

So lügt man mit Statistik

Anita Maas

06. August 2004

Was ist Statistik?

- ❑ **Statistik** ist die Gesamtheit aller Methoden, die für die Untersuchung einer Vielzahl von Einzeltatsachen verwendet werden.
- ❑ Sie ist angewandte Mathematik.
- ❑ Als eine Statistik bezeichnet man auch das Zahlenmaterial, das der zusammenfassenden Beschreibung von Einzeltatsachen dient.

❑ **deskriptive Statistik:**

(beschreibende, empirische Statistik)

dient der Verdichtung der Daten zu Tabellen, graphischen Darstellungen und Kennzahlen.

❑ **induktive Statistik:**

(schliessende, mathematische Statistik)

aus den vorliegenden Daten schliesst man

(mit Hilfe von Wahrscheinlichkeitstheorie) auf eine Struktur.

Mögliche Ursachen der statistischen Fehler

- Fehlerhafter Sprachgebrauch
- Fehlerhafte Bilddarstellung
- Unklarheit der statistischen Begriffe

(Umfrage von Prof. Dr. E. Noelle-Neumann:)

1. Frage:

“Finden Sie, dass in einem Betrieb alle Arbeiter in der Gewerkschaft sein sollten?”

Resultat:

- Dafür: 44 %
- Dagegen: 20 %
- Unentschieden: 36 %

2. Frage:

wie 1., jedoch mit dem Zusatz:

“...oder muss man es jedem einzelnen überlassen, ob er in der Gewerkschaft sein will oder nicht?”

Ergebnis:

- Dafür: 24 %
- Dagegen: 70 %
- Unentschieden: 6 %

Fehlerhafter Sprachgebrauch: Definitionen

Laut einem “Bericht zur Bekämpfung des Analphabetismus in der BRD”
gibt es heute in Deutschland

ca **500 000** Analphabeten ...

Zu Kaiser Wilhelms Zeiten gab es

weniger als 10 000...

Definition zu Kaiser Wilhelms Zeiten:

Analphabet war jeder, der seinen Namen nicht schreiben konnte.

Definition heute:

Analphabet ist eine Person, die sich nicht beteiligen kann an all den zielgerichteten Aktivitäten ihrer Gruppe und ihrer Gemeinschaft, bei denen Lesen, Schreiben und Rechnen erforderlich ist und an der weiteren Nutzung dieser Kulturtechniken für ihre weiteren Entwicklung und die der Gesellschaft”.

Fehlerhafter Sprachgebrauch: Synthetischer Superlativ

Spiel	Turnier	Erdteil	Platz	Ausgang
1	1	Australien	Rasen	verloren
2	2	Amerika	Hartplatz	verloren
3	3	Europa	Rasen	gewonnen
4	3	Europa	Rasen	gewonnen
5	3	Europa	Rasen	gewonnen
6	4	Europa	Hartplatz	verloren
7	5	Europa	Rasen	gewonnen
8	5	Europa	Rasen	gewonnen
9	5	Europa	Rasen	gewonnen
10	5	Europa	Rasen	gewonnen

Figure 1: Ergebnisse des Tennisspielers X

Mögliche Interpretation dieser Tabelle:

- Spieler X im Aufwärtstrend! Mehr als 75 Prozent der letzten acht Spiele gewonnen.
- Phänomenal! X in Europa auf Rasen ohne Niederlage.

Fehlerhafte Bilddarstellung: betrügerische Basis

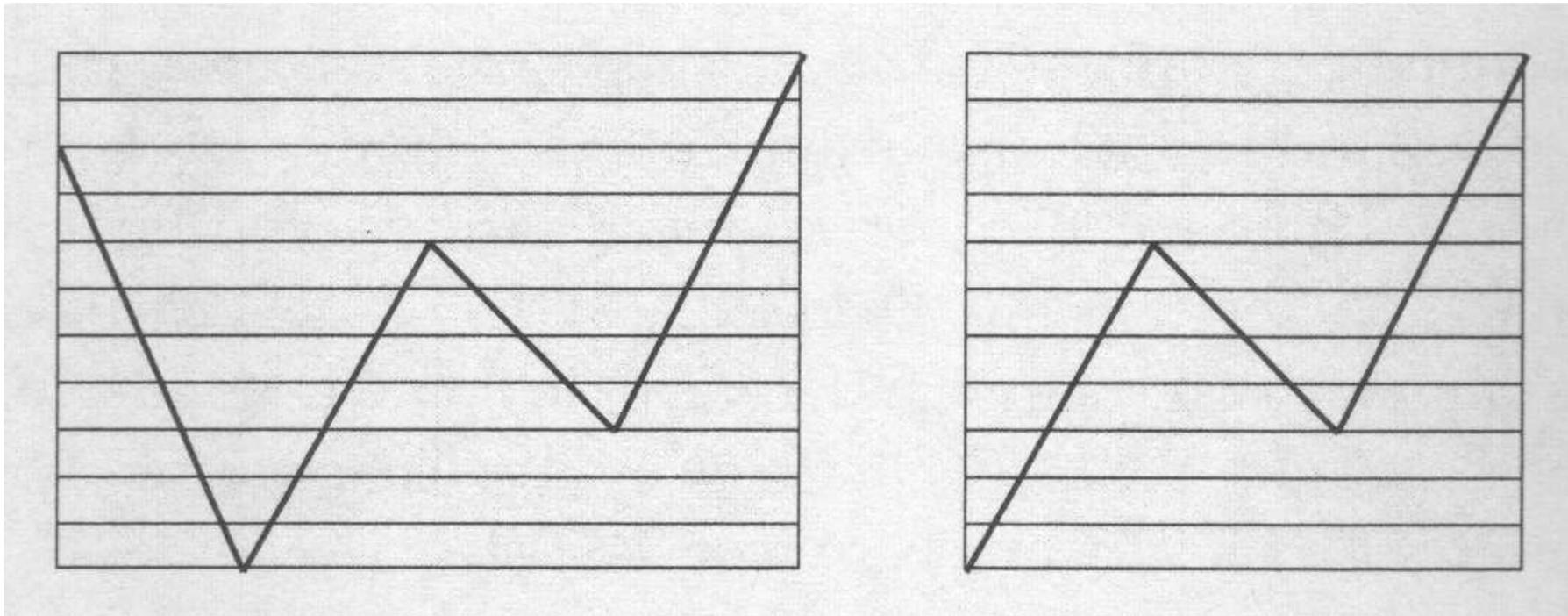


Figure 2: wo fängt die Kurve an?

Fehlerhafte Bilddarstellung: Strecken der Kurve

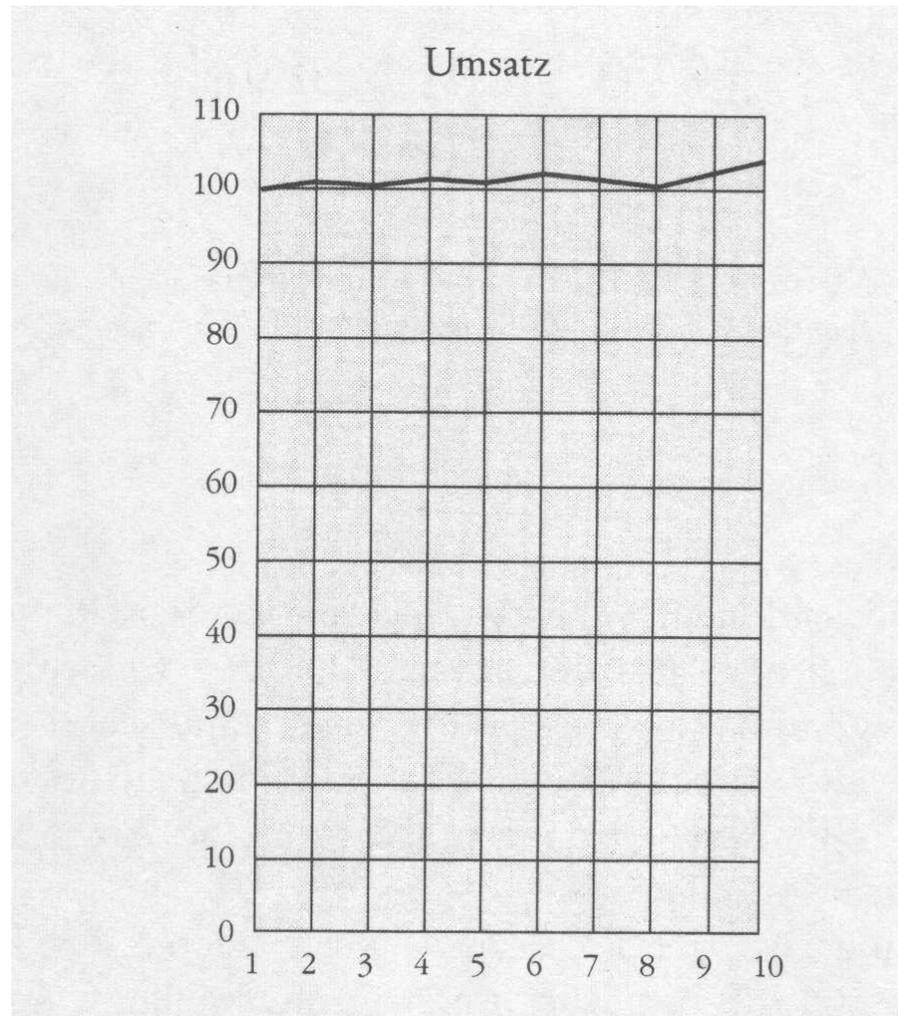


Figure 3: richtige Darstellung

Fehlerhafte Bilddarstellung: Strecken der Kurve

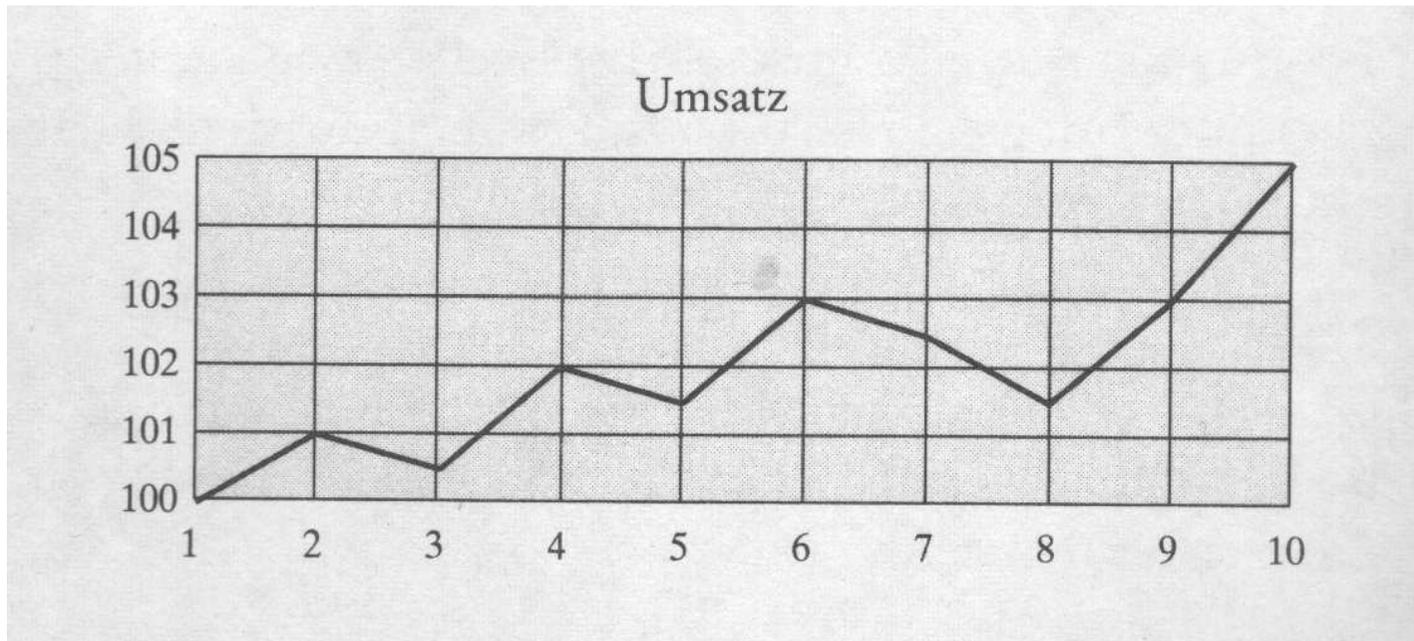


Figure 4: gestreckte Darstellung

Fehlerhafte Bilddarstellung: Balkendiagramm

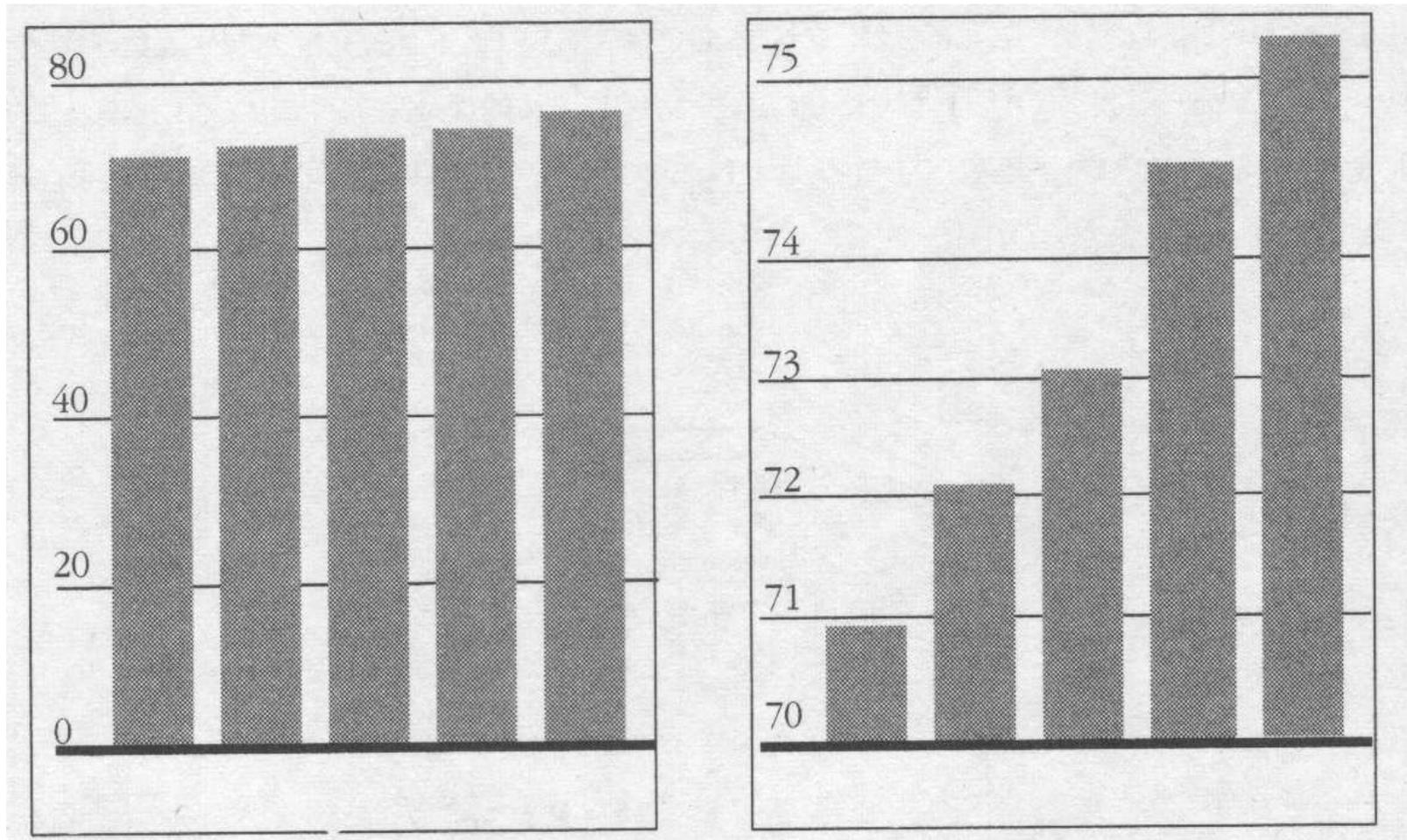


Figure 5: Strecken der Balken

Fehlerhafte Bilddarstellung: Piktogramm

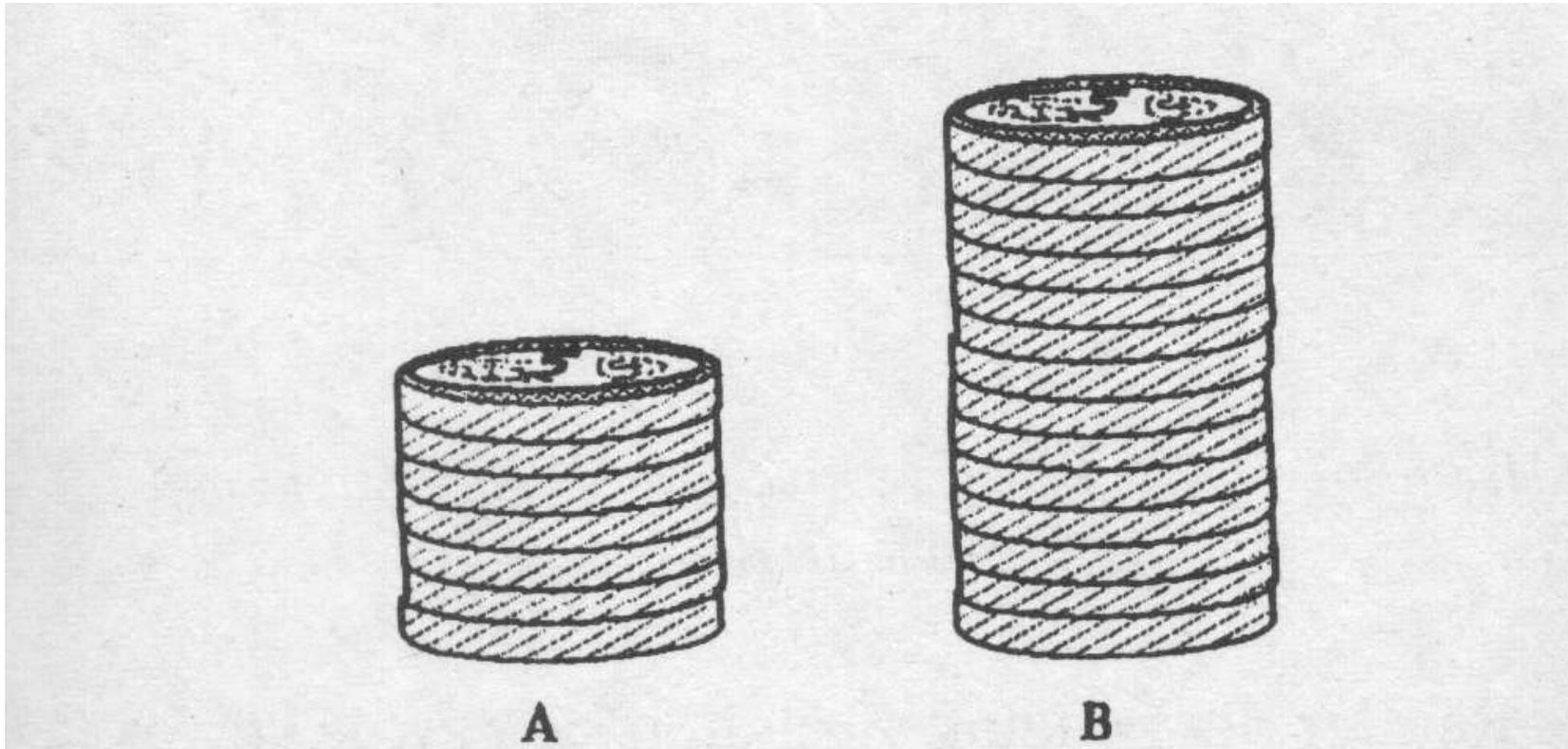


Figure 6: richtige Darstellung

Fehlerhafte Bilddarstellung: Piktogramm



Figure 7: falsche Darstellung

Fehlerhafte Bilddarstellung: Karten

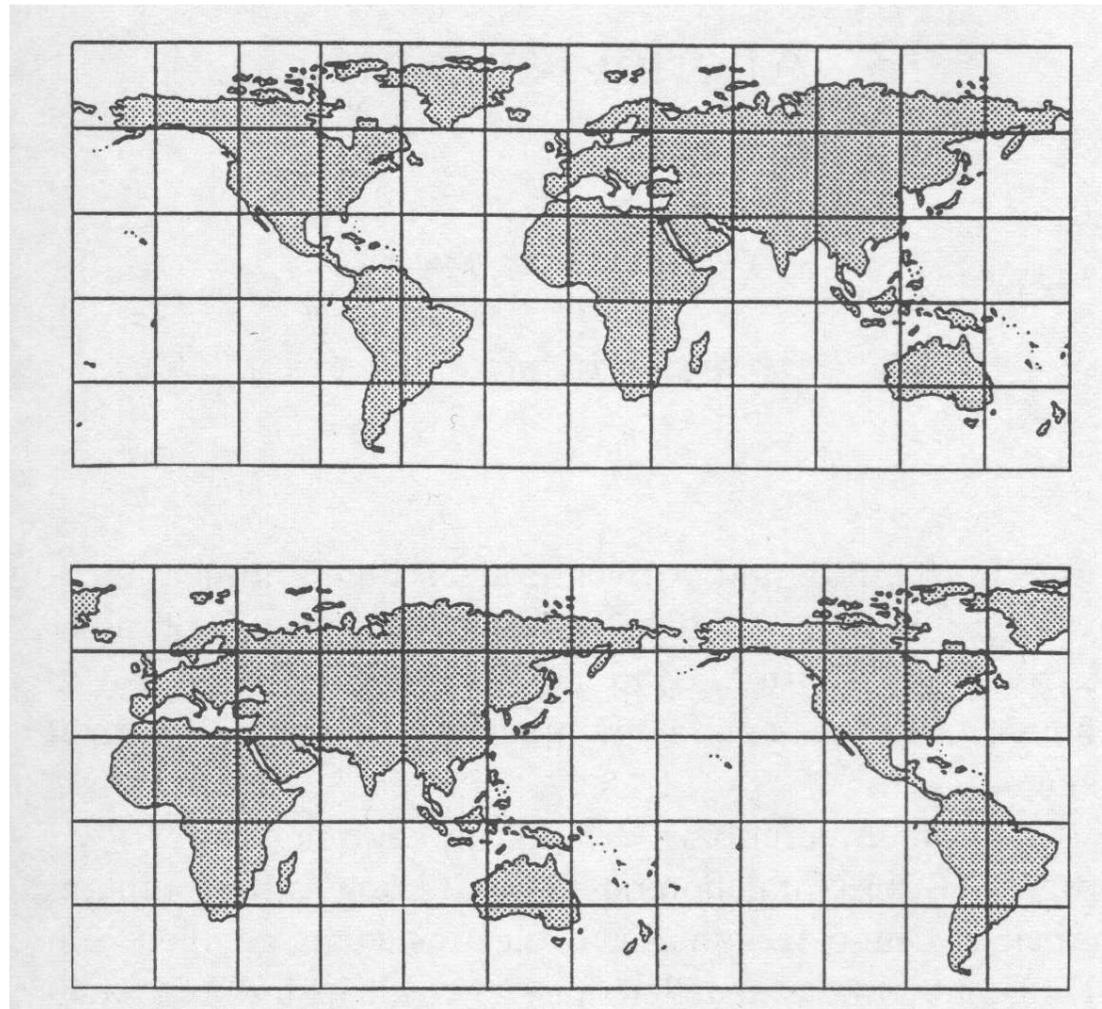


Figure 8: wo ist das Zentrum der Welt?

Fehlerhafte Bilddarstellung: Karten

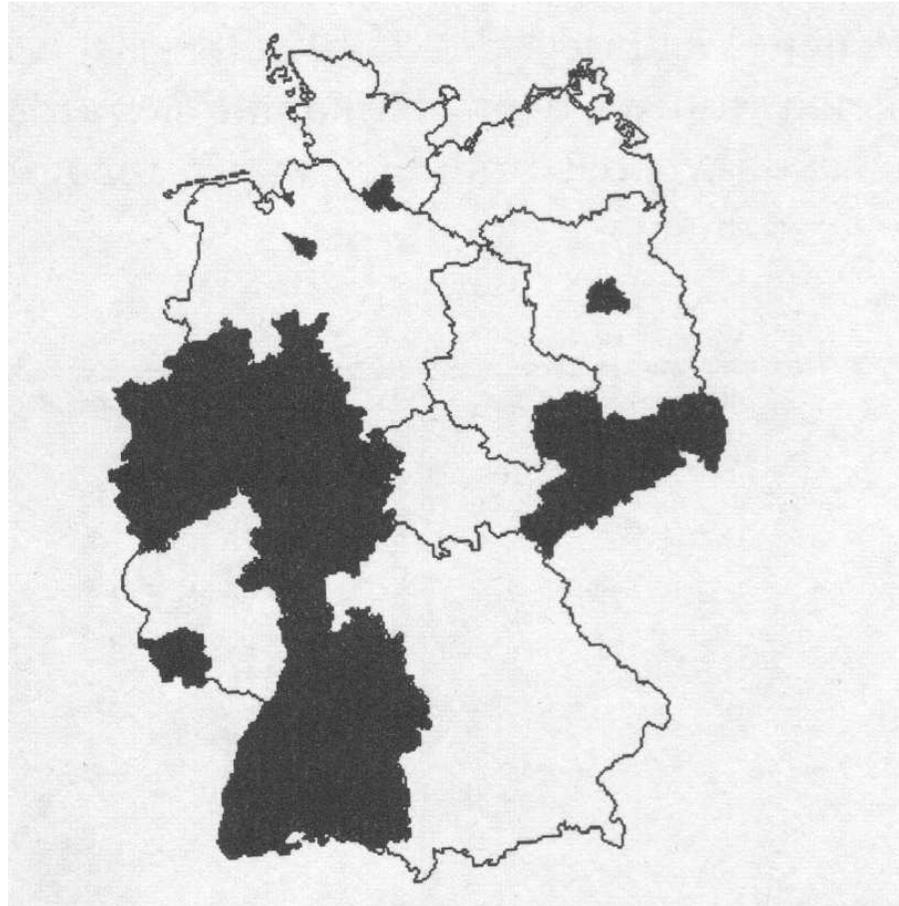


Figure 9: Bevölkerungsdichte in Deutschland

Fehlerhafte Bilddarstellung: Karten

Folgendes könnte man aus dieser Karte (z. B.) ableiten:

1. in der BRD herrscht Unterbevölkerung: die meisten Länder sind “weiß”, also hier unter 200 Menschen pro Quadratkilometer...
2. in den meisten Ländern Deutschlands gibt es kaum Kriminalität, “nur” in den Ballungszentren...
3. BRD wäre mit Rettungsdiensten unterversorgt: in den meisten Ländern muss man länger auf einen Rettungswagen warten...

Unklarheit der statistischen Begriffe: Interpretation der Prozente

- ❑ Prozent (von lat. pro = für; centum = hundert) steht für einen in Hundertstel ausgedrückten Bruchteil
- ❑ Prozente können Informationen liefern
- ❑ Prozente können Informationen schlucken
- ❑ Prozente sehen oft - vor allem bei kleinen Zahlen - besser aus

Unklarheit der statistischen Begriffe: Interpretation der Prozente

Gegeben Umsätze: 100 101 102,5

- ❑ Umsatz-Zuwachs: zuerst um 1 %, dann um $1,5/101 = 1,49$ %
- ❑ und in Wachstumsraten von Wachstumsraten: 49 % !

Unklarheit der statistischen Begriffe: Median kontra Mittelwert

- ❑ Sortiert man eine Reihe von Messwerten der Größe nach, so ist der Wert, der in der Mitte dieser Reihe liegt, der **Median**.

- ❑ Im Gegensatz zum arithmetischen **Mittelwert**, auch Durchschnitt genannt, verändert sich der Median durch einzelne Extremwerte kaum.

Unklarheit der statistischen Begriffe: Median kontra Mittelwert

- So ist der **Median** der Zahlenreihen

1, 2, 3, 4, 5 und **1, 2, 3, 4, 100**

jeweils **3**.

- Der **Mittelwert** ist im ersten Fall ebenfalls **3**,

bei der zweiten Reihe verschiebt der Ausreißer 100 den Mittelwert

auf **22**.

Unklarheit der statistischen Begriffe: Median kontra Mittelwert

Gegeben: eine Firma, in der Frauen 20 Euro und Männer 30 Euro die Stunde verdienen. Wie hoch ist der Durchschnittsgehalt?

- ❑ das gewöhnliche arithmetische Mittel

Durchschnittslohn = **25 Euro**

(korrekt, falls Anzahl d. Frauen = Anzahl d. Männer)

- ❑ das gewogene arithmetische Mittel

z. B. bei 7 Frauen und 3 Männern:

$$0,7 * 20 + 0,3 * 30 = \mathbf{23 \text{ Euro}}$$

Kausalität bezeichnet die Beziehung zwischen Ursachen und Wirkungen.

Die **Korrelation** ist eine Beziehung zwischen zwei oder mehr statistischen Variablen. Es gibt positive und negative Korrelationen.

Die Korrelation beschreibt **nicht unbedingt** eine Ursache-Wirkungs-Beziehung.

Unklarheit der statistischen Begriffe: Kausalität und Korrelation

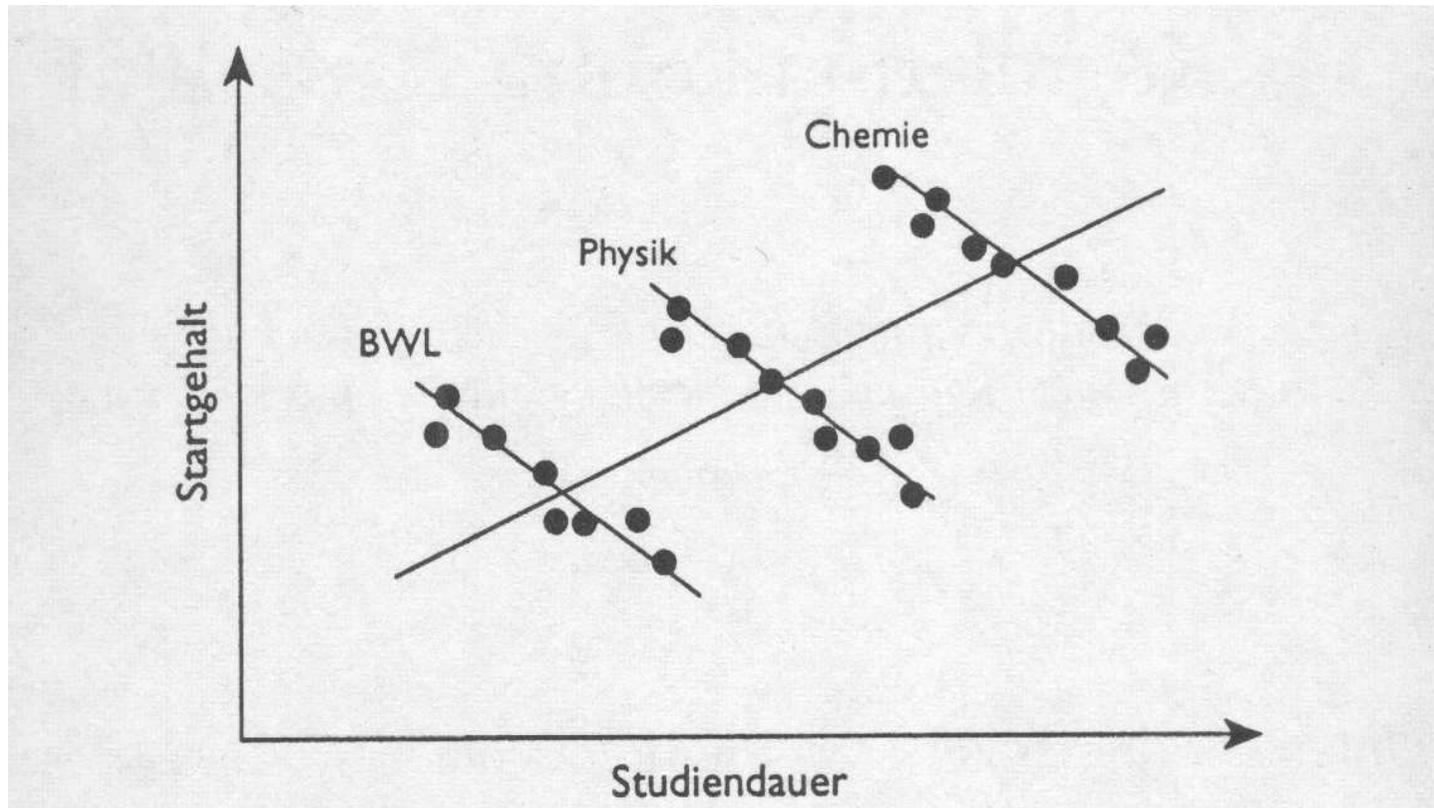


Figure 10: Ein langes Studium zahlt sich aus?

Wahrscheinlichkeit ist ein Maß

1. für die Unsicherheit zukünftiger Ereignisse oder zweifelhafter Aussagen
2. für die relative Häufigkeit des Auftretens von Ereignissen bei Auswahl aus mehreren Möglichkeiten

Unklarheit der statistischen Begriffe: Wahrscheinlichkeit

□ bedingte: $P(A|B) = P(A \cap B)/P(B)$

□ nicht bedingte: $P(A|B) = P(A)$

□ Bayes-Theorem: $P(A|B) = (P(B|A) * P(A))/P(B)$

Die **Stichprobe** ist eine mit einer bestimmten Methode erzeugte **Teilmenge** einer als Grundgesamtheit bezeichneten Menge.

Arten der Stichprobe:

- Nach dem Zurücklegen bereits gezogener Elemente:
 - Ziehen mit Zurücklegen
 - Ziehen ohne Zurücklegen

□ Nach der Auswahl von Elementen für die Stichprobe:

➤ **Zufalls-Stichprobe**

Jedes Element kann mit einer ganz bestimmten Wahrscheinlichkeit in die Stichprobe eingehen.

➤ **Willkürliche Stichprobe**

Elemente werden mehr oder weniger willkürlich in die Stichprobe aufgenommen. Angabe der Wahrscheinlichkeit ist nicht möglich.

Unklarheit der statistischen Begriffe: Signifikanz der Stichprobe

- ❑ Die Größe und Qualität der Stichprobe orientieren sich in erster Linie an der Problemstellung der Untersuchung.
- ❑ Die Größe der Stichprobe ist so zu bemessen, dass sie die vertretbare statistische Fehlertoleranz berücksichtigt.

Unklarheit der statistischen Begriffe: Signifikanz der Stichprobe

Flughafen	AIDS-Fälle
Berlin	0
Frankfurt	2
Hannover	0
Leipzig	1
München	0
Hamburg	4
Düsseldorf	0

Figure 11: Fluglärm erzeugt AIDS...

“Beweis”

- ❑ wir lassen alle Plätze außer Hamburg weg
(eine nicht zufällige Stichprobe)
- ❑ wir tun so, als wären die 10 Personen zufällig ausgewählt
- ❑ nun haben wir 4 von 10 Personen, die mit AIDS infiziert sind
(40 %)
- ❑ es ist höher als der Durchschnitt (10 %), wir lehnen den Zufall ab

Unklarheit der statistischen Begriffe: Signifikanz der Stichprobe

Der Trick:

Für die Ausgangshypothese wird möglichst ungünstige Teilstichprobe genommen.

Jedoch die Wahrscheinlichkeit wird so berechnet, als wäre die Stichprobe zufällig gewählt...

Wo kann man bei Statistik Fehler machen?

- beim Sammeln der Daten
- bei der Auswertung
- bei der Darstellung
- bei der Interpretation

Fazit

- mit Statistik haben wir oft zu tun
- es gibt viele Fehlermöglichkeiten
- Statistik ist schwierig
- sie ist nicht intuitiv
- Statistik wird oft unterschätzt
- falls man sie versteht, kann man Betrug erkennen
- oder betreiben...

- ❑ Alle Definitionen aus “Wikipedia - die freie Enzyklopädie”
<http://de.wikipedia.org>
- ❑ Alle Beispiele aus “So lügt man mit Statistik” Walter Krämer;
Piper Verlag, München, 2002