

BM-Datenanalyse eine Exploration

Marcello Robbiani

ZHAW und EBMK

www.zuguttenberg.de/dokumente/zuguttenberg_presseportrait.jpg

Ein Web-basierter Bildervortrag



Allgemeingut, Eigengewächs:
Keine Quellenangabe

Web-Material:
**Quellenangabe im jeweiligen
Folienkopf**

www.zhaw.ch/fileadmin/php_includes/popup/person-detail.php?kurzz=sick

Inspiriert durch einen Kurzvortrag von

Beate Sick

**Institut für Datenanalyse
und Prozessdesign**

**Zürcher Hochschule für
Angewandte Wissenschaften**



www.mathematikzentrum.ch/fachkonferenz-mathematik/

Eckdaten zu meiner Person:

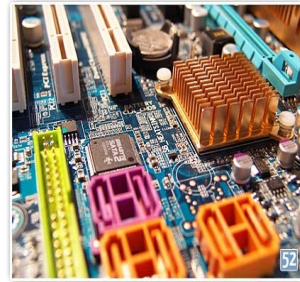
Fachkonferenz Mathematik 2009

Die künftige Berufsmaturität - offene Probleme

Wovon handelt das Folgende?

Technik & IT, Chemie & Life Sciences,
Architektur, Bau- & Planungswesen

10 % Datenanalyse
im Grundlagenfach
(20 von 200 Lektionen)



Im Folgenden:

TDA (Technik - Daten Analyse)

www.np-tinguely.ch/Chamoson/index.html

Wovon handelt das Folgende **nicht**?

Ausrichtungen:

Gesundheit, Soziale Arbeit,
Land- & Forstwirtschaft et al.

Deskriptive Statistik &
Wahrscheinlichkeitsrechnung



Inhalt

1. Statement zur TDA
2. Was sollte TDA nicht sein?
3. Was sollte TDA sein?
4. TDA Pros und Cons

Was sollte TDA sein ?

Explorativ !

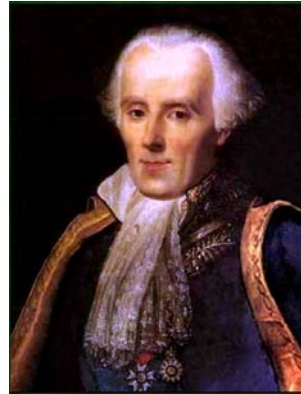
- Qualitativ
- Deskriptiv
- Graphisch
- Datengetrieben



Im Gegensatz zu ...

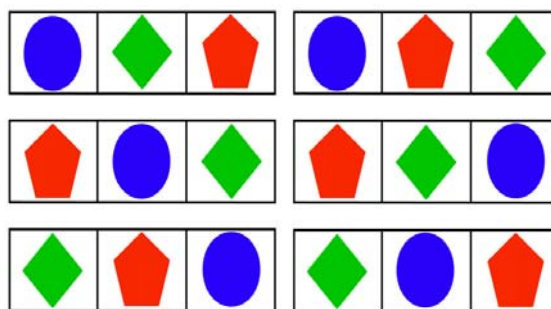
Klassisch !

- Quantitativ
- Normativ
- Mathematisch
- Modellgetrieben



htw.activemath.org/ActiveMath2/search/show.cmd?id=mbase://mathebrHTW/binom/fak_intro

Was TDA nicht ist:

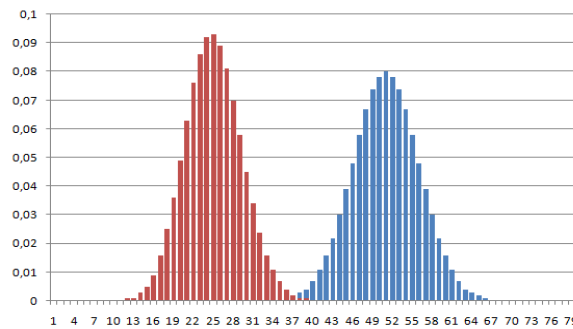


Kombinatorik ...



cheind.wordpress.com/2009/12/13/on-coin-tossing-games-part-2/

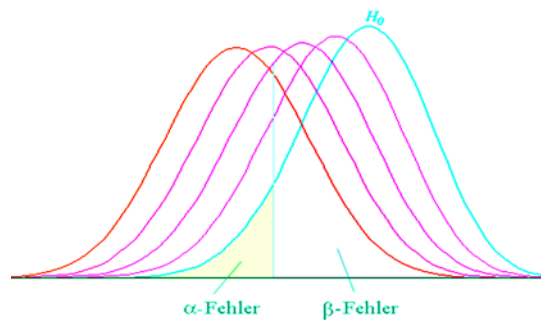
Was TDA nicht ist:



Wahrscheinlichkeitsverteilungen ...

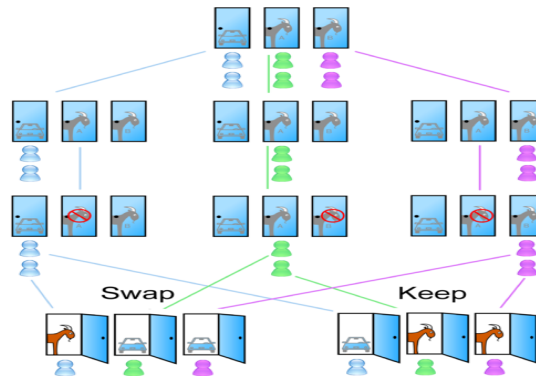
www.matheboard.de/archive/124835/thread.html

Was TDA nicht ist:



Schliessende Statistik ...

Was TDA nicht ist:



Entscheidungstheorie ...

Was ist dann TDA?

Explorativ!

Was heisst "explorativ"?

Wer steckt hinter "explorativ"?

Woher kommt "explorativ"?

Weshalb "explorativ"?

Nicht zu verwechseln mit

"explorative Didaktik"!



John W. Tukey
1915 - 2000
Exploratory Data Analysis

"Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise."

American Statistician 40 (1986)

John W. Tukey
über Explorative Datenanalyse:

"... , I would suggest that it is an **attitude** and a **flexibility** and some **graph paper** ..."

www.ams.org/notices/200202/fea-tukey.pdf

wikipedia.org/wiki/John_Tukey

**Explorative Datenanalyse
ist nicht alternativ!**

**Confirmatory Data Analysis (CDA)
sollte in **Lehre und Praxis** auf
Exploratory Data Analysis (EDA)
aufbauen ...**



cbl.stanford.edu/~langley/cogsys/behrens97pm.pdf

**John W. Tukey
über Explorative Datenanalyse:**

**"... substantive concerns
must take precedence over
statistical convenience."**

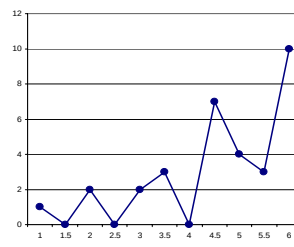
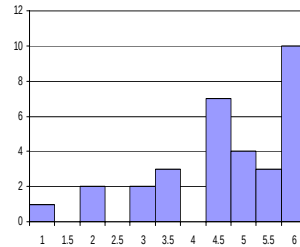
*Gesundheit, Soziale Arbeit,
Land- & Forstwirtschaft et al.*



M.Sc.E. 2009

Eine mögliche Lösung: Noten

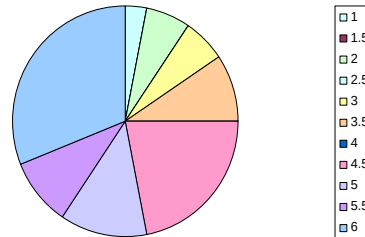
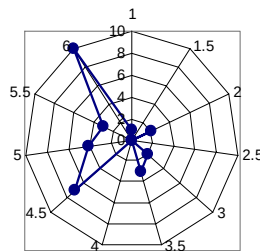
1	1	1
1.5	0	1
2	2	3
2.5	0	3
3	2	5
3.5	3	8
4	0	8
4.5	7	15
5	4	19
5.5	3	22
6	10	32



Mittelwert 4.67

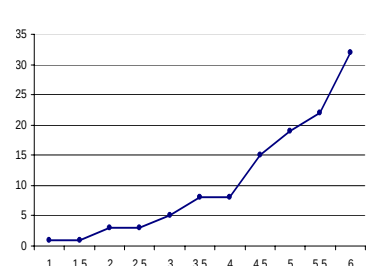
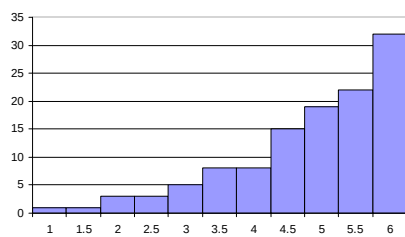
Median 4.5

Modus 6



EXCEL

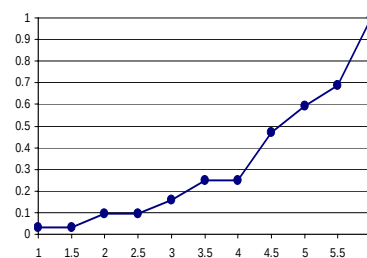
Eine mögliche Lösung: Noten



0.25-Quantil ~ 3.5

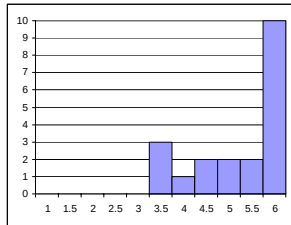
0.50-Quantil ~ 4.5

0.75-Quantil ~ 5.6

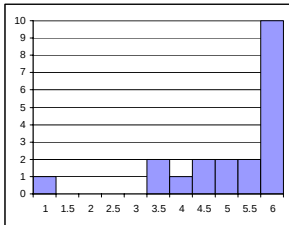


Eine mögliche Lösung: Noten

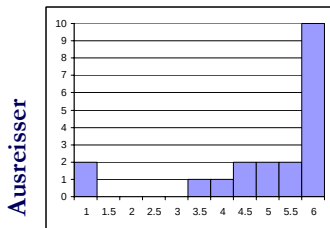
M.Sc.E. 2010



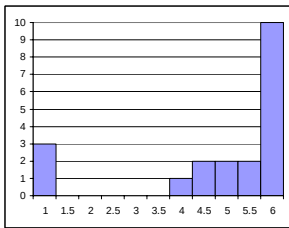
Mittelwert: 5.22
Standardabweichung: 0.97



Mittelwert: 5.10
Standardabweichung: 1.30



Mittelwert: 4.98
Standardabweichung: 1.56



Mittelwert: 4.85
Standardabweichung: 1.77

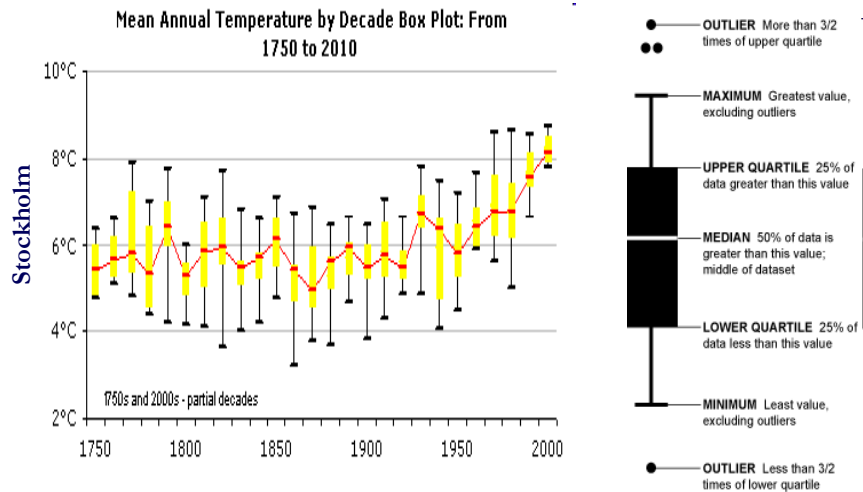
psycnet.apa.org/index.cfm?fa=buy.optionToBuy&id=2002-00947-001

John W. Tukey: Stem-and-Leaf-Plot

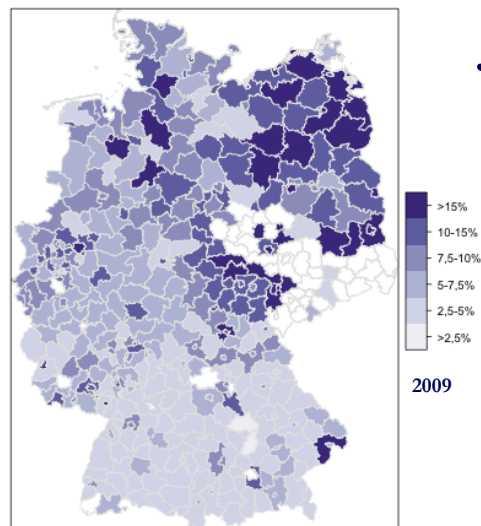
Stem and Leaf Plot of Mean Proportion Cross-Cultural Accuracy of Emotion Recognition (N = 162)

Stem	Leaf
.9	1, 2, 3
.8	5, 6, 6, 7, 7, 9, 9
.8	0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 3, 3, 4, 4, 4,
.7	6, 6, 7, 7, 7, 7, 8, 9, 9, 9, 9
.7	0, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4
.6	5, 5, 5, 5, 5, 7, 8, 8, 9, 9
.6	0, 0, 0, 0, 1, 1, 2, 2, 2, 2, 3, 3
.5	6, 6, 6, 6, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9
.5	0, 0, 0, 0, 0, 1, 2, 2, 3, 3, 3, 4, 4, 4
.4	6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 9
.4	0, 0, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4
.3	5, 6, 7, 8, 8, 9, 9
.3	0, 1, 2, 4,
.2	6, 6, 6, 6, 7, 7, 8
.2	1, 3, 4, 4
.1	8
.1	1, 2
.0	
.0	0, 3

John W. Tukey: Box-and-Whisker-Plot

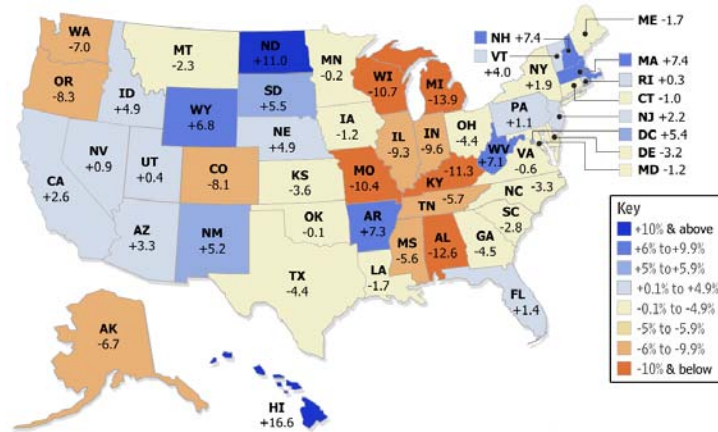


Unemployment in Germany by district



John W. Tukey: Coded Maps

Fallen ...



Fallen ...

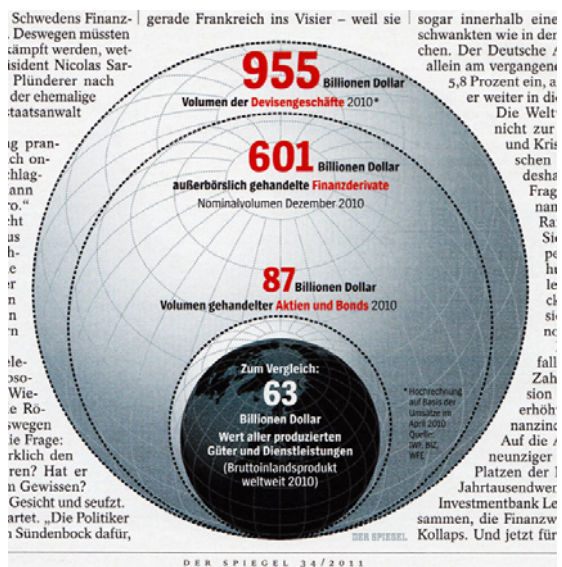
Kubische Wurzel:

63 4.0

87 4.4

601 8.4

955 9.8



Fallen ...

Das sicherste Verkehrsmittel ...

Tote pro Reise	Tote pro Reisezeit	Tote pro Reisestrecke
Bus: 4.3	Bus: 11.1	Flug: 0.05
Zug: 20	Zug: 30	Bus: 0.4
Auto: 40	Flug: 30.8	Zug: 0.6
Laufen: 40	Auto: 130	Auto: 3.1
Flug: 117	Laufen: 220	Velo: 44.6
Velo: 170	Velo: 550	Laufen: 54.2

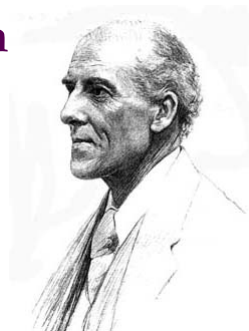
Bett: ∞

Weltweit / 1998

Sollte TDA bivariat sein?

Simpson-Paradoxon

Karl Pearson
1857 - 1936



1892 *The Grammar of Science*

Berkeley gender bias case 1973

Bewerbungen	Angenommen	Abgelehnt
Männer	3714	4728
Frauen	1512	2809

Kreuztabelle / Kontingenztafel

**Erfolgschancen
(Odds)**

Männer 1 : 1.27

Frauen 1 : 1.86

**Relative
Erfolgschancen
(Odds ratio)**

Frauen 1 : 1.46

	Männer		Frauen	
	Bewerbungen	Erfolgreich	Bewerbungen	Erfolgreich
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

Die 6 grössten Departemente von Berkeley

Fazit: Frauen wurden in keiner Weise diskriminiert!

Grund: Das Denken in Odds ist schwierig ...

	Männer		Frauen	
	Bewerbungen	Erfolgreich	Bewerbungen	Erfolgreich
A	10	0%	100	0%
B	10	0%	100	0%
C	10	0%	100	0%
D	10	0%	100	0%
E	10	0%	100	0%
F	950	100%	500	100%

Annals of Eugenics. 7, 1936

Sollte TDA bivariat sein?

Grenzen der Diskriminanzanalyse



Ronald A. Fisher
1890 - 1962

**The Use of Multiple
Measurements in
Taxonomic Problems**

Anderson's Iris data set

50 Schwertlilien der Arten

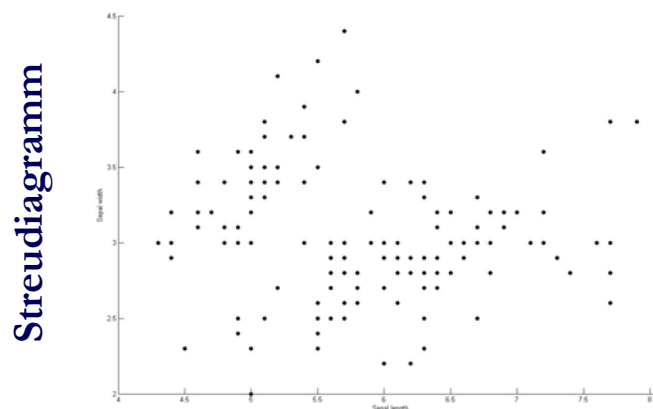
Iris Setosa, Iris Virginica, Iris Versicolor

Merkmale:

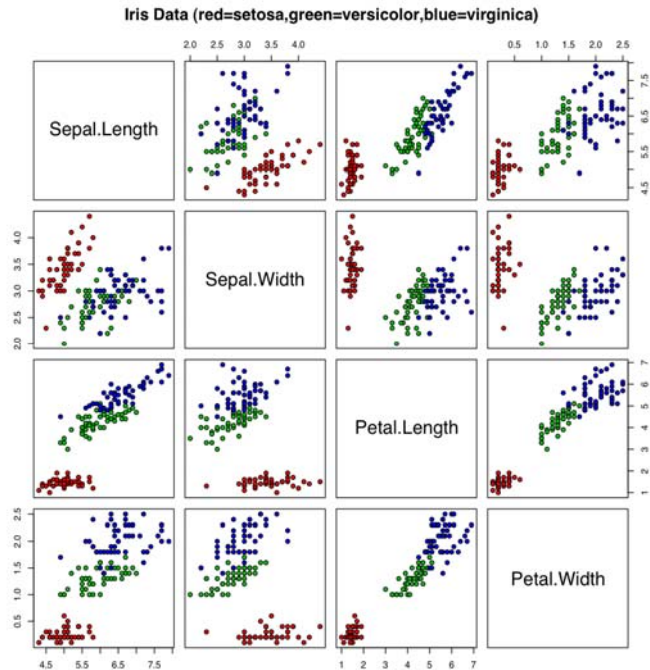
**Länge & Breite von
Kelchblatt - Sepalum
Kronblatt - Petalum.**



**Virginica und Versicolor
sind graphisch nicht diskriminierbar!**

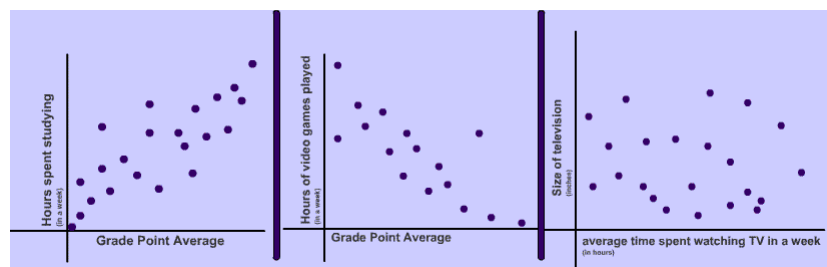


Streudiagramm-Matrix



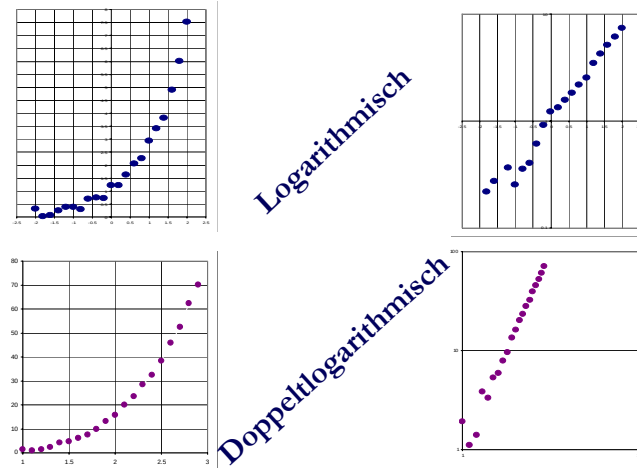
<http://www.mathwarehouse.com/statistics/correlation-coefficient>

Sollte TDA bivariat sein? Zusammenhangsanalyse ...



John W. Tukey war Mitglied der
Charlie Winsor's Society for the
Suppression of Correlation Coefficients ... !!!

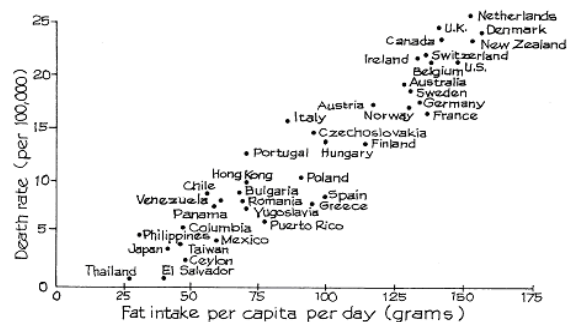
Koordinatentransformationen



www.burns.com/wcbspurcorl.htm

Scheinzusammenhänge

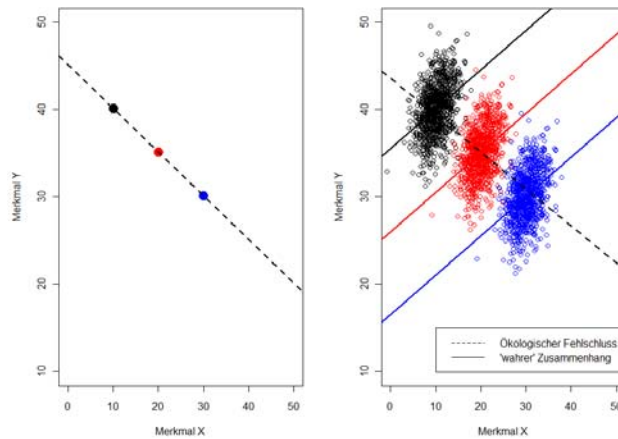
Figure 8. Cancer rates plotted against fat in the diet, for a sample of countries.



Source: K. Carroll, "Experimental evidence of dietary factors and hormone-dependent cancers," *Cancer Research* vol. 35 (1975) p. 3379. Copyright by Cancer Research. Reproduced by permission.

Korrelation vs. Kausalität

Der Ökologische Fehlschluss



Statistik der Kennzahlen ...

Wie betrüge ich mit Statistik ...



Cyril L. Burt
1883 - 1971

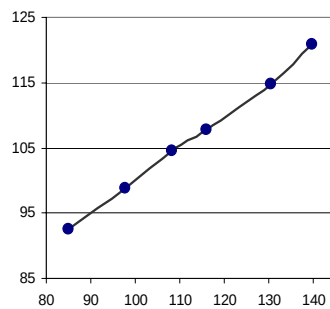
Table 1
Intelligence Quotient (IQ) Data Provided by Burt (1961)

Social class	Adult mean IQ	Child mean IQ
Higher professional	139.7	120.8
Lower professional	130.6	114.7
Clerical	115.9	107.8
Skilled	108.2	104.6
Semiskilled	97.8	98.9
Unskilled	84.9	92.6

Note. From *The Elements of Graphing Data* (p. 97), by W. S. Cleveland, 1985, Monterey, CA: Wadsworth. Copyright 1985 by W. S. Cleveland. Reprinted with permission.

John W. Tukey:
"The greatest value of a picture is when it forces us to notice what we never expected to see."

flowingdata.com/2007/09/14/wise-words-from-john-tukey/



John W. Tukey zur Linearen Ausgleichsrechnung ...

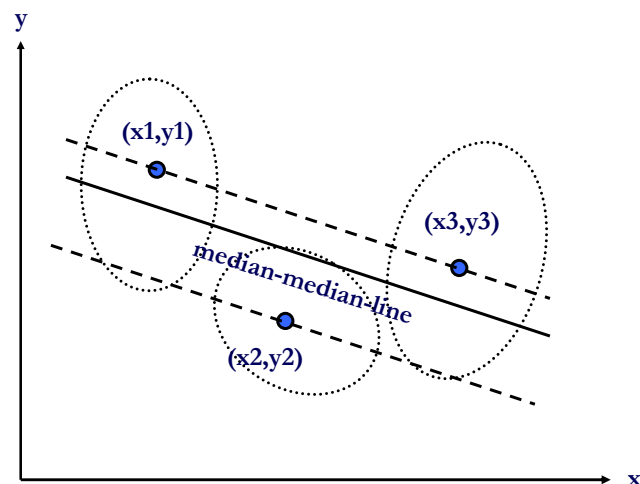
- Daten in Abhängigkeit von x in drei Klassen einteilen.
- x_1, x_2, x_3 Mediane der x-Werte in jeder Klasse
- y_1, y_2, y_3 Mediane der y-Werte in jeder Klasse
- Die Ausgleichsgerade ist $y = a x + b$ mit

$$a = (y_3 - y_1) / (x_3 - x_1)$$

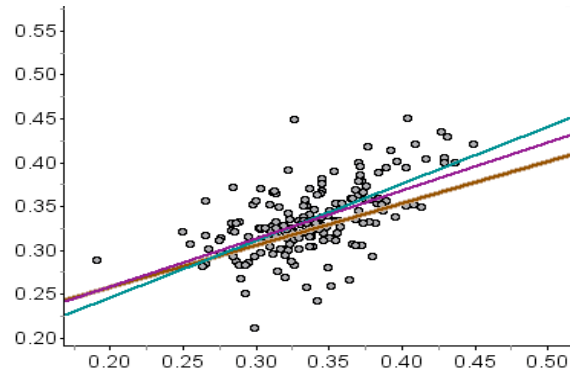
$$b = (y_1 + y_2 + y_3 - a (x_1 + x_2 + x_3)) / 3$$

.

Ideeskizze:



<http://exploratorydataanalysis.blogspot.com/>



Die Erste wurde von Auge bestimmt;
die Zweite durch Gaussche lineare Regression;
die Dritte durch das Verfahren von Tukey;
Welche ist welche?

TDA - Pros & Cons

(ohne Anspruch auf Vollständigkeit oder Wissenschaftlichkeit)

Was spricht dagegen:

- 20 Lektionen zu viel ...
- 20 Lektionen zu wenig ...
- Zu ambitioniert für BM-Lernende ...
- Zu anspruchsvoll für BM-Didaktik ...

Replik:

All dies gilt nur für das klassische Modell!

Was spricht dafür:

- Gesellschaftliche Relevanz ...
- Berufliche Relevanz ...
- Zubringerdienst FH ...
- Viele Chancen attraktiver Didaktik ...
- Erfolgserlebnis auch für Mathe-Muffels ...
- Ideal für projektbasierten Unterricht ...
- Ideal für interdisziplinäres Arbeiten ...

Replik:

Falls TDA nicht zu Betty Bossi entartet!

Mein liebstes Argument:

Elisabeth
Noelle-Neumann
1916 - 2010



"Statistik ist für mich das
Informationsmittel der Mündigen.
Wer mit ihr umgehen kann,
kann weniger leicht manipuliert werden."