

FREQUENCIES VARIABLES = var1 /BARChart /STATISTICS ALL

ergibt Häufigkeitsauszählung von var1, Balkendiagramm und umfangreiche Statistikergebnisse

---

Weitere Hinweise: **SPSS-Kurs** als sinnvolle Ergänzung des Studienangebots

## SAS

SAS ist auf Mainframe und PCs (DOS, Windows) und unter UNIX verfügbar.

Interface mit Fenstern, interaktivem und Batch-Mode.

SAS enthält eine Vielzahl statistischer und ökonomischer Methoden und unterstützt verschiedene Fileformate (ASCII, DBASE, LOTUS, ORACLE etc.)

SAS-Programm: DATA  
PROCEDURE  
mit statistischen, grafischen Analysen  
(SAS-IML Matrixlanguage)

### Beispiel:

---

```
PROC MEANS DATA = SAS dataset options;
  options: MAXDEC = n | statistics
  statistics: N | MEAN | STD | MIN | MAX | RANGE | SUM | VAR | SKEWNESS |
  KURTOSIS
VAR variables;
```

---

weitere Hinweise: SAS-Kurs

## 3 ET, LIMDEP, STATA, TSP und RATS

ET (Econometrics Toolkit) haben wir hier in Statistik I - Deskription als ein einfaches aber schon mächtiges, menuegesteuertes PC-Programmpaket verwendet. Mit der ebenso möglichen Befehlssprache können viele Dinge für alle Beobachtungen alleine über Variablenamen berechnet werden (Tests, Matrizen, Inverse von Matrizen etc.).

Die gleiche Befehlssprache ist Grundlage von LIMDEP, einem Programmpaket speziell für beschränkt abhängige Variablen-Probleme (Limited Dependent Variables, LIMDEP). ET/LIMDEP ist deshalb auch Grundlage meiner weiteren Lehrveranstaltungen im Rahmen einer empirischen Wirtschaftsforschung (z.B. für Diskrete Entscheidungsmodelle - Mikro-ökonomie).

STATA ist ein mächtiges Paket, das viele neue ökonomische Module hat. TSP und RATS sind zwei Programmpakete, die sich auf Zeitreihenanalysen konzentrieren. Beide sind interaktiv oder in Batch-Mode ausführbar.

Enjoy it!

### *Keyconcepts*

*EXCEL*

*SPSS*

*SAS*

*ET*

*LIMDEP*

*STATA*

*TSP, RATS*



# Anhang: Formelsammlung Statistik I - Deskription

## I Wirtschafts- und Sozialstatistik

### Bevölkerung und Fläche

$$\text{allg. Bevölkerungsdichte} = \frac{\text{Bevölkerung}}{\text{Fläche in Quadratkilometern}}$$

### Allgemeine Geburten- und Sterbeziffer

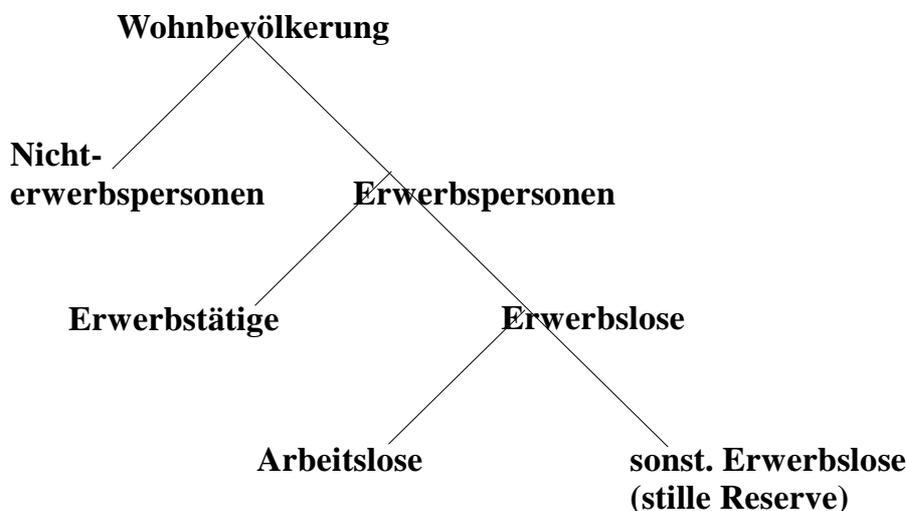
$$\text{Allg. Geburtenziffer} = \frac{\text{Zahl der Lebendgeborenen im Jahr } i \text{ (} G \text{)}}{\text{Mittlerer Bevölkerungsbestand im Jahr } i \text{ (} B \text{)}} \cdot 1000$$

### Besondere Geburten- und Sterbeziffer

$$\text{allg. Fertilitätsrate} = \frac{\text{Lebendgeborene im Jahr } i \cdot 1000}{\text{Anzahl der Frauen zw. 15 u.45 Jahren im Jahr } i}$$

### Erwerbstätigkeit

Gliederung der **Wohnbevölkerung** nach der **Beteiligung am Erwerbsleben** (Erwerbskonzept)



$$\text{Allgemeine Erwerbsquote} = \frac{\text{Erwerbspersonen}}{\text{Wohnbevölkerung}}$$

$$\text{spezifische Erwerbsquote} = \frac{\text{Zahl der Erwerbspersonen}}{\text{Bevölkerung im Alter von 15 - 65 Jahren}}$$

$$\text{Arbeitslosenquote} = \frac{\text{Arbeitslose}}{\text{abhängige (!) Erwerbspersonen}}$$

$$\begin{aligned} \text{Abhängige, zivile Erwerbspersonen} = & \text{ Sozialversicherungspflichtige Beschäftigte} \\ & \text{ (Arbeiter, Angestellte)} \\ & + \text{ Geringfügig Beschäftigte} \\ & + \text{ Beamte} \\ & + \text{ Arbeitslose} \end{aligned}$$

## Exkurs: Summen, Doppelsummen

Häufig hat man es mit Summen endlich vieler Summanden zu tun. Um die Schreibweise zu vereinfachen, können sie mit dem griechischen Sigma  $\sum$  abgekürzt werden.

**Definition:** Das Summenzeichen steht als als Wiederholungszeichen für die fortgesetzte Addition:

$$\sum_{i=k}^m a_i = a_k + a_{k+1} + \dots + a_m, \quad \begin{array}{l} i, k, m \in \mathbb{N} \\ k < m \end{array}$$

wobei:

i = Summationsindex

k = untere Summationsgrenze

m = obere Summationsgrenze

$a_i$  = allg. Summationsglied

**Beispiele:**

$$\text{a) } \sum_{i=1}^3 i^2 = 1^2 + 2^2 + 3^2$$

$$\text{b) } \sum_1^4 c = c + c + c + c = 4c$$

$$\text{c) } \sum_{i=1}^4 x^i = x^1 + x^2 + x^3 + x^4$$

## Zerlegungsregeln für einfache Summen

1. Summe gleicher Summanden

$$\sum_{i=k}^m a = (m-k+1)a \quad , \quad \sum_{i=1}^n a = na$$

2. Summen mit gleicher Summationsvorschrift

$$\sum_{i=k}^m (a_i + b_i) = \sum_{i=k}^m a_i + \sum_{i=k}^m b_i$$

3. Summen mit additiven Konstanten

$$\sum_{i=k}^m (a_i + c) = \sum_{i=k}^m a_i + (m-k+1)c$$

4. Summen mit multiplikativen Konstanten

$$\sum_{i=k}^m ca_i = c \sum_{i=k}^m a_i$$

5. Summenzerlegung

$$\sum_{i=k}^m a_i = \sum_{i=k}^l a_i + \sum_{i=l+1}^m a_i \quad , \quad k \leq l \leq m$$

## II Eindimensionale Häufigkeitsverteilungen

### Qualitative (nominalskalierte) Merkmale

Ausprägung eines qualitativen Merkmals	$A_i$
absolute Häufigkeit eines Merkmals	$n_i = n(A_i)$
Anzahl der verschiedenen Ausprägungen	$k$
Anzahl der Beobachtungen	$n = \sum_{i=1}^k n_i$
relative Häufigkeit eines Merkmales (Häufigkeitsverteilung)	$h(A_i) = \frac{n_i}{n}$

## Quantitative (metrisch skalierte) Merkmale

### Diskrete Merkmale

Merkmalwert	$x_i$
absolute Häufigkeit eines Merkmals	$n_i = n(x_i)$
Anzahl der verschiedenen Merkmalswerte	$k$
Anzahl der Beobachtungen	$n = \sum_{i=1}^k n_i$
relative Häufigkeit eines Merkmalswertes (Häufigkeitsfunktion, -verteilung)	$h(x_i) = \frac{n_i}{n}$
kumulierte absolute Häufigkeit	$n(x \leq x_i) = \sum_{j=1}^i n_j$
kumulierte relative Häufigkeit	$h(x \leq x_i) = \sum_{j=1}^i h(x_j)$
Verteilungsfunktion	$F(x_i) = h(x \leq x_i) = \sum_{j=1}^i h(x_j), \quad i = 1, \dots, k$

### Stetige Merkmale

Merkmalwert	$x$
Klassenuntergrenze der Merkmalsklasse $i$	$x_i^u$
Klassenobergrenze der Merkmalsklasse $i$	$x_i^o$
Klassenbreite	$\Delta x_i = x_i^o - x_i^u$
absolute Häufigkeit der in der Klasse $i$ liegenden Merkmalswerte	$n_i = n(x_i^u \leq x < x_i^o)$
Anzahl der Klassen	$k$
Anzahl der Beobachtungen	$n = \sum_{i=1}^k n_i$
relative Häufigkeit der in der Klasse $i$ liegenden Merkmalswerte (Häufigkeitsverteilung, $i = 1, \dots, k$ )	$h(x_i) = h(x_i^u \leq x < x_i^o) = \frac{n_i}{n}$
normierte relative Häufigkeit (Dichtefunktion, $i = 1, \dots, k$ )	$f(x_i) = \frac{n_i}{n \Delta x_i}$
kumulierte relative Häufigkeit	$h(x \leq x_i^o) = \sum_{j=1}^i h(x_j)$
Interpolation innerhalb der Klasse $i$	$F(x) = F(x_i^u) + \frac{x - x_i^u}{\Delta x_i} \cdot h(x_i)$

## III Lageparameter

### Mittelwerte

#### Häufigster Wert (Modus)

- ungruppiertes Datenmaterial

$$D = x_i \left| \frac{n(x_i)}{n} = \max \right.$$

- gruppiertes Datenmaterial

$$D = x_i^* \left| \frac{n_i}{n\Delta x_i} = \max \right.$$

#### Median (Zentralwert)

- ungruppiertes Datenmaterial

$$\text{falls } n \text{ ungerade} \quad Z = x_{\frac{n+1}{2}}$$

$$\text{falls } n \text{ gerade} \quad Z = \frac{1}{2} \left( x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right)$$

- gruppiertes Datenmaterial: Der Median ist nur approximativ mit Hilfe der Verteilungsfunktion erhältlich. Es gilt:  $h(x \leq Z) = F(Z) = 0,5$

lineare Interpolation bei metrisch skalierten, stetigen Merkmalen:

$$Z = x_i^u + \frac{F(Z) - F(x_i^u)}{h_i} \Delta x_i$$

#### Arithmetisches Mittel

- ungruppiertes Datenmaterial

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- gruppiertes Datenmaterial

$$\text{bekannte Gruppenmittel} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^k \bar{x}_i n_i = \sum_{i=1}^k \bar{x}_i \cdot h(x_i)$$

$$\text{unbekannte Gruppenmittel} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^k x_i^* n_i \quad \text{mit} \quad x_i^* = \frac{1}{2} (x_i^u + x_i^o)$$

#### Geometrisches Mittel

$$GM = \sqrt[n]{\prod_{i=1}^n x_i}, \quad (x_i > 0) \quad \log GM = \frac{1}{n} \sum_{i=1}^n \log x_i$$

#### Harmonisches Mittel

$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

## Streuungsmaße

### Spannweite (range) R

$$R = x_{\max} - x_{\min} = x_{(n)} - x_{(1)}$$

### Quartilsabweichung

$$QA = \frac{1}{2} (x_{0,75} - x_{0,25})$$

### p-Quantile

Interpolationsformel bei gruppiertem Datenmaterial:

$$x_p = x_i^u + \frac{F(x_p) - F(x_i^u)}{\frac{n_i}{n}} \cdot \Delta x_i$$

### Mittlere absolute Abweichung

- ungruppiertes Datenmaterial

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

- gruppiertes Datenmaterial

$$d = \frac{1}{n} \sum_{i=1}^k |x_i^* - \bar{x}| \cdot n_i = \sum_{i=1}^k |x_i^* - \bar{x}| \cdot h_i, \quad x_i^* = \text{Klassenmitte der Klasse } i$$

### Varianz

- ungruppiertes Datenmaterial

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

- gruppiertes Datenmaterial

$$s^2 = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x})^2 \cdot n_i = \frac{1}{n} \sum_{i=1}^k (x_i^*)^2 \cdot n_i - \bar{x}^2 \quad x_i^* = \text{Klassenmitte der Klasse } i$$

### Standardabweichung

$$s = \sqrt{s^2}$$

### Variationskoeffizient

$$V = \frac{s}{\bar{x}} \cdot 100(\%)$$

### Konzept der Momente

Durchschnittliche potenzierte Abweichungen der Merkmalswerte um einen Bezugspunkt a:

Bezugspunkt Null (a = 0)

Momente um Null

Bezugspunkt arithmetisches Mittel (a =  $\bar{x}$ )

Momente um das arithmetische Mittel

- ungruppiertes Datenmaterial

$$m_r^a = \frac{1}{n} \sum_{i=1}^n (x_i - a)^r$$

- gruppiertes Datenmaterial

$$m_r^a = \frac{1}{n} \sum_{i=1}^k (x_i^* - a)^r \cdot n_i, \quad x_i^* = \text{Klassenmitte der Klasse } i$$

### Standardisierte Schiefe

Momente 3. Ordnung (r=3) ergeben die Schiefe. Die Schiefe (skewness) ist ein Asymmetriemaß

$$sm_3 = \frac{m_3}{s^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[ \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right]^3}$$

### Exzeß (Kurtosis, Wölbung)

Momente 4. Ordnung (r=4) ergeben die Wölbung

$$sm_4 = \frac{m_4}{s^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[ \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right]^4}$$

### Konzentration einer Verteilung

Die Merkmale werden für Konzentrationsanalysen grundsätzlich nach ihrer Größe geordnet

Verteilungsfunktion:

Abszissenwerte der Lorenzkurve

$$F(x_j) = \sum_{i=1}^j \frac{n_i}{n}$$

kumulierte relative Merkmalssumme:

Ordinatenwerte der Lorenzkurve

$$MS(x_j) = \frac{\sum_{i=1}^j x_i^* \cdot n_i}{n\bar{x}}$$

Gini Koeffizient:

Gruppiert:

$$G = \left[ \sum_{i=1}^k \left\{ \left[ F(x_{i-1}) + F(x_i) \right] \cdot \frac{n_i \cdot x_i^*}{n \cdot \bar{x}} \right\} \right] - 1$$

Ungruppiert ( $x_i$  geordnet!)

$$G = \frac{2 \cdot \sum_{i=1}^n i \cdot x_i - (n+1) \cdot \sum_{i=1}^n x_i}{n \cdot \sum_{i=1}^n x_i}$$

## IV Zweidimensionale Häufigkeitsverteilung

### Darstellung

Ausprägung des 1. Merkmals (x)  $x_i$   $i = 1, \dots, k$

Ausprägung des 2. Merkmals (y)  $y_j$   $j = 1, \dots, m$

**absolute Häufigkeit des Merkmalspaares ( $x_i, y_j$ )**

$$n_{ij} = n(x_i, y_j)$$

**Anzahl der Beobachtungen**

$$n = \sum_{i=1}^k \sum_{j=1}^m n_{ij}$$

**relative Häufigkeit des Merkmalspaares ( $x_i, y_j$ )**

$$h(x_i, y_j) = \frac{n_{ij}}{n}$$

**Randverteilung des 1. Merkmals (x) (Zeilensumme)**

marginale absolute Häufigkeiten  $n_{i\cdot} = \sum_{j=1}^m n_{ij}$

marginale relative Häufigkeiten  $h_{i\cdot} = \frac{n_{i\cdot}}{n}$

**Randverteilung des 2. Merkmals (y) (Spaltensummen)**

marginale absolute Häufigkeiten  $n_{\cdot j} = \sum_{i=1}^k n_{ij}$

marginale relative Häufigkeiten  $h_{\cdot j} = \frac{n_{\cdot j}}{n}$

**Häufigkeitsverteilung von x bei gegebenem y (bedingte Verteilung)**

$$h(x_i | y_j) = \frac{n_{ij}}{n_{\cdot j}}$$

**Häufigkeitsverteilung von y bei gegebenem x (bedingte Verteilung)**

$$h(y_j | x_i) = \frac{n_{ij}}{n_{i.}}$$

**Korrelationsrechnung****Häufigkeit bei Unabhängigkeit**

$$\tilde{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

**Quadratische Kontingenz**

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}$$

**Kontingenzkoeffizient**

$$K^* = \sqrt{\frac{\chi^2}{n + \chi^2} \cdot \frac{M}{M - 1}} \quad M = \min(k, m)$$

**Rangkorrelationskoeffizient nach Spearman**

$R_i$  Rangnummer des 1. Merkmals (ordinalskaliert)

$R'_i$  Rangnummer des 2. Merkmals (ordinalskaliert)

$$r_{sp} = 1 - \frac{6 \cdot \sum_{i=1}^n (R_i - R'_i)^2}{(n-1)n(n+1)}, \quad (-1 \leq r_{sp} \leq +1)$$

**Kovarianz**

- ungruppiertes Datenmaterial

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y}$$

- gruppiertes Datenmaterial

$$s_{xy} = \frac{1}{n} \sum_{i=1}^k (x_i^* - \bar{x}) (y_i^* - \bar{y}) \cdot n_i = \frac{1}{n} \sum_{i=1}^k x_i^* y_i^* \cdot n_i - \bar{x} \bar{y}$$

**Bravais-Pearson-Korrelationskoeffizient**

$$r = \frac{s_{xy}}{s_x s_y}, \quad (-1 \leq r \leq +1)$$

## Regressionsrechnung

beobachtete Werte	$y_i, x_{1i}, x_{2i}, \dots$
unterstellte funktionale Abhängigkeit	$y = f(x_1, x_2, \dots)$
Regressionskoeffizienten	$\beta_0, \beta_1, \beta_2, \dots$
Schätzwerte für $\beta_0, \beta_1, \beta_2, \dots$	$b_0, b_1, b_2, \dots$

### Modell der linearen Regression (stochastischer Ansatz)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K + \varepsilon$$

### Deterministischer Ansatz

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_K x_K$$

### Methode der kleinsten Quadrate (Ordinary Least Squares -OLS)

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \dots)^2 = \min!$$

$$\Rightarrow \frac{\partial Q}{\partial b_0} = 0, \quad \frac{\partial Q}{\partial b_1} = 0, \quad \frac{\partial Q}{\partial b_2} = 0, \dots$$

### Schätzwerte bei einfacher linearer Regression

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x}$$

### OLS-Schätzer bei multipler Regression

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

### Bestimmtheitsmaß bei einfacher linearer Regression

$$B = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = r^2$$

## V Indexzahlen

Preis des Gutes i zur Basiszeit	$p_0^i$
Preis des Gutes i zur Berichtszeit	$p_t^i$
Menge des Gutes i zur Basiszeit	$q_0^i$
Menge des Gutes i zur Berichtszeit	$q_t^i$

### Preisindex

- nach Laspeyres

$$I_{0,t}^L(p) = \frac{\sum_{i=1}^n p_t^i \cdot q_0^i}{\sum_{i=1}^n p_0^i \cdot q_0^i} \cdot 100$$

- nach Paasche

$$I_{0,t}^P(p) = \frac{\sum_{i=1}^n p_t^i \cdot q_t^i}{\sum_{i=1}^n p_0^i \cdot q_t^i} \cdot 100$$

### Mengenindex

- nach Laspeyres

$$I_{0,t}^L(q) = \frac{\sum_{i=1}^n p_0^i \cdot q_t^i}{\sum_{i=1}^n p_0^i \cdot q_0^i} \cdot 100$$

- nach Paasche

$$I_{0,t}^P(q) = \frac{\sum_{i=1}^n p_t^i \cdot q_t^i}{\sum_{i=1}^n p_t^i \cdot q_0^i} \cdot 100$$

### Umsatzindex (Wertindex)

$$I_{0,t}(p, q) = \frac{\sum_{i=1}^n p_t^i \cdot q_t^i}{\sum_{i=1}^n p_0^i \cdot q_0^i} \cdot 100$$

### Umbasierung von Basisjahr 1 auf Basisjahr 2

Basisjahr 1      01

Basisjahr 2      02

$$I_{02,t} = \frac{I_{01,t}}{I_{01,02}} \cdot 100$$

## VI Zeitreihenanalyse

Zeitreihenwerte                       $y_i$

Zeitpunkte                               $t_i$

### Trendgerade

$$y = \beta_0 + \beta_1 t + \varepsilon$$

### Ermittlung der Trendgeraden nach der Methode der kleinsten Quadrate

$$\hat{y} = b_0 + b_1 t$$

$$b_0 = \bar{y} - b_1 \bar{t}$$

$$b_1 = \frac{\frac{1}{n} \sum_{i=1}^n t_i y_i - \bar{t} \bar{y}}{\frac{1}{n} \sum_{i=1}^n t_i^2 - \bar{t}^2}$$

### transformierte Zeitwerte

bei ungeradem n                       $t'_i = t_i - \bar{t}$

bei geradem n                          $t'_i = 2(t_i - \bar{t})$

wenn  $\sum_{i=1}^n t'_i = 0$  gilt, folgt die Vereinfachung für die Ermittlung der Trendgeraden:

$$b'_0 = \bar{y}$$

$$b'_1 = \frac{\sum_{i=1}^n y_i t'_i}{\sum_{i=1}^n t'^2_i}$$

### Methode der gleitenden Durchschnitte

geschätzter Trendwert zum Zeitpunkt t                       $y'_t$

Anzahl der Zeitreihenwerte vor und nach dem Zeitpunkt t, die in die Berechnung von  $y'_t$  einbezogen werden.                      k

- ungerade Ordnung

$$\hat{y}'_t = \frac{1}{2k+1} \sum_{\tau=t-k}^{t+k} y_\tau, \quad t = k+1, k+2, \dots, n-k$$

- gerade Ordnung

$$\hat{y}'_t = \frac{1}{2k} \left( \frac{1}{2} y_{t-k} + \sum_{\tau=t-(k-1)}^{t+(k-1)} y_\tau + \frac{1}{2} y_{t+k} \right), \quad t = k+1, k+2, \dots, n-k$$

### Exponentielle Glättung

Glättungsfaktor  $a$ ;  $0 \leq a \leq 1$

- 1. Ordnung

$$\hat{y}'_t = a y_t + (1-a) \hat{y}'_{t-1}$$

- 2. Ordnung

$$\hat{y}''_t = a \hat{y}'_t + (1-a) \hat{y}''_{t-1}$$

### Prognose mit exponentieller Glättung

- 1. Ordnung

$$\hat{y}_{t+1} = \hat{y}'_t$$

- 2. Ordnung

$$\hat{y}_{t+r} = \bar{\alpha}_t + \bar{\beta}_t r \quad \text{mit} \quad \bar{t} = \frac{1-a}{a}, \quad \bar{\beta}_t = \frac{\hat{y}'_t - \hat{y}''_t}{\bar{t}}, \quad \bar{\alpha}_t = 2\hat{y}'_t - \hat{y}''_t$$

### Saisonbereinigung bei konstanter Saisonfigur

um die glatte Komponente bereinigte Zeitreihe  $y_{i,j} - \hat{y}'_{i,j}$   $i = \text{Jahr}, j = \text{Monat}$

$$\text{Saisonindexziffer} \quad \tilde{S}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} (y_{i,j} - \hat{y}'_{i,j})$$

$$\text{normierte Saisonindexziffer} \quad \hat{S}_j = \tilde{S}_j - \frac{1}{\lambda} \sum_{i=1}^{\lambda} \tilde{S}_j \quad \text{mit}$$

$\lambda = \text{Ordnung der gleitenden Durchschnitte}$

$$\text{saisonbereinigte Zeitreihe} \quad y_{i,j} - \hat{S}_j$$

### Saisonbereinigung bei variabler Zeitreihe

um die glatte Komponente bereinigte Zeitreihe  $\frac{y_{i,j}}{\hat{y}'_{i,j}}$

$$\text{Saisonindexziffer} \quad \tilde{I}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} \frac{y_{i,j}}{\hat{y}'_{i,j}}$$

$$\text{normierte Saisonindexziffern} \quad \hat{I}_j = \tilde{I}_j \cdot \frac{\lambda}{\sum_{j=1}^{\lambda} \tilde{I}_j}$$

$$\text{saisonbereinigte Zeitreihe} \quad \frac{y_{i,j}}{\hat{I}_j}$$

# Literatur

## A EINIGE STANDARDWERKE

- Anderson, David, R., Sweeney, Dennis, J., Williams, Thomas, A., Freeman, Jim und Essie Shoemith (2007), *Statistics for Business and Economics*, Thomson Publisher, London (mit CD)
- Anderson, O., Popp, W., Schaffranek, M., Stenger, H. und K. Szameitat (1988), *Grundlagen der Statistik*, Springer-Verlag, 2. Auflage, Berlin
- Bamberg, G. und F. Baur (2002), *Statistik*, R. Oldenbourg Verlag, 12. Auflage, München
- Bleymüller, J., Gehlert, G. und H. Gülicher (2004), *Statistik für Wirtschaftswissenschaftler*, 14. Auflage, Vahlen, München
- Blossfeld, H.-P., Hamerle, A. und K. U. Mayer (1986), *Ereignisanalyse: Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften*, Campus-Verlag, Frankfurt/New York
- Bortz, Jürgen (2004, 2010), *Statistik für Human- und Sozialwissenschaftler*, Springer-Verlag, Berlin
- Buttler G. und N. Fickel (2002), *Statistik mit Stichproben*, Rowohlt Taschenbuch Verlag, Reinbeck bei Hamburg
- Fahrmeier, L., Künstler, R., Pigeot, I. und G. Tutz (2004, 2009), *Statistik - Der Weg zur Datenanalyse*, 5./7. verbesserte Auflage, Springer-Verlag, Berlin
- Ferschl, F. (1985), *Deskriptive Statistik*, Physica-Verlag, 3., korrigierte Auflage, Würzburg
- Grohmann, H. (1986a), *Statistik - Allgemeine Methodenlehre I (ohne Wahrscheinlichkeitsrechnung)*, 2. Auflage, dipa-Verlag, Frankfurt a.M.
- Hansen, G. (1985), *Methodenlehre der Statistik*, 3. Auflage, München, Vahlen
- Hartung, J., Elpelt, B. und K.-H. Klösener (2005), *Statistik: Lehr- und Handbuch der angewandten Statistik*, 14., unwesentlich veränderte Auflage, R. Oldenbourg Verlag, München
- Hochstädter, D. (1996), *Statistische Methodenlehre*, 8., überarbeitete Auflage, Verlag Harri Deutsch, Frankfurt a.M.
- Hujer, R. (2001), *Statistik - Manuskript zur Vorlesung*, Frankfurt a.M.
- Hujer, R. und R. Cremer (1998), *Methoden der empirischen Wirtschaftsforschung*, 2. Auflage, Vahlen, München
- Kellerer, H. (1976), *Statistik im modernen Wirtschafts- und Sozialleben*, 14. Auflage, Rowohlt, Reinbek bei Hamburg
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) (2001), *Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*, Baden-Baden.
- Kreyszig, E. (1989), *Statistische Methoden und ihre Anwendungen*, 7. Auflage, Vandenhoeck & Ruprecht, Göttingen

- Krug, W. und M. Nourney (2001), Wirtschafts- und Sozialstatistik: Gewinnung von Daten, R. Oldenbourg Verlag, 6. Auflage, München
- Kunz, D. (1987), Praktische Wirtschaftsstatistik, Kohlhammer, Stuttgart.
- Lippe von der, P. (1996), Wirtschaftsstatistik, 5., völlig neubearbeitete und erweiterte Auflage, Lucius & Lucius, Stuttgart
- Litz, H.P. (2003), Statistische Methoden in den Wirtschafts- und Sozialwissenschaften, 3., vollständig überarbeitete und erweiterte Auflage, R. Oldenbourg Verlag, München, Wien
- Merz, J. (2011), Statistik I - Deskription, Skriptum zur Vorlesung, 10. erweiterte Auflage, Lüneburg
- Mittag, H.J. und D. Stemann, Statistik – Beschreibende Statistik und explorative Datenanalyse, 5., verbesserte und erweiterte Auflage, Fachbuchverlag Leipzig im Carl-Hanser-Verlag, Leipzig
- Moore, D.S. (1997), Statistics – Concepts and Controversies, 5. Auflage, W.H. Freeman and Company, New York
- Neubauer, W., Bellgardt, E. und A. Behr (2002), Statistische Methoden, 2. Auflage, Vahlen, München
- Pfanzagl, J. (1983), Allgemeine Methodenlehre der Statistik I, 6., verbesserte Auflage, Walter de Gruyter, Berlin/New York
- Sachs, L. (2004), Angewandte Statistik - Anwendung statistischer Methoden, 11., überarbeitete und aktualisierte Auflage, Springer Verlag, Berlin
- Scharnbacher, K. (2004), Statistik im Betrieb - Lehrbuch mit praktischen Beispielen, 14., aktualisierte Auflage, Gabler Verlag, Wiesbaden
- Schira, J. (2009), Statistische Methoden der VWL und BWL Theorie und Praxis, 3. aktualisierte Auflage, Pearson Studium, München
- Schlittgen, R. (2003), Einführung in die Statistik, 10., durchgesehene Auflage, R. Oldenbourg Verlag, München, Wien
- Schwarze, J. (2005, 2009), Grundlagen der Statistik I - Beschreibende Verfahren, 10./11. Auflage, Verlag Neue Wirtschafts-Briefe, Herne/Berlin
- Wetzel, W. (1971, 1973), Statistische Grundausbildung für Wirtschaftswissenschaftler, Teil I und Teil II, Walter de Gruyter, Berlin
- Yamane, T. (1981), Statistik - Ein einführendes Lehrbuch, Teil I und Teil II, Fischer Taschenbuch Verlag, Frankfurt
- Zöfel, P. (1992), Statistik in der Praxis, 3., überarbeitete und ergänzte Auflage, UTB 1293, Stuttgart

## **B BÜCHER MIT ÜBUNGSAUFGABEN**

- Bamberg, G. und F. Baur (2000, 2007), Statistik Arbeitsbuch: Übungsaufgaben, Fallstudien, Lösungen, 6./8. Auflage, R. Oldenbourg Verlag, München
- Bihn, W.R. und K.A. Schäffer (1986), Übungsaufgaben zur Grundausbildung in Statistik für Wirtschaftswissenschaftler, J.C. Witsch Nachf., Köln

- Fahrmeier, L., Künstler, R., Pigeot, I., Tutz, G., Caputo, A. und S. Lang (2003), Arbeitsbuch Statistik, 3., überarbeitete und erweiterte Auflage, Springer-Verlag, Berlin
- Hartung, J. und B. Heine (1999), Statistik Übungen: Deskriptive Statistik, 6. Auflage, R. Oldenbourg Verlag, München
- Hochstädter, D. (1993), Aufgaben mit Lösungen zur statistischen Methodenlehre, 2. Auflage, Verlag Harri Deutsch, Frankfurt
- Lippe von der, P. (2006), Formeln, Aufgaben, Klausurtraining in Statistik, R. Oldenbourg Verlag, 7. Auflage, München/Wien
- Merz, J. (2011), Statistik I - Deskription, Übungs- und Klausuraufgaben mit Lösungen, 10. Auflage, Lüneburg
- Spiegel, M.R. (2003), Statistik, 1. Auflage, Mc Graw Hill, Düsseldorf

## **C FORMEL- UND TABELLENWERKE**

- Bihn, E.R. und K.A. Schäffer (1987), Formeln und Tabellen zur Grundausbildung in Statistik für Wirtschaftswissenschaftler, J.C. Witsch Nachf., Köln
- Bleymüller, J. und G. Gehlert (1999), Statistische Formeln, Tabellen und Programme, 9. Auflage, Verlag Franz Vahlen, München
- Bohley, P. (1998), Formeln, Rechenregeln und Tabellen zur Statistik, R. Oldenbourg Verlag, 7. Auflage, München
- Rinne, H. (1997), Statistische Formelsammlung, 2. Auflage, Verlag Harri Deutsch, Frankfurt
- Vogel, F. (2000), Beschreibende und schließende Statistik - Formeln, Definitionen, Erläuterungen, Stichwörter und Tabellen, R. Oldenbourg Verlag, 12. Auflage, München

## **D WEITERFÜHRENDE LITERATUR**

- Allgemeines Statistisches Archiv (1992), Band 76
- Arminger, G. und F. Müller (1990), Lineare Modelle zur Analyse von Paneldaten, Opladen
- Backhaus, K., Erichson, B., Pinke, W., Schuchard-Fischer, Chr. und R. Weiber (2006), Multivariate Analysemethoden, 11., überarbeitete Auflage, Springer Verlag, Berlin/Tokyo
- Bates, B.C., Z.W. Kundzewicz, S. Wu, J.P. Palutikof, Eds. (2008), Climate Change and Water. Technical Paper of Intergovernmental Panel on Climate Change, IPCC Secretariat, 210 pp., Geneva
- Berntsen, R. (1991), Dynamik in der Einkommensverteilung. Eine empirische Längsschnittuntersuchung der Strukturen der Einkommensverteilung privater Haushalte in der Bundesrepublik Deutschland, Dissertation, Frankfurt
- Diewald, M. (1984), Das 'SPES-Indikatoren-Tableau' 1976 - Fortschreibung bis zum Jahr 1982, Sfb 3-Arbeitspapier Nr. 150, Frankfurt/Mannheim
- Esenwein-Rothe (1978), Modelle für eine Bevölkerungsprojektion und die Grenzen der Aussagekraft, in: Jahrbuch für Nationalökonomie und Statistik, Band 193, Heft 1

- Esser, H., Grohmann, H., Müller, W. und K.A. Schäffer (1989), Mikrozensus im Wandel - Untersuchungen und Empfehlungen zur inhaltlichen und methodischen Gestaltung, Forum der Bundesstatistik, Bd. 11, Metzler-Poeschel Verlag, Stuttgart
- Galler, H.P. and G. Wagner (1986), The Microsimulation Model of the Sfb 3 for the Analysis of Economic and Social Policies, in: Orcutt, G.H., Merz, J. and H. Quinke (eds.), Microanalytic Simulation Models to Support Social and Financial Policy, S. 227-247, North Holland, Amsterdam
- Galler, H.P. und N. Ott (1994), Das dynamische Mikrosimulationsmodell des Sonderforschungsbereichs 3, in: Mikroanalytische Grundlagen der Gesellschaftspolitik, Deutsche Forschungsgemeinschaft, Bd. 2, Akademie Verlag, Berlin
- Glatzer, W. und W. Zapf (Hrsg.) (1984), Lebensqualität in der Bundesrepublik, Frankfurt/New York
- Goldberger, A.S. (1991), A Course in Econometrics, Harvard University Press, London
- Greene, W. (1991), Econometric Analysis, Macmillan Publishing Company, New York
- Greene, W. (1992), ET - The Econometrics Toolkit, Version 3.0, Econometric Software, Inc., New York
- Grohmann, H. (1986b), Bevölkerungs- und Wirtschaftsstatistik, dipa-Verlag, 2. Auflage, Frankfurt a.M.
- Habich, R., und H.-H. Noll unter Mitarbeit von W. Zapf (1993), Soziale Indikatoren und Sozialberichterstattung - Internationale Erfahrungen und gegenwärtiger Forschungsstand, Berlin/Mannheim
- Hamer, G. und C. Stahmer (1992), Integrierte Volkswirtschaftliche- und Umweltgesamtrechnung (I): Konzeption, in: ZfU, Heft 1, S. 85-117; Integrierte Volkswirtschaftliche- und Umweltgesamtrechnung (II): (Zahlen-)Beispiel und Realisierungsmöglichkeiten, in: ZfU, Heft 2, S. 237-256
- Hansen, G. (1993), Quantitative Wirtschaftsforschung, Franz Vahlen, München
- Hsiao, C. (1986), Analysis of Panel Data, Cambridge (Mass.)
- Huff, D. (1978), How to Lie with Statistics, Penguin Books, Harmondsworth (UK)
- Johnson, J.D. (1992), Applied Multivariate Data Analysis – Volume II: Categorical and Multivariate Methods, Springer Verlag, New York
- Krämer, W. (1991), So lügt man mit Statistik, Campus-Verlag, Frankfurt/New York
- Krupp, H.-J. und W. Zapf (1977), Sozialpolitik und Sozialberichterstattung, Frankfurt/New York
- Leipert, C. (1975), Unzulänglichkeiten des Sozialprodukts in seiner Eigenschaft als Wohlstandsmaß, Tübingen
- Leipert, C. (1989), Die heimlichen Kosten des Fortschritts, Frankfurt
- Maddala, G.S., Rao, C.R. und H.D. Vinod (Hrsg.) (1993), Econometrics, 11. Auflage, North-Holland, New York
- Malinvaud, E. (1980), Statistical Methods in Econometrics, 3. Auflage, American Elsevier, New York
- Mátyás, L. und P. Sevestre (1992), The Econometrics of Panel Data, Handbook of Theory and Applications, Kluwer academic publishers, Dordrecht

- Merz (2002), Freie Berufe im Wandel der Märkte, FFB-Schriftenreihe, Band 13, Nomos Verlag, Baden-Baden
- Merz, J. (1980a), Die Ausgaben privater Haushalte - Ein mikroökonomisches Modell für die Bundesrepublik Deutschland, Frankfurt/New York
- Merz, J. (1980b), Prognosegüte und Spektraleigenschaften ökonomischer Modelle, in: Stöppler, S. (Hrsg.), Dynamische ökonomische Systeme - Analyse und Steuerung, 2. Auflage, Gabler-Verlag, Wiesbaden, S. 31-66
- Merz, J. (1987), Mathematik II für Wirtschaftswissenschaftler, Skriptum zur Vorlesung, Johann Wolfgang Goethe-Universität Frankfurt, Frankfurt a.M.
- Merz, J. (1991a), Markt- und nichtmarktmäßige Aktivitäten privater Haushalte - Theoretischer Ansatz, repräsentative Mikrodaten, mikroökonomische Analyse und Mikrosimulation wirtschafts- und sozialpolitischer Maßnahmen für die Bundesrepublik Deutschland, Frankfurt a.M.
- Merz, J. (1991b), Microsimulation - A Survey of Principles, Developments and Applications, in: International Journal of Forecasting, 7, S. 77-104
- Merz, J. (1994a), Statisches Mikrosimulationsmodell: Mainframe und PC-Version, in: Hauser, R., Ott, N. und G. Wagner (eds.), Deutsche Forschungsgemeinschaft: Mikroanalytische Grundlagen der Gesellschaftspolitik - Erhebungsverfahren, Analysemethoden und Mikrosimulation, Akademie Verlag, Berlin
- Merz, J. (1994b), Microdata Adjustment by the Minimum Information Loss Principle, Forschungsinstitut Freie Berufe der Universität Lüneburg, FFB-Diskussionspapier Nr. 10, Lüneburg
- Merz, J. (2008), Statistik II - Wahrscheinlichkeitsrechnung und induktive Statistik, Skriptum zur Vorlesung, 7. verbesserte Auflage, Lüneburg
- Merz, J. und H. Stolze (2006), Adjust for Windows Version 1.1 - A Software Package to Achieve Representative Microdata by the Minimum Information Loss Principle - Manual, FFB-Dokumentation Nr. 13, Forschungsinstitut Freie Berufe, Fakultät II Wirtschafts-, Verhaltens- und Rechtswissenschaften, Universität Lüneburg, Lüneburg
- Merz, J. und J. Faik (1992), Equivalence Scales Based on Revealed Preference Consumption Expenditure Microdata - The Case of West Germany, Forschungsinstitut Freie Berufe der Universität Lüneburg, FFB-Diskussionspapier Nr. 3, Lüneburg
- Merz, J., Helberger, C. und H. Schneider (1985), Nebenerwerbstätigkeitsumfrage 1984, Dokumentation, bearbeitet von Klaus Wolff, Frankfurt
- Merz, J., Stolze, H. und M. Zwick (2002), Professions, entrepreneurs, employees and the new German tax (cut) reform 2000 - A MICSIM microsimulation analysis of distributional impacts, Forschungsinstitut Freie Berufe der Universität Lüneburg, FFB-Diskussionspapier Nr. 34, Lüneburg
- Mueller, U. (1993), Bevölkerungsstatistik und Bevölkerungsdynamik, Walter de Gruyter, Berlin/New York
- Noll, H.-H. (1990), Sozialindikatorenforschung in der Bundesrepublik - Konzepte, Forschungsansätze und Perspektiven, in: Timmermann, H. (Hrsg.), Lebenslagen, Sozialindikatorenforschung in beiden Teilen Deutschlands, S. 69-87, Saarbrücken
- Noll, H.-H. (1993), Lebensbedingungen in der Europäischen Gemeinschaft gleichen sich nur langsam an - ökonomische und soziale Indikatoren im EG-Vergleich, in: ZUMA, Informationsdienst Soziale Indikatoren (ISI), Heft 4, S. 11-15

- Noll, H.-H. (Hrsg.) (1997), Sozialberichterstattung in Deutschland - Konzepte, Methoden und Ergebnisse für Lebensbereiche und Bevölkerungsgruppen, Juventa, Weinheim und München
- Nullau, B. u.a. (1969), Das Berliner Verfahren, DIW- Beiträge zur Strukturforchung, Berlin
- Orcutt, G., Merz, J. and H. Quinke (eds.) (1986), Microanalytic Simulation Models to Support Social and Financial Policy, North Holland, Amsterdam
- Rapin, H. (Hrsg.) (1990), Der private Haushalt - Daten und Fakten, Stiftung 'Der Private Haushalt', Campus, Frankfurt a.M.
- Rinne, H. (1994), Wirtschafts- und Bevölkerungsstatistik - Erläuterungen, Erhebungen, Ergebnisse, R. Oldenbourg Verlag, München/Wien
- Rockwell, R. (1986/87), Prospects for Social Reporting in the United States: A Receding Horizon, in: The Tocqueville Review, Vol. 8, S. 251-262
- Schlittgen, R. und B.H.J. Streitberg (1987), Zeitreihenanalyse, 2. Auflage, R. Oldenbourg Verlag, München/Wien
- Schwarze, J. (2006), Grundlagen der Statistik II – Wahrscheinlichkeitsrechnung und induktive Statistik, Verlag Neue Wirtschafts-Briefe, 8. Auflage, Herne/Berlin
- Sheldon, E. B. and R. Park (1975), Social Indicators, in: Science, American Association for the Advancement of Science, Vol. 188, S. 693-699
- Sheldon, E. B. and W.E. Moore (eds.) (1968), Indicators of Social Change, Concepts and Measurement, New York
- Spanos, A. (1986), Statistical Foundation of Econometric Modelling, Cambridge
- Stahmer, C. (1992), Integrierte Volkswirtschaftliche- und Umweltgesamtrechnung, in: Wirtschaft und Statistik, Heft 9, S. 577-593
- Statistisches Bundesamt (1980), Fachserie 14, Reihe 7.3, Finanzen und Steuern, Lohnsteuern, Wiesbaden
- Statistisches Bundesamt (1990), Fachserie 1, Reihe 4.1.1, Bevölkerung und Erwerbstätigkeit, Wiesbaden
- Statistisches Bundesamt (Hrsg.) (1988), Das Arbeitsgebiet der Bundesstatistik, W. Kohlhammer, Stuttgart/Mainz
- Statistisches Bundesamt (Hrsg.) (1992). Datenreport 1992, in Zusammenarbeit mit dem Wissenschaftszentrum für Sozialforschung, Berlin und dem Zentrum für Umfragen, Methoden und Analysen, Mannheim, Wiesbaden
- Statistisches Bundesamt, Statistisches Jahrbuch, verschiedene Jahrgänge, Wiesbaden
- Statistisches Bundesamt, Zahlen, Fakten, Trends: Monatlicher Pressedienst, Wiesbaden
- Steger, A. (1980), Haushalte und Familien bis zum Jahre 2000 ,Campus, Frankfurt a.M./New York).
- Stobernack, M. (1989), Die Bedeutung der Arbeitslosenversicherung für Arbeitslosigkeit und Arbeitsangebot unter Einbeziehung eines empirischen Arbeitsangebotsvergleichs zwischen der Bundesrepublik und den USA, Berlin.
- United Nations Development Programme (UNDP) (1991), Human Development Report 1991, Oxford

- Vogel, J. (1990), Social Indicators - A Swedish Perspective, in: Journal of Public Policy, Vol. 9, S. 439-444
- Yang, M.C.K. and D. Robinson (1986), Understanding and Learning Statistics by Computer, World Scientific, Singapore
- Zapf, W. (1972), Zur Messung der Lebensqualität, in: Zeitschrift für Soziologie, 1. Jg., S. 353-367
- Zapf, W. (1977), Einleitung in das SPES-Indikatorensystem, in: Zapf, W. (Hrsg.), Lebensbedingungen in der Bundesrepublik. Sozialer Wandel und Wohlfahrtsentwicklung, S. 11-27, Frankfurt/New York
- Zapf, W. (1990), Einleitung, in: WZB-AG Sozialberichterstattung (Hrsg.), Sozialreport 1990, Dokumentation eines Workshops am Wissenschaftszentrum Berlin für Sozialforschung, Arbeitspapier P90-102, Berlin

## **E SONSTIGE LITERATUR**

- Berger-Schmitt, R. (2002), Unterschiede in den Lebensbedingungen in der Europäischen Union kaum verringert, in: Informationsdienst Soziale Indikatoren (ISI), Ausgabe 27, Januar 2002, S. 2
- Club of Rome (1991), Der Blick in die Zukunft, in: natur, Heft 9, S. 31-32
- Der siebte Tag (30.11.2002), Wochenendbeilage zur Hannoverschen Allgemeinen Zeitung, November 2002, Madsack-Verlag, Hannover
- Die ZEIT (02.09.1994), Nr. 36, ZEITVERLAG, Hamburg
- Gonick, L. and W. Smith (1993), The Cartoon Guide to Statistics, HarperCollins Publishers, New York
- Kuhn, T. (1970), Structure of Scientific Revolutions, 2. Auflage, Chicago
- Meadows, D., Randers, D. und J. Randers (1992), Die neuen Grenzen des Wachstums - Die Lage der Menschheit: Bedrohung und Zukunftschancen, Stuttgart
- Merz, J., Rauberger, T.K. und A. Rönnau (1994), Freie Berufe in Rheinland-Pfalz und in der Bundesrepublik Deutschland: Struktur, Entwicklung und wirtschaftliche Bedeutung, Schriften des Forschungsinstituts Freie Berufe der Universität Lüneburg Nr. 7, Lüneburg
- Myers, N. (Hrsg.) (1985), gaia - der öko-Atlas unserer Erde, Fischer Verlag, Frankfurt
- Schwarze, J. (1977), Bibliographie zur Statistik in der Weiterbildung, 2. Auflage, Pädagogische Arbeitsstelle des Deutschen Volkshochschulverbandes, Holzhausenstr. 21, 6000 Frankfurt, Frankfurt/Bonn
- Sozio-ökonomisches Panel (Welle I (1) 1984, Welle I (9), 1992), des Sonderforschungsbereichs 3, Frankfurt/Mannheim und des DIW, Berlin
- Statistisches Bundesamt (1986), Volkszählung '87: Zehn Minuten, die allen helfen - Materialien, Abschnitt 2.2, Wiesbaden
- Statistisches Bundesamt (2003), Energieverbrauch und Luftemissionen des Sektors Verkehr, Band 12 der Schriftenreihe zu den Umweltökonomischen Gesamtrechnungen - Kurzfassung

- Statistisches Bundesamt (2005), Verbraucherpreisindex und Index der Einzelhandelspreise - Jahresdurchschnitte ab 1948, Wiesbaden
- Statistisches Bundesamt (Hrsg.) (1988), Das Arbeitsgebiet der Bundesstatistik 1988, Mainz, Kohlhammer
- United Nations (2005), Population Challenges and Development Goals, Department of Economic and Social Affairs, Population Division, New York
- United Nations Development Programm (UNDP) (Hrsg.), Human Development Report 1991, New York, Oxford, 1991, Oxford University Press
- United Nations Development Programm (UNDP) (Hrsg.), Human Development Report 2003, New York, Oxford, 2003, Oxford University Press
- Wolffs, M. (2002), Bevölkerung zwischen Dynamik und Stillstand - Demographische Entwicklungen im Längsschnitt, Sankt Augustin, Arbeitspapier der Konrad-Adenauer-Stiftung e.V.
- ZEIT-Punkte (1994), Weltbevölkerung - Wird der Mensch zur Plage?, Nr. 4/1994, ZEIT-VERLAG, Hamburg