

Peter von der Lippe

Statistik und Manipulation

Gebrauch und Missbrauch von Statistik



Manipulation ganz ohne Statistik

Vortrag bei der **Klausurtagung**

"Smart Data- Wie sieht ein intelligenter Umgang mit Zahlen und Daten aus?" der Geschäftsstelle **des Wissenschaftsrats** am **31. 10. 2013 in Bonn** (Haus der Geschichte)

Statistik = Mathematik +

Mathematik und Manipulation wäre kein Thema

Was kommt zur Mathematik hinzu und bietet damit die Möglichkeit der Manipulation?

Die Absicht, ein Bild der Wirklichkeit zu zeichnen

Aber:

1. Das Bild soll "**richtig**" sein und den Vorstellungen (Wünschen) entsprechen; es wird "**beurteilt**"
2. man kann es mit **Zahlen und** mit **Worten** zeichnen (Daten → Analyse → Interpretation) Statistik ist **nicht nur Zahlen**
3. Zahlen sprechen **nicht** für sich

Warum Manipulation?

- Wahrnehmung der Wirklichkeit ist tendenziell **kontrovers**
- Statistik soll Bild der Wirklichkeit "objektiv" machen (Streit schlichten)

Wie kann das gemacht werden?

damals völlig ungewohnt, als noch Worte, Bilder, Analogien (Gleichnisse) das Denken dominierten

Das Ideal, wie es anfänglich war

1. Fokus auf **Zahlen** (Messung → Aufklärung)

(Suche nach Regelmäßigkeiten, "soziale Physik", "Descartes Traum")

2. Sammeln von (**vielen**) "**Fakten**"; keine Einzelfälle

(impressionistisch, "anecdotal evidence")

und wie es mit der Zeit mehr und mehr wurde ⇒

Keiner traut sich noch ohne Berufung auf Statistiken (aus dem Bauch heraus) zu entscheiden, Messung ist eine Obsession geworden (⇒ "Computopia")

Manipulation nicht grundsätzlich verwerflich*

amtliche Statistik

Gericht

Parteilichkeit nicht akzeptabel
extremes Beispiel: DDR (SZS)

Agitation und Propaganda \Rightarrow wann ist ein Statistikproduzent "seriös"?

1. Statistik wäre uninteressant wenn sie nicht auch zur Manipulation genutzt werden könnte
2. Sie ist nicht schlecht weil sie manipuliert werden kann, sondern gut, weil ohne Statistik (mit Worten) Manipulation noch viel leichter wäre

Statistikanwender

Rechtsanwalt#

Parteilichkeit gewünscht

Bsp: • Sir Ronald A. Fisher
• Preiselastizität

Aber Rechtsanwälte machen keine Gesetze

Literatur

- Deborah Tannen "The Argument Culture"
- Robert P. Abelson "Statistics as Principled Argument"

[auf meiner Homepage](#)

Statistik für Schaumschläger

Was tun, wenn einem eine Statistik nicht passt? (WT)

* Ich bin mit dieser Auffassung in einer Diss. (Chr. Birnstiel, Bundeswehrhochschule München 2001) kritisiert worden

Regeln der Manipulation

allgemein

- andere beeinflussen, ohne dass diese das merken
- Mächtige → Unterworfenen
aber auch unter Gleichen



1. Geschenke
2. Schuldbewusstsein
3. Ängste und Gruppenzwang
(auch andere machen es: Unterschriftenlisten; wiss. Vorträge
→ S. für Schaumschläger)

mit Statistik

Datenbeschaffung

- Definitionen (Industrieroboter Beispiel DDR) → Messung
- "Repräsentativität", biased sample

Darstellung und Analyse

- Rechenricks (Wachstumsraten) und Trugschlüsse nahelegen
- verbal kommentieren
- Analysen erarbeiten + fremde Analysen kritisieren

Dieser Zweiteilung entspricht auch Aufbau des Vortrags; davor noch eine Kosten/Nutzen Analyse

Kosten-Nutzen Analyse von (manipulierter) Statistik

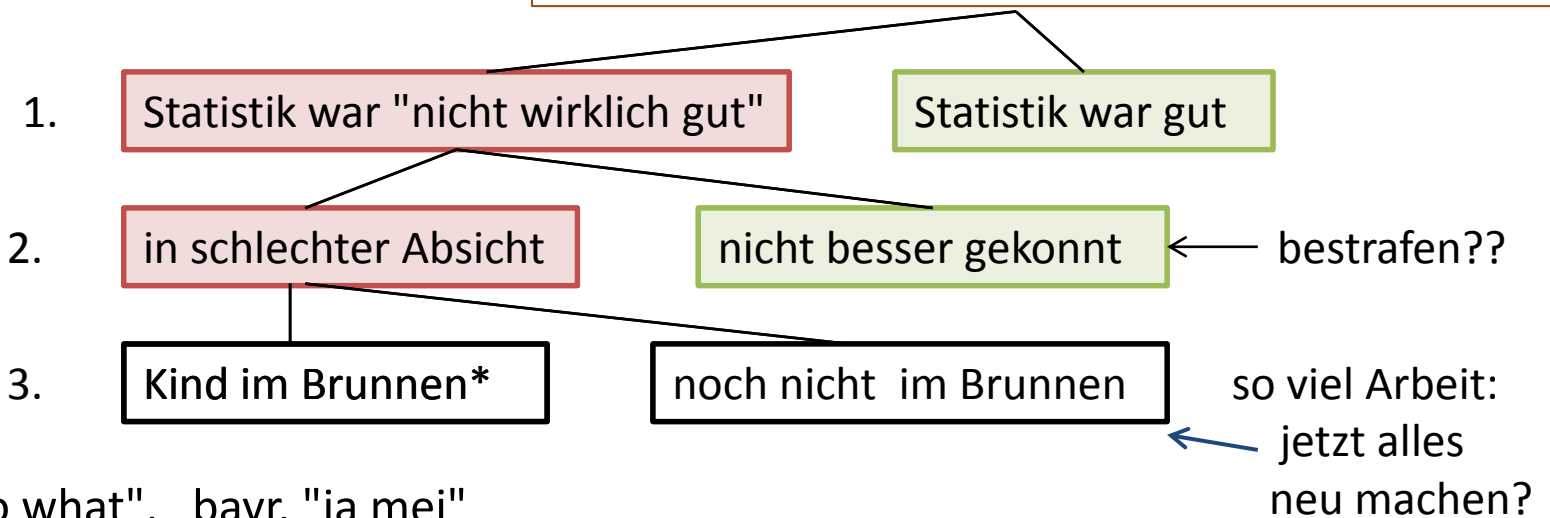
Nettonutzen weitgehend positiv (am Beispiel von wiss. Arbeiten)

Nutzen hoch weil

- heutzutage ein Muss
- Tabellen (T) + Gleichungen (G) = Kompetenz, Fleiß, Exaktheit
- T+G kann man genauso abschreiben wie Worte; wirkt aber besser

Kosten gering weil

- Informationsasymmetrie (schwer, Qualität statistischer Ausführungen zu beurteilen)
- schlechte Statistik weniger gefährlich als schlechtes (oder unterlassenes) Zitieren
- Sanktionen unwahrscheinlich, weil es Beweisprobleme auf den drei Ebenen gibt



* engl. "so what", bayr. "ja mei"

1.

- **Was erfragen** (was will ich darstellen?)
⇒ **Zahlen** (was besagen sie?)
- **Wen, wie und wie viele befragen**

2.

- **Daten auswerten** (Instrumentenkasten)
- **Auswertung kommunizieren
und kritisieren**

"Kategorien" (Kant, Erkenntnisvoraussetzungen) für Statistik

Datengewinnung *

Ausprägungen von **Merkmalen** werden an **Einheiten** einer **Gesamtheit** "erhoben"

3. **wie viele** befragen Stichprobe "repräsentativ"?

Warum Wahrscheinlichkeitsrechnung?

2. **wen** über **was** befragen?


1. diskret **zählen**/stetig **messen**

engl. hierfür zwei Wörter **discreet** (Verhalten), **discrete** (Statistik)

genauer **Skalentyp** (was definiert? welche Transformation zulässig?)

Zahlen \mathbb{N} oder \mathbb{R} haben unterschiedlichen Informationsgehalt. Durch Einführung einer Maßeinheit wird **messen auf zählen zurückgeführt**

zählen und rechnen

wie viele Erbsen hier?  (**zählen**)

und hier?  (**rechnen**)

* weniger allgemein akzeptierte Grundsätze bei Interpretation (Auswertung), d.h. Beschreibung + Induktion

modern talking

(überall nur Zahlen; die aber nicht selbsterklärend sind)

mit dem Arzt

42 0,92 141 56 4,3 2,35 8,8 14 41 86

mit dem Piloten

09 162₀ 172₀ 290 13 09 - 03 3000 FEW 040 1013

mit wem ?

51 163 313 346 17,3 11,35 5,8 3014 81 1099

1. Beim Rechnen erhält man immer Zahlen (also ein "Ergebnis")
2. aber die Zahlen sind nicht immer sinnvoll [Gesamtwicht von Brüssel \(Quetelet\)](#) und man kann es den Zahlen nicht ansehen, ob sie sinnvoll sind
3. es sind erst die Worte (42 = Harnstoff, 0,92 = Phosphat) , die den obigen Zahlen einen **Sinn geben**, bzw. nur FEW (Bewölkung) versteht man auf Anhieb
4. mit Worten (methodischen Einwänden) kann man auch einer Zahl **Sinn**, Bedeutung, Wichtigkeit etc. **nehmen**
5. auch wenn alle erklärende Worte nicht befriedigen hat man lieber irgendeine Zahl (Messung) als keine Zahl

dagegen [If you have nothing to say, don't say anything](#) [Abelson's 5th Law](#)

Promise and Pitfalls of "Big Data" und Geschichte hinter den Zahlen

The numbers have no way of speaking for them. **We** speak for them. We imbue them with meaning.

Big Data: Information is no longer a scarce commodity; we have more of it than we know what to do with it.

Nate Silver, *The Signal and the Noise: Why so Many Predictions Fail but Some Don't*, 2012

Statistik über Qualität der Lehre Rektorat Uni Essen (90er Jahre):

ein Student hat sagt, dass ein Professor des Fachbereichs seit 40 Jahren genau dieselbe Vorlesung hält.

Frage: wie ist alt der Student?

wie kann es jemand fertig bringen, sich 40 Jahre lang die gleiche Vorlesung anzuhören

Diskret (kategoriales Denken) → Zählen

Mann = Mann → 2 Männer

Nominalskala, Äquivalenzrelation = bzw. \neq definiert, noch nicht Ordnungsrelation $<$, $>$

Dr.A = Dr.B oder Dr.A \neq Dr.B ?

antikapitalistisches Pamphlet bei Par-
teihochschule; Fuß-
noten o.k.

aufwändige empi-
rische Studie, aber
nicht immer zitiert, Fuß-
noten schlecht

Grenze zwischen **zählen** und **erzählen**
(tale, tell ⇒ Zahl)

1976

als ich nach
Essen kam

Universität

Fachhochschule

aber mit der Zeit
hat sich das verändert

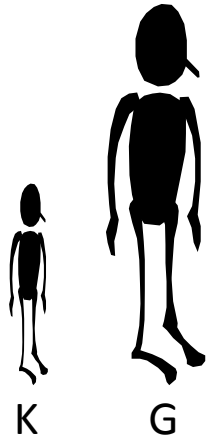
Universität

Fachhochschule

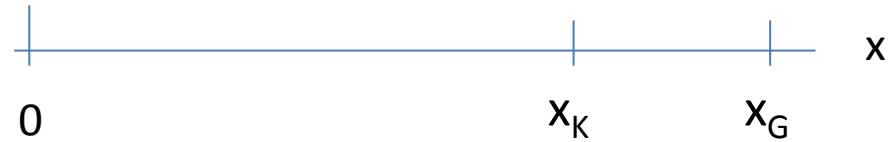
gilt noch $=/\neq$ oder
gar $>$ $<$ oder gilt
jetzt \approx ?

Messung: Dimensionen

einfach bei **einer**
Dimension



Körpergröße $x_K = 140$, $x_G = 180$



aber bei zwei
Dimensionen?



Ranking von Professoren A, B, C, D

	F -	F +
L -	A --	B -+
L +	C +-	D ++

klar ist $x_A < x_D$

aber wo die "hybriden" Fälle B und C?

L = Lehre

- schlecht, + gut

F = Forschung

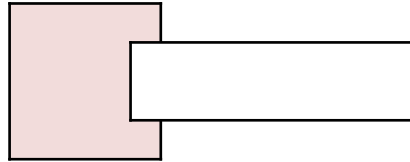
- schlecht, + gut

Bei $m = 3$ Dimensionen $2^3 = 8$ Kombinationen und 6 "hybride" Profs

Auf weitere Dimensionen und hybride Fälle hinweisen, wenn einem das Ranking nicht gefällt

Messung: aus zwei Dimensionen eine machen

Nur wenige Menschen kommen mit mehr als einer Dimension zurecht



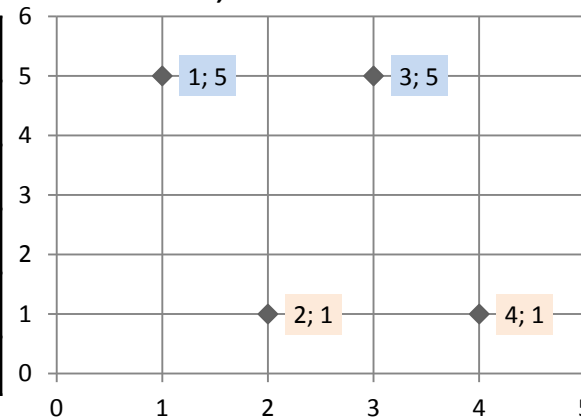
noch schwerer 3 Dimensionen (Tricks mit graphischen Darstellungen) und wie ist es mit 4 Dimensionen??

Welche Fläche (Häufigkeit) ist größer?

zwei Dimensionen zu einer machen Punktsummenverfahren

Vier Objekte mit Messungen in **zwei** Dimensionen, X und Y

	X	Y	$(X+Y)/2$
A	2	1	1,5
B	4	1	2,5
C	1	5	3
D	3	5	4
	A < B	C < D	A < B < C < D



Punktsummenverfahren sehr beliebt und (weil?) einfach

Ranking beliebt weil "Reduktion der Komplexität"

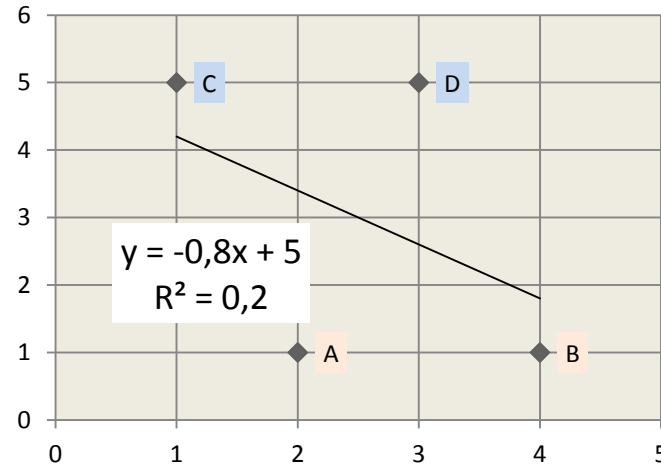
Man traut sich nicht / kann nicht/darf nicht aus dem Bauch heraus entscheiden

Gewichtung	A	B	C	D	Bemerkung
X:Y 2/3, 1/3	1,67	3,00	2,33	3,67	A < C < B < D
4/5, 1/5	1,8	3,4	1,8	3,4	A = C und B = D
5/6, 1/6	1,83	3,50	1,67	3,33	A > C < B > D

"Gewichtung" der Dimensionen

ohne Gewichtung (bzw. $\frac{1}{2}$ und $\frac{1}{2}$)

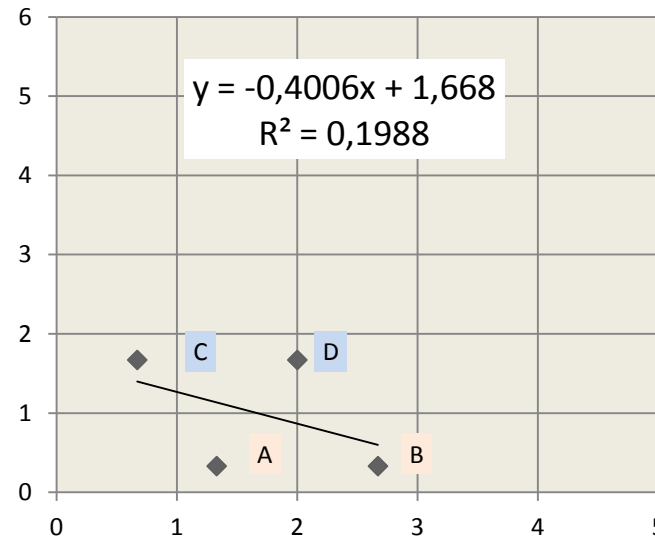
	X	Y	$(X+Y)/2$
A	2	1	1,5
B	4	1	2,5
C	1	5	3
D	3	5	4
	A < B	C < D	A < B < C < D



$$s^2 = 1,2638$$

mit Gewichtung $\frac{2}{3}$ für X zu $\frac{1}{3}$ für Y

	$\frac{2}{3} X$	$\frac{1}{3} Y$	$\frac{2}{3} X + \frac{1}{3} Y$
A	$\frac{4}{3}$	$\frac{1}{3}$	$\frac{5}{3} = 1,67$
B	$\frac{8}{3}$	$\frac{1}{3}$	$\frac{9}{3} = 3,00$
C	$\frac{2}{3}$	$\frac{5}{3}$	$\frac{7}{3} = 2,33$
D	$\frac{6}{3}$	$\frac{5}{3}$	$\frac{11}{3} = 3,67$
	A < C < B < D		



$$s^2 = \frac{5}{9} = 0,5555$$

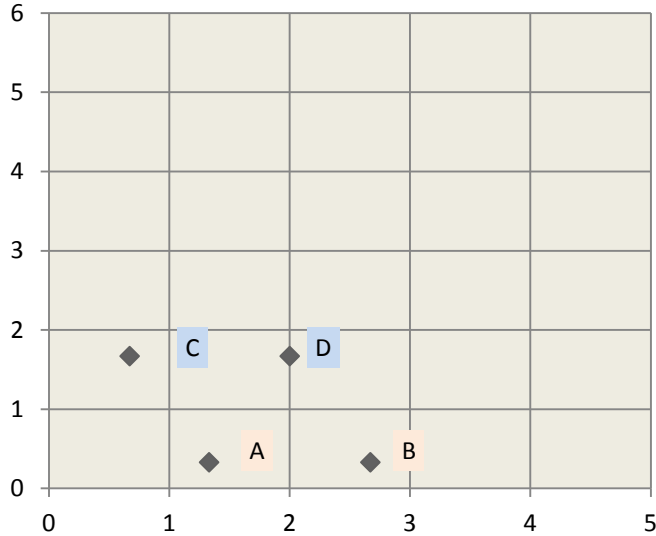
Varianz der Gesamtpunktzahl wird kleiner
Punkte nähern sich mehr der X-Achse

Regressionsgerade ändert sich, aber nicht die Korrelation

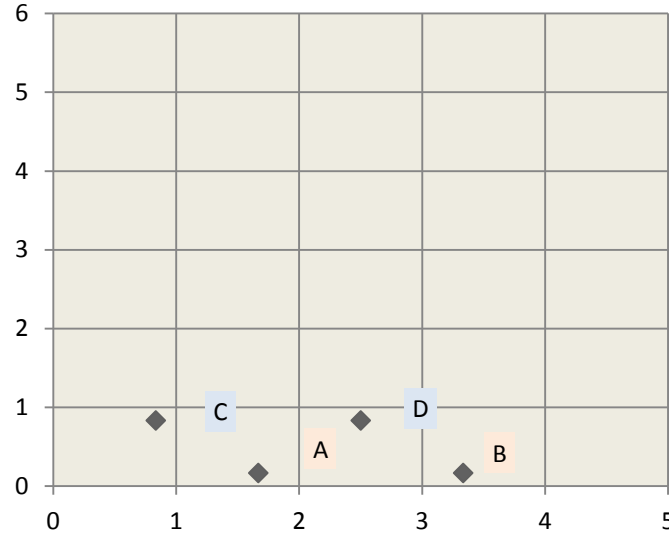
(sie ist invariant gegenüber linearen Transformationen der Variablen X und Y)

Gewichtung und Korrelation bei zwei Dimensionen

Gewichtung 2/3 für X zu 1/3 für Y



Gewichtung 5/6 für X zu 1/6 für Y



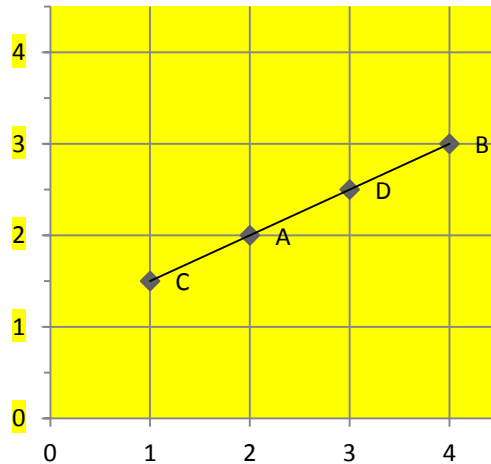
Daten nähern sich
mehr und mehr der
X-Achse

	X
C	1
A	2
D	3
B	4

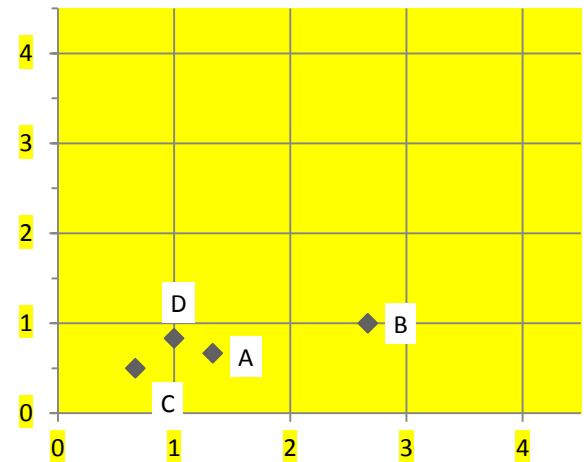
Bei Korrelation $r = 1$ der beiden Dimensionen

	X	Y	$(X+Y)/2$
A	2	2	4
B	4	3	7
C	1	1,5	2,5
D	3	2,5	5,5
	C < A < D < B		

funktionaler Zusammen-
hang $Y = 1 + \frac{1}{2} X$



Gewichtung 2/3 für X zu 1/3 für Y



Messung: Definitionen, vier Beispiele (aktuell BIP/GDP in den USA)

1. **BIP** Inhalt der neuen BIP Festlegungen des BEA in den USA: Wirkung auf das BSP: + 3%

**Data shift
to give 3%
lift to US
economy**
Intangible assets add billions to GDP

Financial Times 4/22/2013

The revision is "equivalent to adding a country as big as Belgium to the estimated size of the world economy"

	bn \$	% of US GDP
R & D investment (fixed capital formation)	300	2.1
Artistic originals	70	0.5
Housing transfer costs	60	0.4
Accounting for pensions*	30	0.2

* Zusagen, nicht Auszahlungen bei Betriebsrenten

BIP/Wohlfahrt eine endlose Geschichte → "GDP and beyond" (Nutzen/Glück ...)

Der wichtigste Unterschied: **bei GDP** "fundierte Schätzung", **bei beyond** nicht

Messung: Definitionen (Boskin Commission)

2. Armutsmessung

nicht **absolute** Armut, sondern **relativ** in ... % des Medianeinkommens (als Äquivalenzeinkommen)

3. Vermögensverteilung

indir. Eigentum d. Bankkunden

4. "Nutzen" bei Inflationsmessung

nicht Warenkorb sondern Ausgaben für den gleichen Nutzen



BOSKIN Report 1995/96: Budgetdefizite (Indexierung in den USA) zu hoch weil Inflation um ca 1,1 Prozentpunkte überzeichnet (= 148 Mrd \$ mehr Neuverschuldung 2006 wegen "Indexfehler")

statt	sollte man betrachten
teure Operationen	billigere medikamentöse Behandlung
Preise in Restaurants	jetzt billige Fertiggerichte im Supermarkt
hohe Buchpreise	Ausleihen in Bibliothek (noch keine E-Books)

auch inflationsmindernd nach Boskin etc.

- *tanken* jetzt auch *mit Kreditkarte* möglich
- *mehr Auswahl* (jetzt auch deutsches Bier im U.S. Supermarkt)

Problem mit Nutzen

- prinzipiell grenzenlos,
- subjektiv
- nicht nur durch (käufliche) Güter bestimmt \Rightarrow Glück

Messprobleme: ist das Manipulation?

nach Krämer ist dies (z.B. auch Fehlinterpretation von Signifikanztests) "*indirektes Lügen*" mit Stat.

"GDP and beyond"

es soll auch einbezogen werden (neben Umwelt etc.) Gesundheit, Chancengleichheit, Rechtsstaatlichkeit, "social connections" * etc.

Eurostat 111 deutsch/franz. SVR 25 Indikatoren *; immerhin erfreulich: jetzt Verzicht auf Gewichtung und Aggregation

"**Glück**" (happiness) : was tun, wenn man etwas nicht messen kann?

1. die Leute befragen
2. Kommissionen bilden
3. eine Flut von Grundsatzpapieren

Kein "measurement of **societal progress**" wird mit Fliegen ohne Kompass verglichen! **Computopia**
Wie so oft: **keine Zahl = Katastrophe**

Befragung in England (mit Skala 0 bis 10):

1. Overall, how satisfied are you with your life nowadays ?
- 2/3 Overall, how **happy/ anxious** did you feel **yesterday** ? *was kann man daraus folgern???*
4. Overall, to what extent do you feel that the things you do in your life are worthwhile?

Wie haben das die Menschen bloß früher ganz ohne Statistik ausgehalten?

* u.a. frequency of time spent with people at sport events

Messung, Wirtschaftsdaten: Adäquation und "big data"

Neue Herausforderungen an die amtliche Statistik

"big data"

prozessgenerierte Daten, nicht nur – wie schon bisher – in der Verwaltung, auch "technische Protokolle" automatische Verkehrsdaten (Internet-, Handy-Nutzung, GPS; Scannerdaten)

Private Konkurrenz

Bsp.: Preisindizes mit sehr **vielen, schnell** verfügbaren Scannerdaten; amtl. Statistik kann nur mit besserer Methodik (z.B. reiner Preisvergleich) dagegenhalten.

Problem der Adäquation
(Operationalisierung)



Früher Adäquation **vor** der (Primär)erhebung
jetzt

zuerst hat man die **Daten** und
dann macht man sich **Gedanken** ,
was sie besagen und wofür man sie
brauchen kann

Wen und wie viele befragen, biased sample

biased sampling frame

Fragen nach Krankheiten (Daten aus Patientendateien),
"Krebsatlas"

Selbstselektion durch nonresponse

"Nehmen Sie gern an statistischen
Befragungen teil?" ja $\approx 100\%$ ↓

Wen über was befragen?

"Sind Sie dafür, dass die Reichen höhere
Steuern zahlen?" ca. 90% antworten mit ja

Angenommen nur 10% der Bevölkerung sind "reich"; dann hätte die Frage auch lauten
können: "Sind Sie dafür, dass die **Anderen** höhere Steuern zahlen?" →

"zahlen"? ein ganz schlechtes Wort, dafür jetzt sagen "Verantwortung über-
nehmen"; oder "in ein solidarisches Europa 'investieren'" (Zwangsanleihe)

survivor bias: Lohnende Geldanlage in (alten) Gemälden weil mit ihnen bei
Auktionen enorme Preise erzielt werden.

Nicht repräsentativ für **alle** Gemälde "...die Stichprobe ist deutlich zugunsten hoher Preis-
sprünge verzerrt; genauso können wir auch zeigen (wir fragen nur die Gewinner), dass
Lottoscheine gute Kapitalanlagen sind" (W. Krämer, So lügt man mit Statistik, Neuauflage 2011, S. 111)

Wen und wie viele befragen, biased sample

biased sample

Vergleich auch methodisch verfehlt

Dirigenten leben im Schnitt Jahre länger als die "normale" Bevölkerung (Musik ist gesund und verlängert das Leben)*

Auswahl verzerrt, weil betrachtete Gesamtheit (Dirigenten) ein bestimmtes Alter erreicht haben muss (auch hier survivor bias)

* nach Abelson, auch bei Krämer S. 110 erwähnt



Verglichen wird durchschnittl. Sterbealter der Dirigenten mit der Lebenserwartung e_0 (gem. Sterbetafel)

Repräsentativität

wenn es in der Grundgesamtheit gleich viele Männer und Frauen gibt

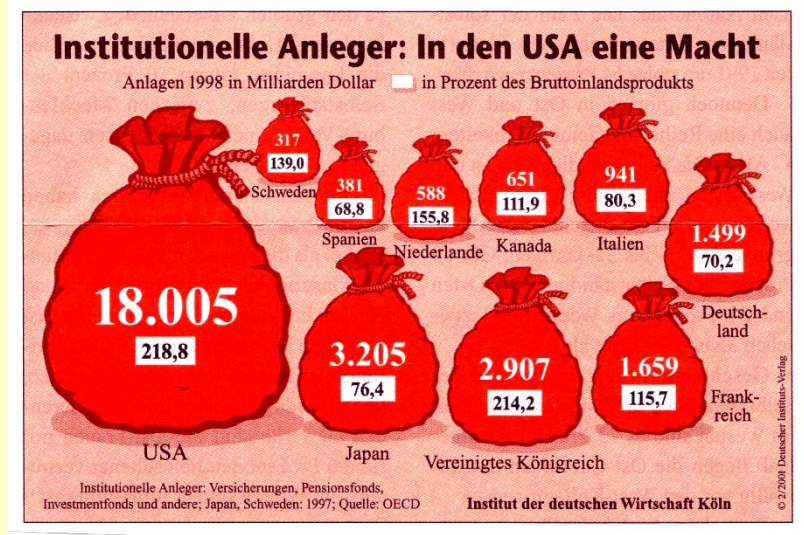
kein Fachausdruck; richtig: Stichprobenfehler

dann wäre eine Stichprobe mit 5 Männern und 5 Frauen "repräsentativ" aber eine Stichprobe mit 508 Männern und nur 492 Frauen nicht (oder weniger) "repräsentativ"

2.

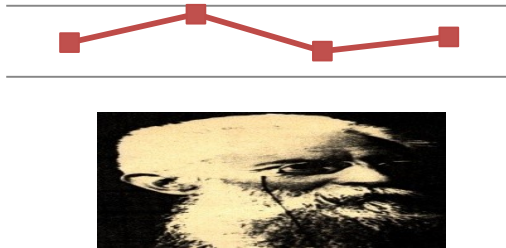
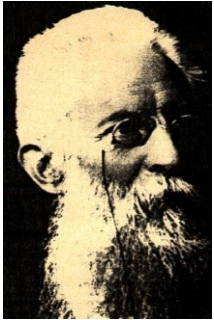
- **Daten auswerten** (Instrumentenkasten)
- **Auswertung kommunizieren und kritisieren**

- Jetzt auch mehr über "zählen" und "erzählen" und
- was man tun kann, wenn einem das Ergebnis einer Statistik, bzw. ein entsprechender Vortrag mit Statistiken nicht gefällt (vgl. entspr. Satire auf meiner Homepage)



Drei einfache Tricks

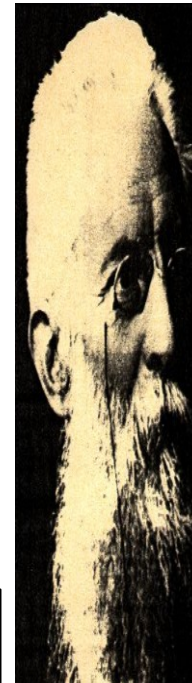
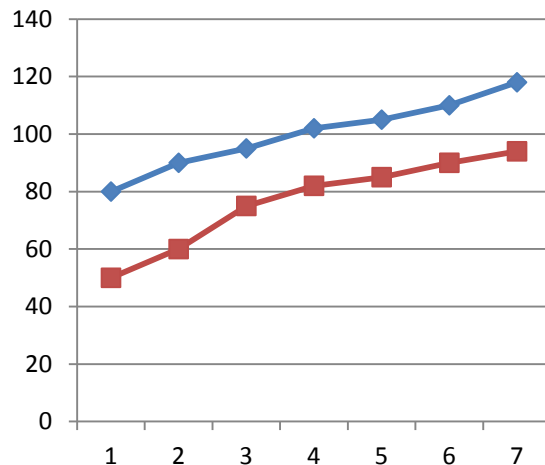
1. Maßstäbe auf den Achsen



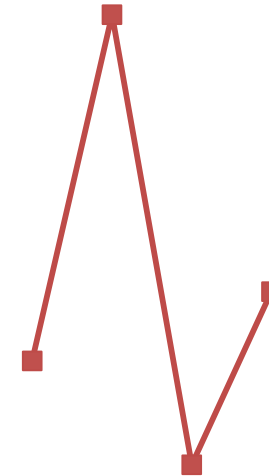
Beim Gesicht gibt es ein *gegebenes* (natürliches) Verhältnis von Höhe und Breite

2. Basis bei Messzahlen und Indizes

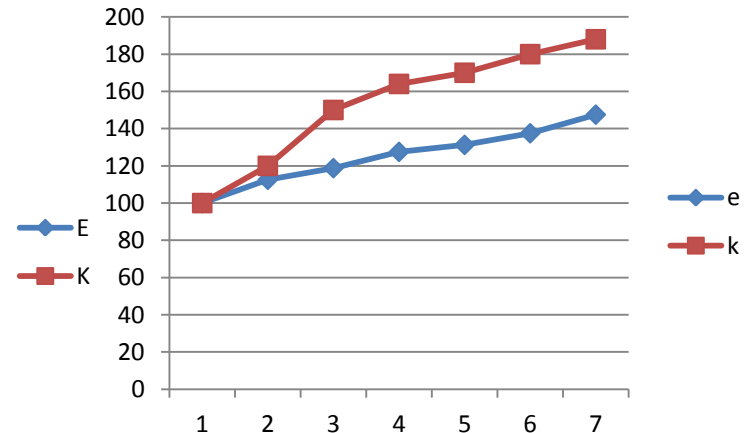
Erlös und Kosten als absolute Größen **E, K**
als Messzahlen **e, k**



Etienne Laspeyres
1834 - 1913



3. später: Die "richtigen" Zahlen auswählen \Rightarrow Statistik und Astrologie



weitere einfache Probleme

Prozente

oft Schwierigkeiten schon mit ganz einfacher Mathematik

10% Preissteigerung = 10% weniger Kaufkraft?

100% Preissteigerung = ??

München um 20% teurer als Bayern, aber Bayern nicht um 20% billiger als München

Zunahme von 4% auf 6% ist Zunahme **um 50 Prozent**, aber nur um **zwei Prozentpunkte**

Preise können um **100%** oder mehr **steigen**,
aber sie können nicht um **100%** oder gar mehr **fallen**

(bei 100% ist Preis dann ja 0)

100% von nichts ist eben auch nichts

Wahl der Bezugsgröße

Um das Todesrisiko des Reisens im Flugzeug mit dem der Deutschen Bahn zu vergleichen ist die richtige Bezugsgröße:

- die Anzahl der transportierten Personen (P),
- die geflogenen (gefahrenen) Kilometer (= Strecke S),
- die Personenkilometer (Produkt von P und S)
- die Zeit, die man im Zug/Flugzeug verbringt,
- die Anzahl der Flüge (Bahnfahrten),
- die Anzahl der entsprechenden Fahrzeuge (Züge bzw. Flugzeuge)?

Logik: je länger die Strecke, desto mehr Tote sind auch gerechtfertigt (Sterben beim Flug nach Tokyo weniger schlimm, als auf Bahnfahrt nach Hannover)

Implizite Annahmen 1

K. Rudzio (Schlagzeile ↓)

"Wer wird hier wirklich ausgebeutet? Viele Medien beklagen, dass der Staat immer öfter die Löhne aufstocken muss. Doch das stimmt so nicht"*

in **Die Zeit** vom 16.05.2013 (Nr. 21, Zeit-Online)

Annahme:

Aufstocker arbeiten voll, aber können davon nicht leben, weil Stundenlöhne zu niedrig sind .

Wenn sie beim Job-Center Geld bekommen ist das staatlich (mit Steuergeldern) finanziertes Lohndumping

Tatsache ist aber: Es gab

1. nur bei einem Teil der Aufstocker einen Zuwachs
2. andererseits aber eine Zunahme bei der soz.vers.pfl. Beschäftigung

3. Aufstocker oft bei Teilzeitarbeit und weil Partner Job verloren hat.
4. von 1,2 – 1,4 Mill Aufstockern nur ca. 35000
2,7% vollzeit erwerbstätig, dar. ca. 5000 Azubis

* weiter heißt es in dem Artikel: "ein Beispiel dafür, wie man mit Statistiken ein nahezu beliebiges Bild der Wirklichkeit erzeugen kann ... einfach indem man bestimmte Zahlen herausgreift und andere ausblendet" K. Rudzio

Aber: die Gegenrechnung im *gleichen* Artikel (!)

Aufstocken doch eine Frage des geringen Lohns?

Im **gleichen** Artikel von K. Rudzio heißt es dann aber auch "Oft trifft man Aufstocker in Putzkolonnen, in der Gastronomie oder in privaten Haushalten. In diesen Bereichen sind bis zu 15 Prozent aller Arbeitnehmer auf Hartz IV Leistungen angewiesen, während es in der gesamten Wirtschaft nur drei Prozent sind"

Aufstocker (A),
Putzkolonnen (B),
gesamte Wirtschaft (keine bedingte Wahrscheinlichkeit)
 $P(A|B) = 0,15 > P(A) = 0,03$
und Löhne in B niedriger als in der **übrigen** Wirtschaft

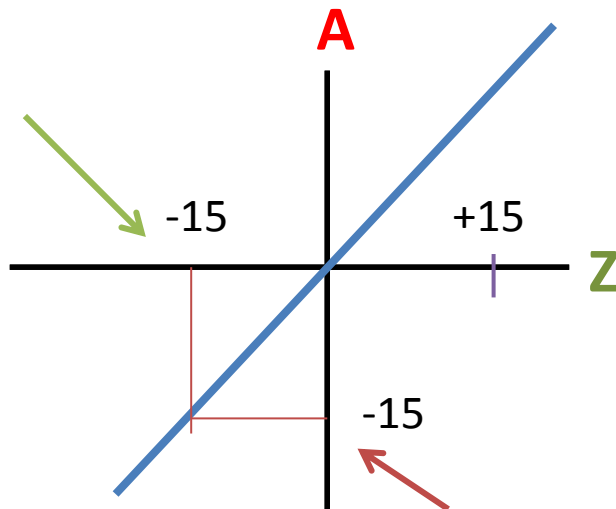
Statistik und Astrologie

1. auch wenn eine Sache mit Mathematik zu tun hat, kann sie trotzdem Humbug sein
2. weil alles (auch jeder Charakter ☿, ♀, ♁, ...) zwei (oder noch mehr) Seiten hat, kann man auch für alles eine passende Beschreibung (Statistik) finden (oder einen passenden Gutachter)
Man kann sich also aussuchen was man über ... (z.B. über sich) gerne hören/sagen möchte

Implizite Annahmen 2 (bzw. Milchmädchenrechnung)

Märkische Zeitung 90er Jahre (also ältere Meldung)

"Nach Ansicht von Brandenburgs Sozialministerin Regine Hildebrand (SPD) sollte die **Arbeitszeit** in der Bundesrepublik generell **um 15 Prozent gesenkt werden**, statt eine **Arbeitslosigkeit** von rund 15 Prozent zu **finanzieren**"



Fragen:

1. wenn Arbeitszeit (Z) um 100% reduziert wird ist dann die Arbeitslosigkeit (A) 0% oder ist sie 100%?
2. ein Rezept für Spanien, Griechenland usw.?

Regressionsgleichung $A = Z + u$ obige Gerade (**Kausalität** $Z \rightarrow A$). Im Text ist auch impliziert ein **operativer** Zusammenhang: Finanzierung (F) und A (also $F \rightarrow A$).

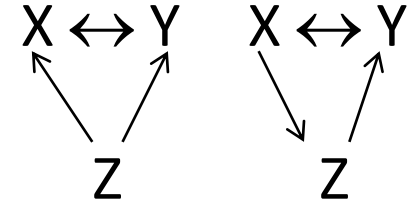
(neben $Z \rightarrow A$): Einstellen der Zahlungen (Finanzierung F), wann nimmt dann A ab??

Scheinkorrelation und Milchmädchenrechnung (1)

Abhängigkeit von einer dritten Variable Z wird nicht gesehen

Scheinkorrelation

direkter Zusammenhang wird behauptet; er besteht aber nur indirekt (über Z)



X	Y	Aussage	Z
Störche	Geburten	positive Korrelation	Trend
Geschlecht	Fahrkunst (Unfälle)	Frauen sind die besseren Autofahrer	Fahrhäufigk.
Sterberate von Priestern höher als von Bergarbeitern			Alter



Milchmädchenrechnung

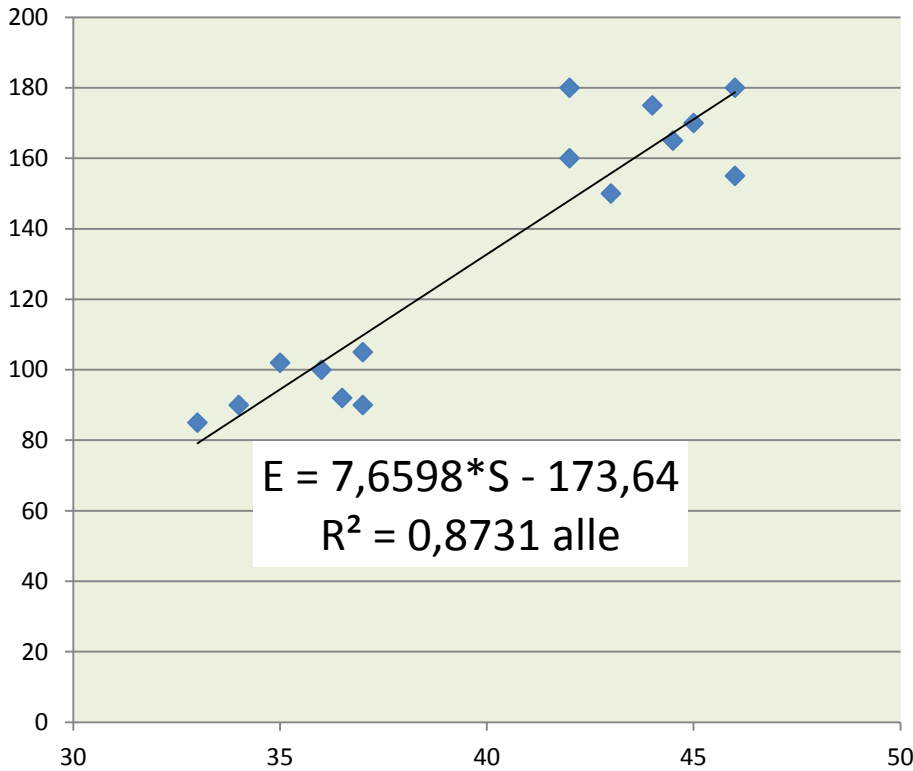
$X \rightarrow Y$ angenommen obgleich ein Zusammenhang nicht, oder so nicht besteht
 implizit angenommen, dass alle (bzw. alle anderen) Einflussfaktoren konstant bleiben

X	Y	Aussage	Z
Preis	Umsatz	Preis 10% höher \Rightarrow 10% mehr Umsatz	Reaktion der Mengen
Arbeitszeit	Arbeitslosigkeit	15% weniger X \Rightarrow 15% weniger Y (kein Einfluss von Löhnen, Reaktion der Arbeitsnachfrage etc.)	

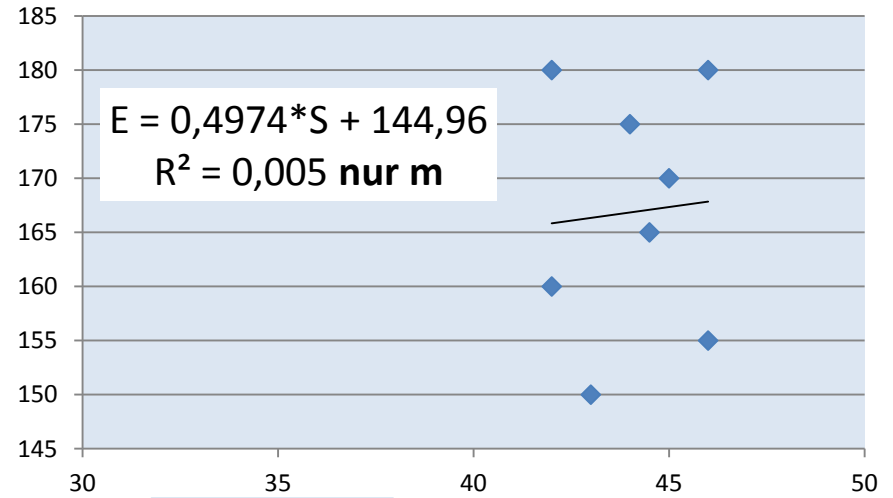
Nonsens Korrelationen (inhomogene Gesamtheit)

Schuhgröße (S) und Einkommen (E)

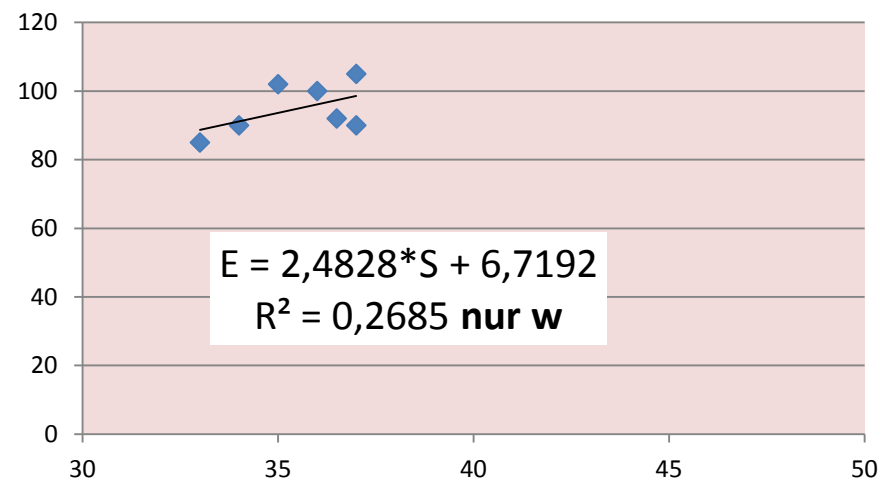
fiktive Zahlen (aus der Zeit vor Emanzipation der Frauen)



Abszisse S, Ordinate E. Hohe positive Korrelation $r = +0,9344$ weil es **zwei Cluster** gibt: Frauen links unten, Männer rechts oben



$r = 0,0707$



$r = 0,5182$

Scheinkorrelation und Milchmädchenrechnung (2)

Unterschiede operativ zu nutzen?

Selbst wenn Zusammenhang kausal wäre: wie kann man ihn nutzen??

Bericht über **Gehaltsreport* 2013** in SPON 24.5.2013 : **In der Schlagzeile steht auch (kein Witz!)**

"..... Er erklärt auch , was Sie tun können, um mehr zu verdienen"

Medianeinkommen gem. obigem SPON Artikel

Berlin	47.500 €
München Stuttgart und Rhein-Main	62.500 €
Sozialwissenschaftler	47.500 €
Mediziner	92.500 €

Tipp für einen Berliner Soziologen:

Umziehen nach München und Mediziner werden/sein

Man könnte auch raten: 20 Jahre jünger werden und dann neu anfangen (und das Richtige studieren)

Operativ nutzbarer Kausalzusammenhang

Essen im
Flugzeug



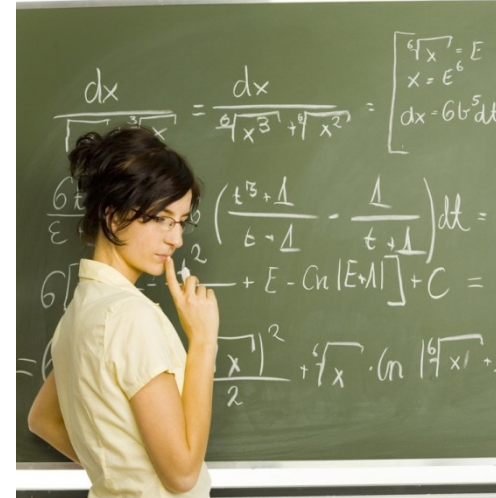
* Befragung mit 107.000 Teilnehmern.
SPON = SPIEGEL online

- **Zum Fach "Statistik" und**
- **wie Statistik kommunizieren?**

wie Statistik gelehrt wird und warum keiner das Fach mag

Warum ist Statistik so unbeliebt?

- (1) deutsches **Bildungsideal**: in D ist es nicht peinlich zu sagen, man habe Mathematik/ Statistik nie gekonnt (es gilt im Gegenteil als menschlich/sympathisch) weil Stat. allgemein als schwer empfunden auch Schaumschlägerei überhaupt erst möglich
- (2) anders als sonst ist bei diesen Disziplinen **verstehen**, aber **nicht auswendig lernen** notwendig (auswendig lernen nützt gar nichts) Stat. fällt aus d. Rahmen



Fortschreitende Perversion

wie macht man eine Vorlesung ?

- (1) mit copy & paste aus U.S. Lehrbücher viele ppt-Folien machen
- (2) sehr selbstbewusst die Folien auf Englisch/Denglisch abspulen (= vortragen)
- (3) Studenten haben nichts verstanden, aber Folien auswendig gelernt. Sie haben
- (4) alles schnell wieder vergessen und es gibt nichts, worauf man aufbauen kann

und einen Seminarvortrag
vgl. Satire "Statistik für Schaumschläger"

- (1) die "Metamorphose des Durchschnitts-Ökonomen"
- (2) Vortrag darf nicht einfach verständlich sein; er soll primär zeigen, wie schlau man ist

Wie gegen (einen Vortrag mit) Statistiken ankämpfen? *

Zwölf Regeln (auch anwendbar, wenn man nicht sehr kompetent in Statistik ist *)

1. keine Zeit für komplizierte Ausführungen die ohnehin die wenigsten Anwesenden interessieren
2. Sich gut vorbereiten
3. Sich mit "Fragen" einbringen; aber sich nicht lächerlich machen
4. "Wer die Erhebung/Auswertung gemacht hat, ist nicht neutral/seriös"
5. "nicht repräsentativ"
6. "nicht vergleichbar" (oder "nicht anwendbar auf.../relevant für...")
7. Mehr Zahlen (eigene ausgewählte Zahlen) präsentieren oder gar keine Zahlen (einen "impressionistischen" Gegenschlag versuchen)
8. Aufgeben, wenn eine Autorität (Typ "elder statesman") den Weltgeist beschwört und gegen Zahlen immun ist ("der € ist weit mehr als eine Währung")
9. Scheinkorrelation und Milchmädchenrechnung aufdecken
10. Eine Statistik/Erhebung gezielt überfordern
11. Den "Sinn" einer Zahl bzw. Aussage bestreiten
12. Die Sache sportlich sehen

♣ aber nicht anwendbar bei der üblichen Schaumschlägerei an der Uni.

* Hierzu die Satire (zum Download) "Was tun ..." auf meiner Homepage

zu 4 und 6: Seriosität und Vergleichbarkeit

4. Warum ist die Seriosität des Statistikproduzenten so wichtig?

- Bei Statistik gibt es eine Manipulationsvermutung (man kann sie parteilich nutzen [Propaganda]), andererseits braucht Statistik aber Vertrauen (denn man kann sie meist nicht als Außenstehender nachprüfen)
- einziger u. leicht verfügbarer Indikator für "nicht manipuliert":
 - 1. Neutralität/Unparteilichkeit** des Statistikproduzenten; daneben auch wichtig
 - 2. Kompetenz** in Methoden (Beweislast)
- in Amtlichkeit liegt auch das eigentliche Kapital der *amtlichen* Statistik (Ist sie sich dessen bewusst oder versucht sie sich anders zu profilieren? Hat sie das Prestige verdient?)

6. Wann sind Zahlen "vergleichbar"?

- in Statistik keine absolute Größen, Zahl einschätzbar immer nur durch Vergleich
- eine Frage danach, wie genau man hinschauen will; mit wie vielen Merkmalen etwas beschrieben/definiert wird
- man muss von der Sache etwas verstehen, um zu erkennen, was die Vergleichbarkeit beeinträchtigen kann (Beispiel Professorengehälter in WT und warum US Unis beim Ranking Vorteile haben [Matth. 13, Vers 12] → DUZ Artikel 31.5.2013)
- eine Variante des Arguments: "interessant" (oder sogar richtig) **aber** nicht anwendbar auf ...

Warum funktioniert Nr. 7 und die impressionistische Masche?

Weil die meisten Menschen ein **ambivalentes** Verhältnis zu Zahlen haben

Es gibt Leute, die es schaffen, den Eindruck zu erwecken, dass wir einer Katastrophe entgehen, weil wir über irgendetwas immer noch keine Statistik haben, oder weil die entsprechende Statistik erst in einer Woche kommt und nicht schon jetzt da ist. Wenn sie dann aber die angeblich so alles entscheidende Zahl "endlich" sehen, dann fällt ihnen kaum mehr dazu ein als "interessant" oder "hätte nicht gedacht, dass es so groß/klein (oder viel/wenig) ist" (vdL "Was tun...").

"There is something very sad in the disparity between our passion for figures and our ability to make use of them once they are in our hands." "My experience is that people have a sort of religious awe about figures." M. J. Moroney, Facts from Figures, S. 458, 459.

zu 11: Warum fragt man nach dem "Sinn" einer Statistik?

Warum wird dem "Gestern-Erklärer" (v.a. beim Börsengeschehen) mit Interesse zugehört, obgleich er doch gestern noch nicht voraussah, was heute passiert ist?

zu 11: Sinnhaftigkeit und Genauigkeitswahn (Folie nicht zentral)

Operationalisierungsprobleme

Sinnlos = nutzlos oder trivial : was für nichts zu gebrauchen ist/ohnehin gedacht hat
Kinder mögen lieber Feiertage, an denen sie etwas geschenkt bekommen (z.B. Geburtstag)

Sinnvoll = interessant: was macht eine Statistik "*interessant*" ?(MAGIC, nachAbelson)

Magnitude effect size

Articulation degree of comprehensible detail in which conclusions are phrased

Generality breadth of applicability of the conclusion

Interestingness potential to change what people believe about important issues

Credibility methodological soundness and theoretical coherence (e.g. no fishingness)

Kann man die Sinnhaftigkeit einer Zahl "von außen" erkennen?

1. Nein, man sieht es v.a. nicht an der "Genauigkeit" der Zahl (Zahl der Nachkommastellen)
2. Fehler in der Erhebung können durch die Auswertung nicht behoben werden, aber
3. Auswertung kann sich von Datenbasis entfernen und reine Fiktionen produzieren (Hätte-Könnte-Sollte [HKS] Rechnungen)

Statistische und impressionistische Methode

Impressionistische Methode: eine statistische Information wird mit einer entgegengesetzten Erfahrung einer Person, die als kompetent ("betroffen") gilt, ausgehebelt

statistisch	impressionistisch
zählen	erzählen (Anekdoten, emotional)
objektiv, nüchtern, formal	subjektiv, plastisch, lebendig
verlangt Vorbildung muss man verstanden haben (haben viele nicht)	kann jeder schnell verstehen (für viele leichter zugänglich und überzeugender)
verallgemeinerungsfähig, wissenschaftliche Methode	nicht repräsentativ, plausibel, gesunder Menschenverstand reicht
verlangt Akzeptanz (man muss glauben, dass richtig gerechnet wurde); kann auch schnell langweilen	wird akzeptiert, wenn Rhetorik gut ist und Person angesehen ist; kann man gebetsmühlenhaft wiederholen
Akzeptanz kann schnell zerstört werden	Fazit: <u>sehr</u> viel besser zur Manipulation geeignet

Keine Statistik, wohl aber Impressionismus, wenn

impressionistische Methode
funktioniert auch (gerade), wenn

zu 1

1. es **keine Daten** (Beobachtungen) gibt, etwas selten ist
2. die Variable **nicht klar definiert** ("Seit ... hat die soziale Kälte (wie messen?) enorm zugenommen")
3. es um Werte und **Emotionen** geht ("Der € ist weit mehr als eine Währung") und wenn
4. es darauf ankommt, **wer** etwas sagt ("**Autoritätsbeweis**")



Ein (akzeptierter [Autorität!] und rhetorisch begabter) Kenner der Verhältnisse kann mit seinen persönlichen Erfahrungen eine Statistik "aushebeln" (unser Erbe "we are story tellers"). Aber: In Statistik **viele** Daten **von no-names**, nicht **ein** statement **von einer** **Autorität**

Womit Eindruck machen: Schaumschläger/elder statesman

Statistik allein wird kaum der Renner sein

- meist interessieren nur (ins Verbale übersetzte) Ergebnisse, nicht Methoden
- viele haben Schwierigkeiten, gute und schlechte Statistik zu unterscheiden
- gerade gute Statistik verlangt viele Überlegungen und daher oft auch viele **Erläuterungen** und das **kann man sich ersparen mit**

Effekthascherei

A: Protzen mit Mathematik

→ Statistik für Schaumschläger

B: Beeindrucken mit Erfahrung
und Detailkenntnissen

= der impressionistische Gegenschlag ;
(schwerer als A ?) es braucht

1. eine respektable Persönlichkeit
2. ein überzeugendes Auftreten

Beides erschwert, gute und schlechte Anwendung der Statistik zu unterscheiden, was aber wichtig ist ⇒

gleichwohl ist die Spreu vom Weizen zu trennen

denn ungeachtet der Effekthascherei gilt:

Es gibt gute und schlechte Statistik. Unterscheidung notwendig, aber wie erkennen? Allgemeine operationale Kriterien hierfür?*

Publizieren ?

Fördern, Finanzieren ?

einzelne Arbeiten
(ex post, quasi retail)

Wissenschaftler, Projekte (ex ante),
Institute + Fachbereiche (wholesale)

methodische Arbeiten

Peer reviews bei Veröffentlichungen

beachtete Ergebnisse

"Nachrechnen" und Streit in
der scientific community über
Ergebnisse eher selten

aktuell **Beispiel:**
Rogoff/Reinhart
Staatsverschuldung

* "**Meta-Analysen**", neue "Metriken" (**Zitatezählen** nicht nur bei tradit. print-Veröffentl. DUZ-Artikel)

schon aus rechtlichen
Gründen unattraktiv

Fazit zu Manipulation mit Statistik (1 von 3)

- Ja, man kann mit Statistik manipulieren und das muss auch nicht gleich verwerflich sein und es kann wie gezeigt positiven Nettonutzen haben
 - Man kann aber mit Worten und ganz ohne Zahlen ("impressionistisch") noch viel mehr/besser/leichter manipulieren (und das kann auch überzeugender wirken und sogar richtige statistische Ausführungen unwirksam machen)
-
- Anders als Mathematik gibt es kein richtig/falsch (man kann gegen **jede** Statistik etwas einwenden)
 - Statistik verlangt **Akzeptanz** und **Vertrauen** (man kann nicht alles nachrechnen oder gar auf eigene Faust nacherheben und die meisten Menschen können auch nicht mathematisch alles nachvollziehen [sie geben es nur nicht zu])
 - ein wunder Punkt für die Statistik, aber (**positiv**)
Man kann gegen **diese Art** Manipulation etwas tun (argumentieren)

"The fact is that despite its mathematical base, statistics is **as much an art as it is a science**" (Darrell Huff: How to lie with Statistics, Pelican Books 1973 ff, p. 109)

Fazit zu **Manipulation mit Statistik (2)**

Das Problem mit der Statistik ist weniger die Manipulation als die folgenden **vier Probleme**

1. unsere **Obsession, überall Zahlen**, Rankings und (zweifelhafte) Messungen zu verlangen ("**Computopia**")
 - und das trotz unserer eher schwachen Fähigkeit, statistische Methoden verstehen und beurteilen zu können
 - und der Gefahr kreativer Buchführung und einer Flut intransparenter privater Statistikprodukte
2. Neigung (in der Politik) , **Statistik selektiv** zur Unterstützung, oder als Sündenbock zu **benutzen, oder** sie impressionistisch ganz **auszuhebeln**

Fazit zu **Manipulation mit Statistik (3)**

3. die **Instrumentalisierung der Statistik** (zur Profilierung in der Wissenschaft) und gedankenlose, ritualisierte Anwendungen der Methoden
4. (in Zukunft) dank **gestiegener technischer Möglichkeiten** Überschwemmung mit **big data** und Analysen per data mining

"More data inevitably results in more time spent arguing, validating, reconciling and replicating. There is a real danger that Big Data moves us backward, not forward. It threatens to take science back to the Dark Ages, as bad theories gain ground by gathering bad evidence and drowning out good theories."*

Nicht nur mehr, auch schlechte Statistiken (\Rightarrow Folie 6; Nettonutzen > 0)

* Kaiser **Fung, Numbersense**, How to use Big Data to Your Advantage, 2013 (Mc Graw Hill+Kindle)

etwas an den Rand gerückt: \Rightarrow **Wahrscheinlichkeitsrechnung** (nächste Folien)
notorisch die größten Verständnisprobleme aber **das** Instrument der Statistik

Abschließend (oder ergänzend nur noch 4 Folien)

ganz kurz etwas zur

Wahrscheinlichkeitsrechnung

und Analyse mit stochastischen Modellen

(und ein Versuch zur Systematisierung der "Lügen" mit Statistik)

Das Problem mit den fortgeschrittenen Gegenständen der Statistik und Ökonometrie ist, dass sich heutzutage jeder Ökonom (und bald auch jeder Soziologe und Politologe usw.) als Fachmann auf diesen Gebieten fühlt oder zumindest glaubt sich so aufführen zu müssen.

Das ist eine ernst zu nehmende Krankheit \Rightarrow v.d.L. "Statistik für Schaumschläger"

Was bringt uns die Wahrscheinlichkeitsrechnung? 1 (von 4)

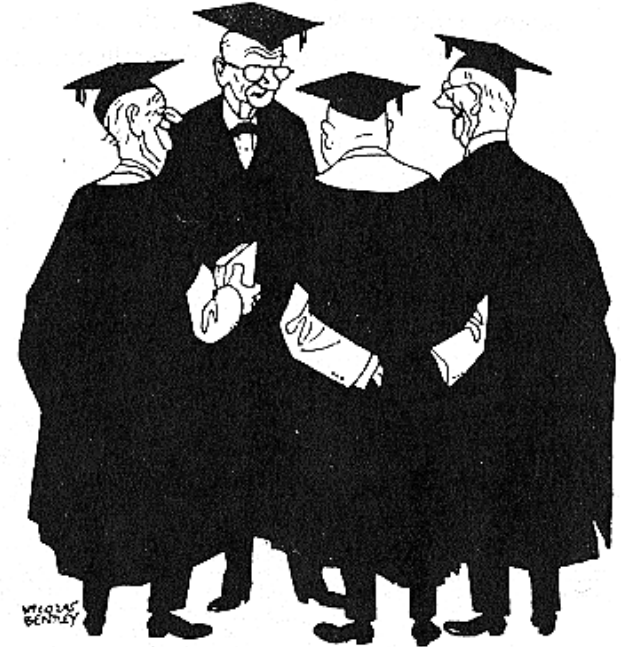
1. Berechnung einer Wahrscheinlichkeit

Die Universität X habe eine Professur zu besetzen: Die für diese Stellenbesetzung eingerichtete Kommission von vier im eigenständigen Denken geübten Herrn gerät so sehr in Streit, dass sie sich bei ihrer Auswahl *von keinem erkennbaren Prinzip mehr leiten lässt*. Vier von den 20 Bewerbern sind jedoch für die ausgeschriebene Stelle absolut ungeeignet:

Wie wahrscheinlich ist es, dass in der Dreierliste

a) ein, b) zwei, c) drei, d) mindestens ein ungeeignete(r) Bewerber/in erscheinen (bzw. erscheint)?

a) 35,79%, b) 8,95%, c) 0,36% d) damit *mindestens ein(e) Ungeeignete(r)* $511/1140 = 0,44825$ also 44,8%



Das sagt uns

- dass dies wahrscheinlicher/weniger wahrscheinlich ist als ...
- aber **nicht**, was bei der kommenden Berufungskommission im Fach/an der Univ. passieren wird

Man könnte aber ein *Modell* in Abhängigkeit von Merkmalen von Profs. bilden (*fragwürdig*)

Was bringt uns ... ? 2 Stichproben: Schätzen und Testen

2. Stichproben

"Die Untersuchung des Psychologen P bei Angela, Bert, Charlie, Doris ... hat ergeben, dass nur jede(r) Vierte die Wahrscheinlichkeitsrechnung mag"

Wenn A, B, C, D eine **Stichprobe** von Wiwi-Studenten aus NRW wären:

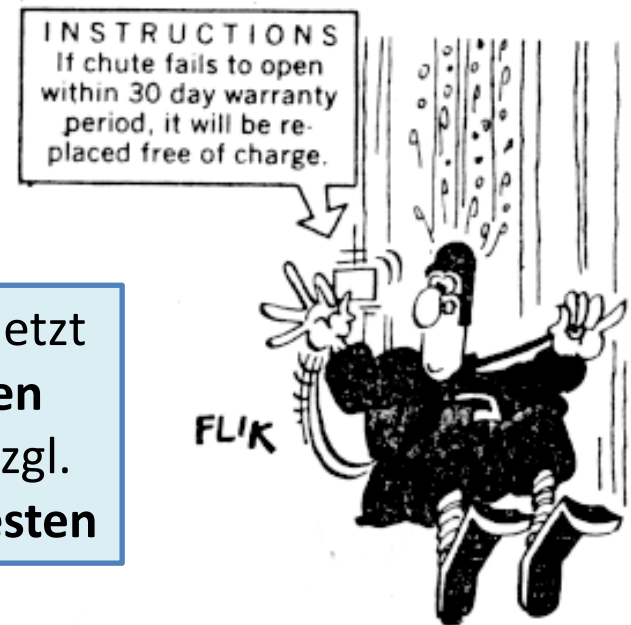
"Die Untersuchung von P hat ergeben, dass nur 25% der Wiwi-Studenten in NRW die Wahrscheinlichkeitsrechnung mögen"

Stichproben oft die einzige Möglichkeit

Beispiel: Qualitätskontrolle (Garantieverpflichtungen), Erntevorausschätzung

Grundgesamtheit (GG) unbekannt: man muss daher jetzt

1. Werte der GG (etwa mittl. Reißfestigkeit μ) **schätzen**
2. über eine Hypothese (H_0 : Fallschirme sind "gut") bzgl. der GG entscheiden (H_0 annehmen/ablehnen) = **testen**



Was bringt uns? **2a** noch Schätzen und Testen von Hypothesen

95% Schwankungsintervall in Abhängigkeit von n

Ist ein Mittelwert von 98,5 bzw. 101,5 "signifikant"?

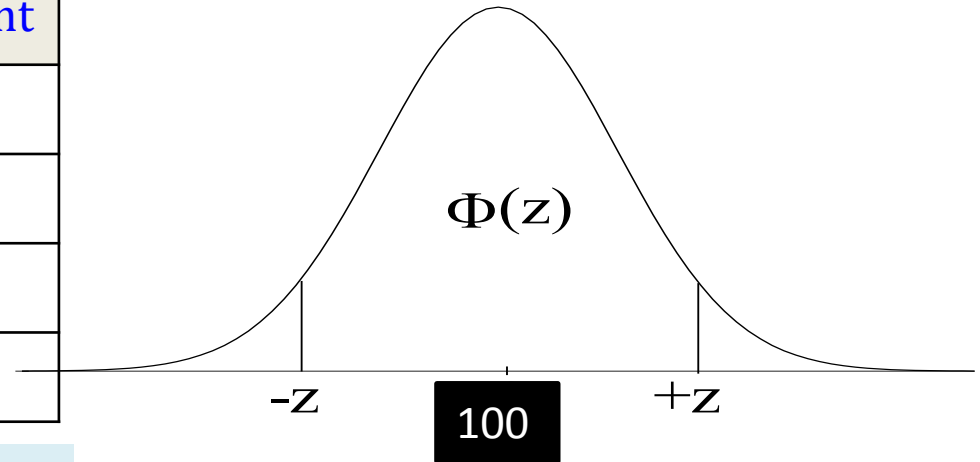
n	von	bis	signifikant
100	96,86	103,12	nein
400	98,43	101,57	nein
1000	99,01	100,99	ja
10000	99,69	100,31	ja

"signifikant"? Kein Problem, wenn nur die Stichprobe groß genug ist

Entscheidungsproblem

H_0	H_0 richtig	H_0 falsch
ablehnen	α Fehler (1.Art)	kein Fehler
annehmen	kein Fehler	β Fehler (2.Art)

Grundgesamtheit $\mu = 100$ und $\sigma = 16$
 Stichprobenverteilung des Mittelwerts (hier z) ist eine Normalverteilung auch wenn x in GG nicht normalverteilt ist



H_0 : Fallschirm gut
 H_0 : Frau richtig

"annehmen" heißt
 Fallschirm **kaufen**
 Frau **heiraten**

α Fehler: einen guten Fallschirm nicht kaufen,
 "richtige" Frau nicht heiraten

β Fehler: einen schlechten Fallschirm kaufen,
 die "falsche" Frau heiraten

Was bringt uns ... ? 3 Stochastische Modelle

Hat man überhaupt eine Stichprobe gezogen?

Zufallsauswahl? Aus welcher GG?

Fünf "Sozialdemokratische Wohlfahrtsstaaten" (S, N, DK, FIN/SF, NL) signifikant höhere Werte bei Verteilungs-, Bedarfs-, Generationen-, Geschlechtergerechtigkeit als ... Länder*

3. Modelle

jetzt Testen von Modellannahmen (neben Parameter des Modells)

a) Messen: Es gibt Dinge, die man nur mit einem Modell messen kann

Beispiel **Intelligenz**: "Sind Sie intelligent?" \Rightarrow Modell der Faktorenanalyse (stoch.)
Lebenserwartung \Rightarrow Modell der stationären Bevölkerung (nichtstochastisch)

b) Suche nach Einflussfaktoren und Prognose

- 1. Erklären** in Statistik = **Zerlegung einer Varianz** (von y) in eine erklärte Varianz (durch x_1, x_2, \dots) und eine (nicht erklärte) Residualvarianz; Sowohl y als auch x_1, \dots müssen variieren: eine Konstante (bzw. mit einer Konstanten als Regressor) kann man nicht "erklären"
- 2. Ursachen** durch Ausschließen(kontrollieren) anderer mögl. Einflussfaktoren (Kontrollgruppenexperiment) **Design of experiments** als ein Teilgebiet der Statistik

* Wagschal et al. in Vanberg (Hrsg.) Marktwirtschaft und soziale Gerechtigkeit 2012 Mohr/Siebeck

Nachtrag: "Lügen" oder nicht lügen?

Versuch einer **Systematisierung** von **Missbrauch und Missverstehen**

Fehlerquelle (1 -2)	Beispiele	Wertung
Was gemessen werden soll ist inhaltlich umstritten und schwer zu operationalisieren (jedes Messkonzept kann mit guten Gründen kritisiert werden)	Arbeitslosigkeit, "Armut", auch schon so vertraute Konzepte wie BIP, "Inflation"...	Kein "Lügen", weil Messung naturgemäß stets kontrovers sein wird ¹⁾
Verzerrte (biased) Stichprobe : betrifft eher Interpretation und Umfrageforschung (dagegen: Zufallsauswahl stellt sicher dass niemand bevorzugt/benachteiligt wird)	unzulässige Verallgemeinerung eines Befunds bei spezieller Teilgesamtheit ²⁾	trotz Stichprobe bias mögl. durch systemat. Nichtbeantwortung ³⁾

1) nicht selten liegt auch die Initiative und das dringende Verlangen, etwas zu "messen", was problematisch ist (z.B. "Glück") bei Statistiknutzern in Politik und Wissenschaft, und nicht bei Statistikern (⇒"Computopia")

2) Beispiel der Kapitalanlage mit Gemälden; ein bias kann auch durch entsprechende (quasi die Antworten implizierende) Fragen entstehen ("Sollen die Reichen mehr Steuern zahlen?")

3) sog. Selbstselektion einer non-response Gruppe, die sich systematisch von response Gruppe unterscheidet

Fehlerquelle (3 -7)	Beispiele	Wertung
kein oder ein falscher "standard of comparison", fehlende oder eingeschränkte Vergleichbarkeit (zwischen verglichenen Gruppen oder über die Zeit); Definition nicht homogener Teilgesamtheiten	Bezugsgrößen bei Unfallhäufigkeit; Strukturabhängigkeit bei Veränderungen beschreib. Statistiken	Problem, wenn auf begrenzte Aussagefähigkeit einer Statistik (Zahl) nicht hingewiesen wird ¹⁾
Naheliegen von Trugschlüssen (Übersehen möglicher Scheinkorrelationen)	eher Problem der Interpretation (unzulässige Kausalinterpretation oft auch kaum erkennbar)	
Wachstumsraten, Prozentsätze etc. ²⁾	Manipulationsmöglichk. aber kein "Lügen" ³⁾	
Signifikanztests als Ritual; Nichtprüfen der Modellvoraussetzungen" ⁴⁾	Nur signifikante Ergeb. mitteilen, Aussagefähig. überbewerten	eher Problem in d. Wissenschaft, weniger bei amtl. Statistik
Intransparenz, unzureichende oder unverständliche methodische Erläuterungen	keine Angabe d. Stichprobenfehlers	Statistik erläuterungsbedürftig; Bringschuld

- 1) Statistiker müssen evtl. auch Standardisierungen und Bereinigungen bei ihren Zahlen vornehmen; die Konstruktion guter Kennzahlen der beschreibenden Statistik ist ein methodisches Problem; es kann hier andauernde Kontroversen und nicht wirklich ein "richtig" oder "falsch" geben (also kaum ein "Lügen")
- 2) hier nicht speziell erwähnt (zu dreist/simpel): Tricks mit Graphiken
- 3) man kann Statistikern nicht vorwerfen dass Nutzer Prozentrechnung nicht beherrschen
- 4) z.B. bei Regressionsrechnung, multivariaten Analysen etc.