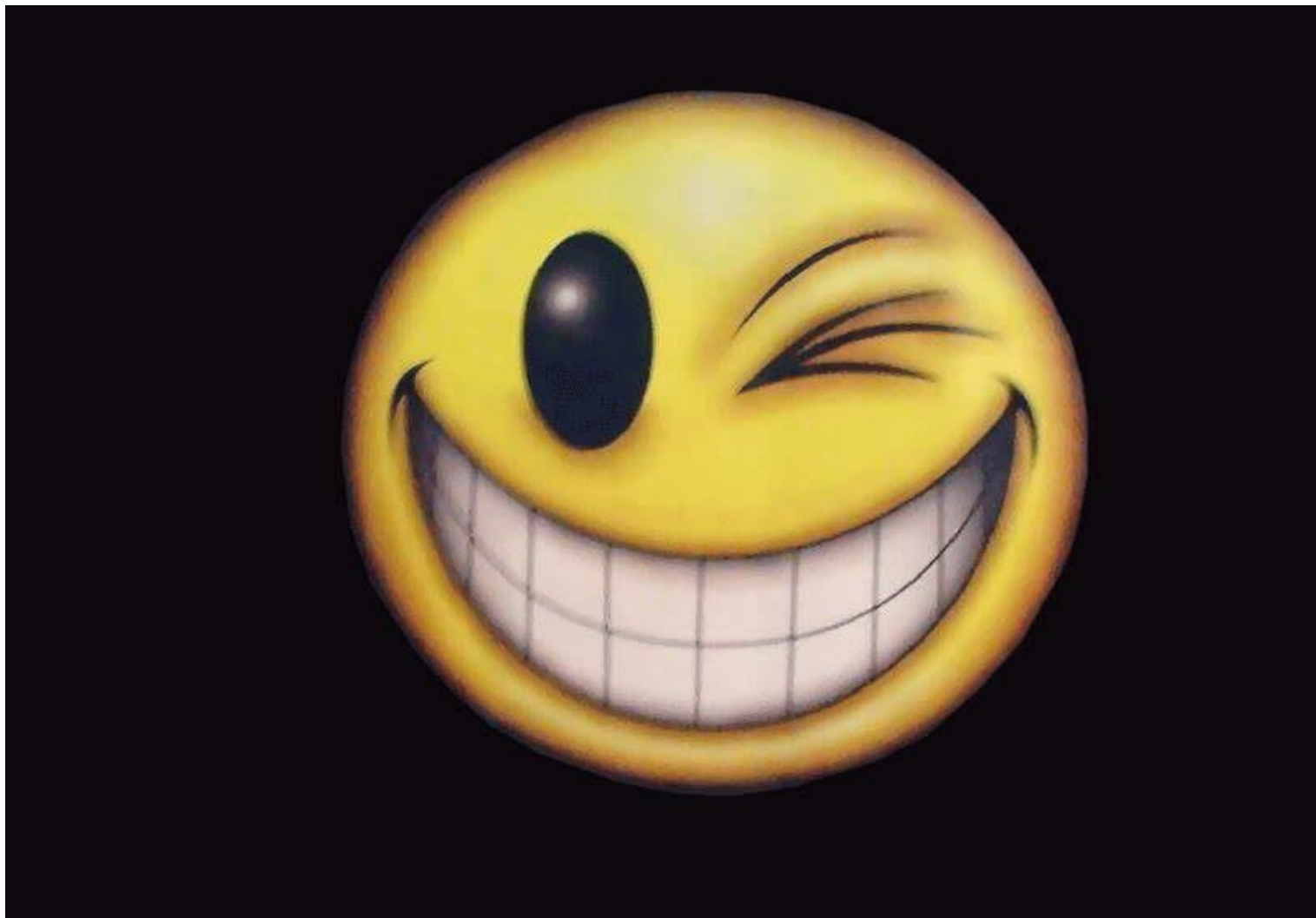


Deskriptive Statistik

Statistik (Biol./Pharm./HST) – Herbst 2013



Halbzeit !



Lernziele

Gemacht

- Grundlagen:
 - Wa.modell
 - Unabh.
 - Bed. Wa.
 - diskrete ZV
 - Verteilungen
 - Kennzahlen
- Am Bsp vom Binomialtest:
 - Test (Fehler 1. Art, Macht)
 - p-Wert
 - Vertrauensintervall

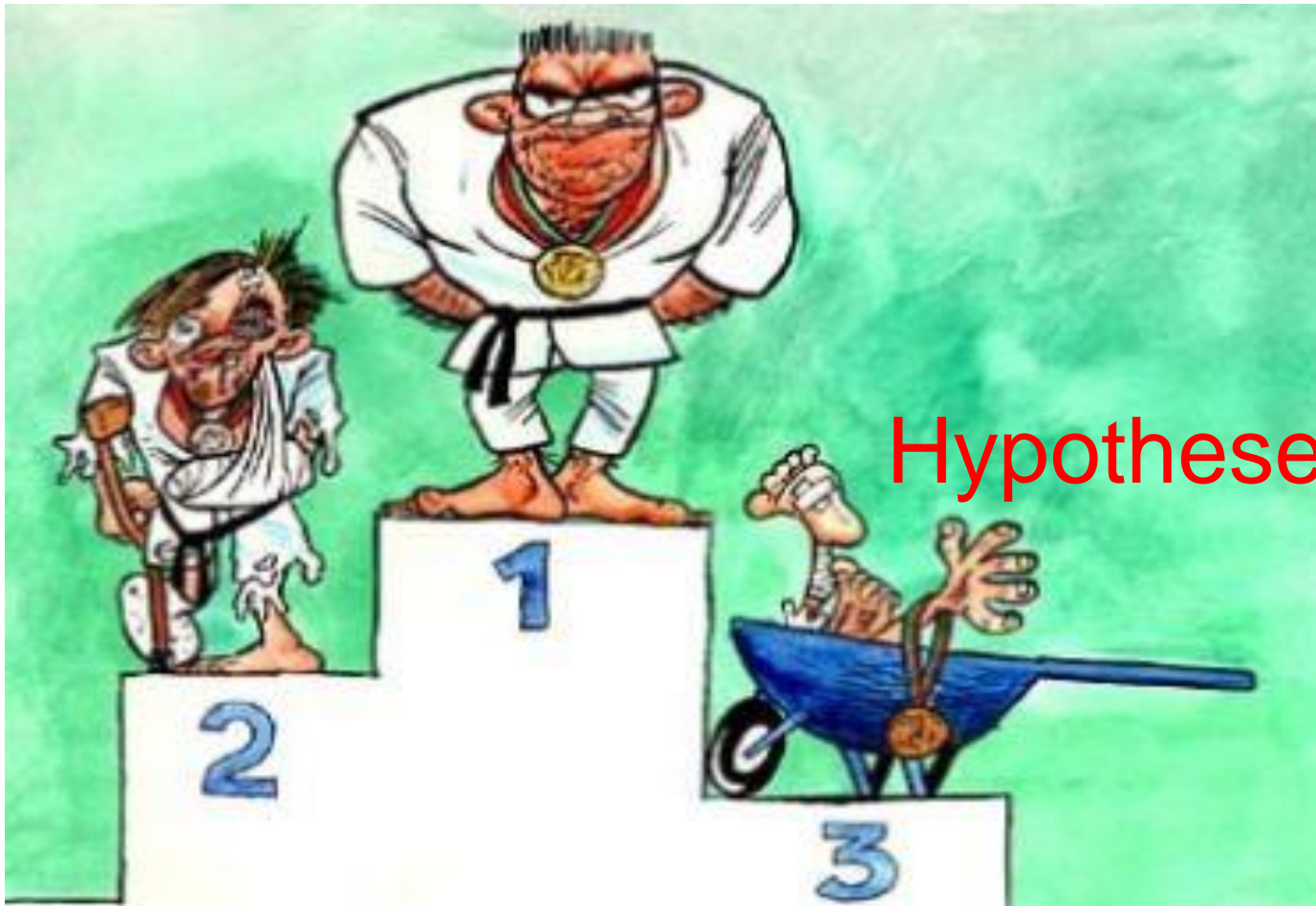
To Do

- Deskriptive Statistik:
 - Zahlen
 - Graphiken
- Am Bsp vom t-Test:
 - Test (Fehler 1. Art, Macht)
 - p-Wert
 - Vertrauensintervall
- Lineare Regression
 - einfach
 - multiple

Statistische Inferenz: Drei Methoden



Statistische Inferenz: Drei Methoden



Hypothesentest

Hypothesentest

(Input: Sign.niveau, Daten)

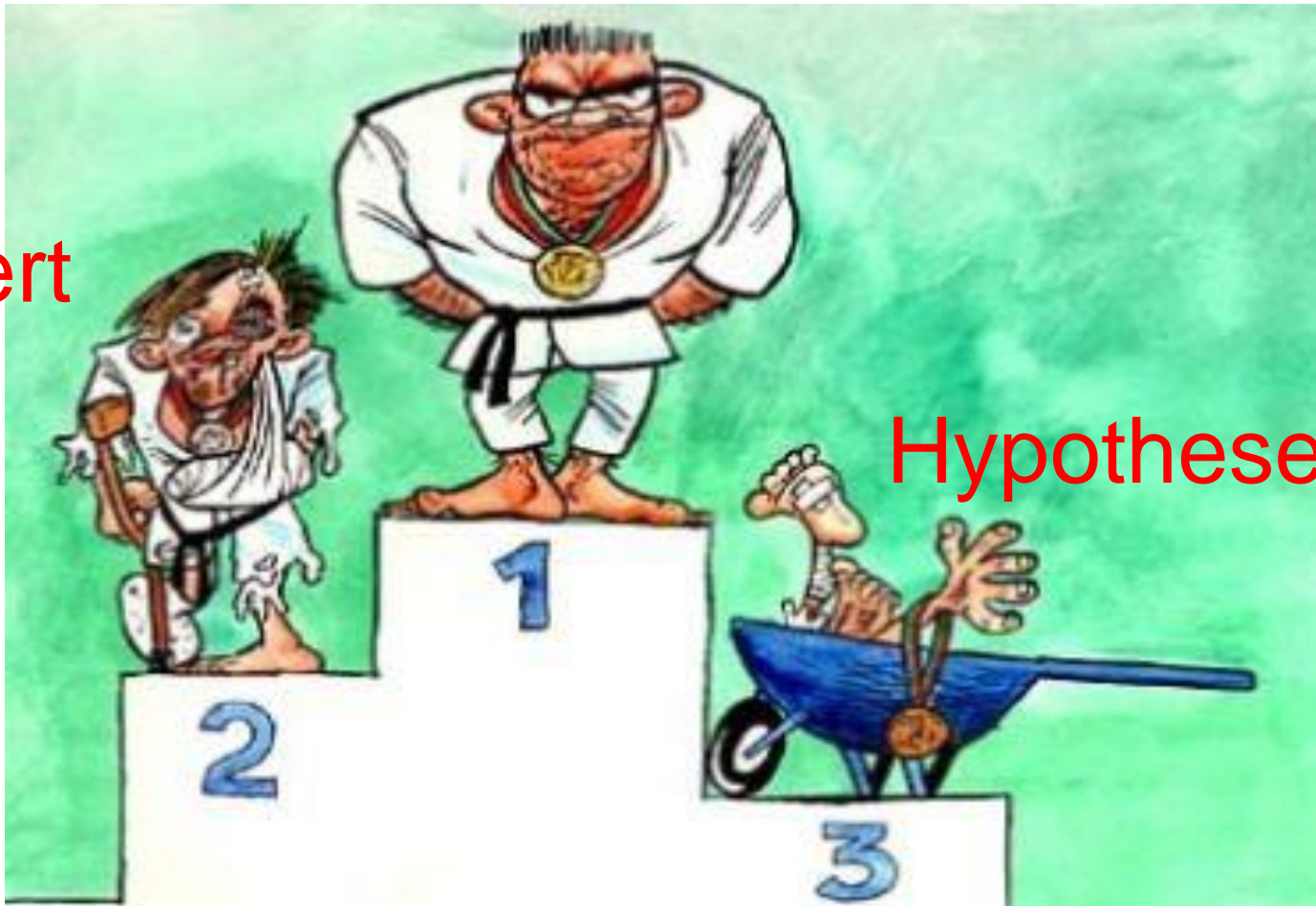
1. Modell
2. H_0 , H_A
3. Teststatistik etc.
4. Sign.niveau
5. Verwerfungsbereich etc.
6. Testentscheid

+ Klares Prozedere
+ Klare Aussage über Fehler 1./2. Art

- Wie deutlich wurde verworfen?
- Wie gross ist der wahre Parameter?

Statistische Inferenz: Drei Methoden

p-Wert



Hypothesentest

p-Wert

(Input: Daten)

- Kleinstes Sign.niveau, bei dem Test gerade noch verwirft
- Gegeben H_0 ; wie plausibel ist die Beobachtung oder etwas noch extremeres?

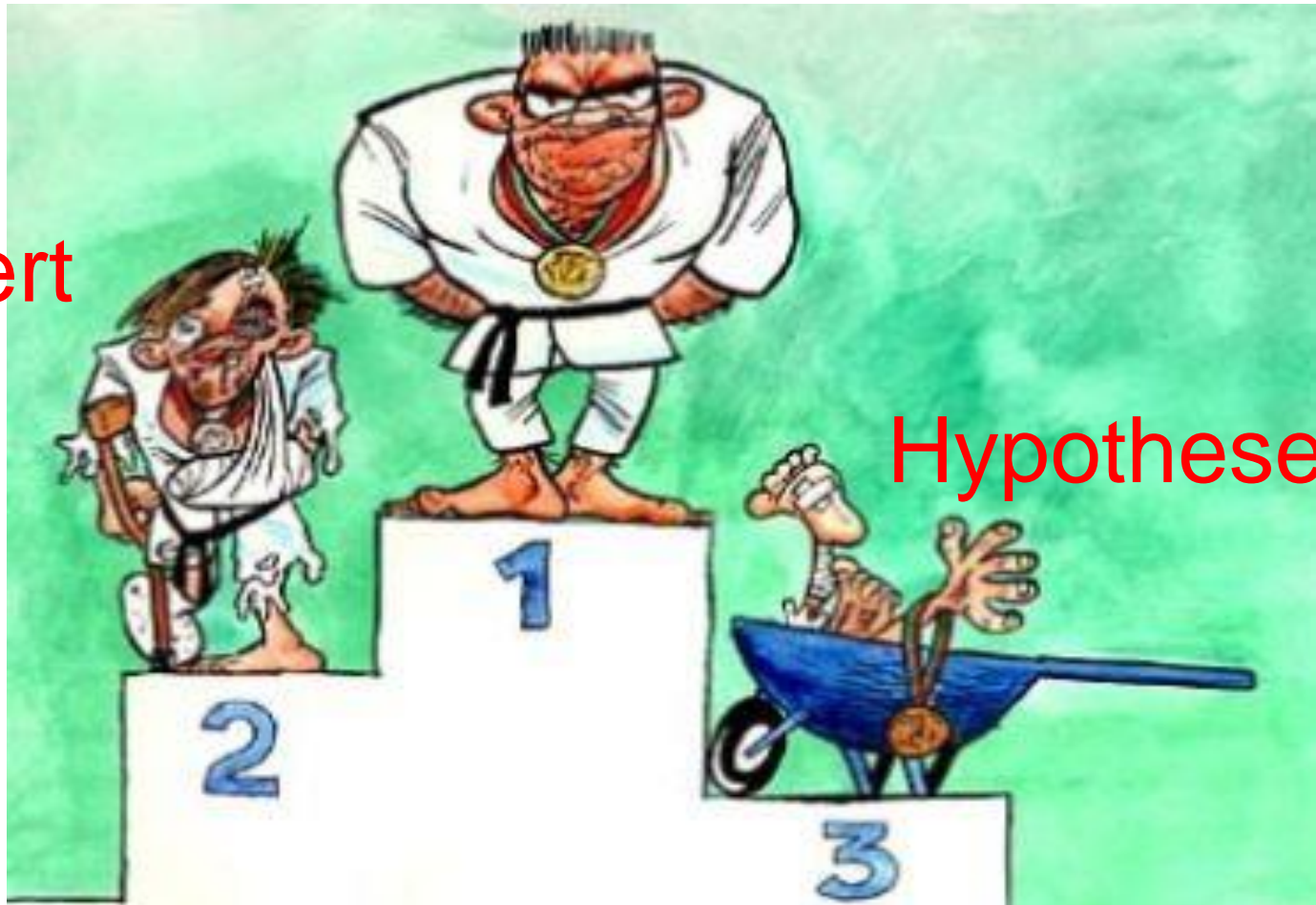
+ Klar, ob und wie deutlich verworfen wird

- Keine klare Aussage über Fehler 1./2. Art
- Wie gross ist der wahre Parameter?

Statistische Inferenz: Drei Methoden

Vertrauensintervall

p-Wert

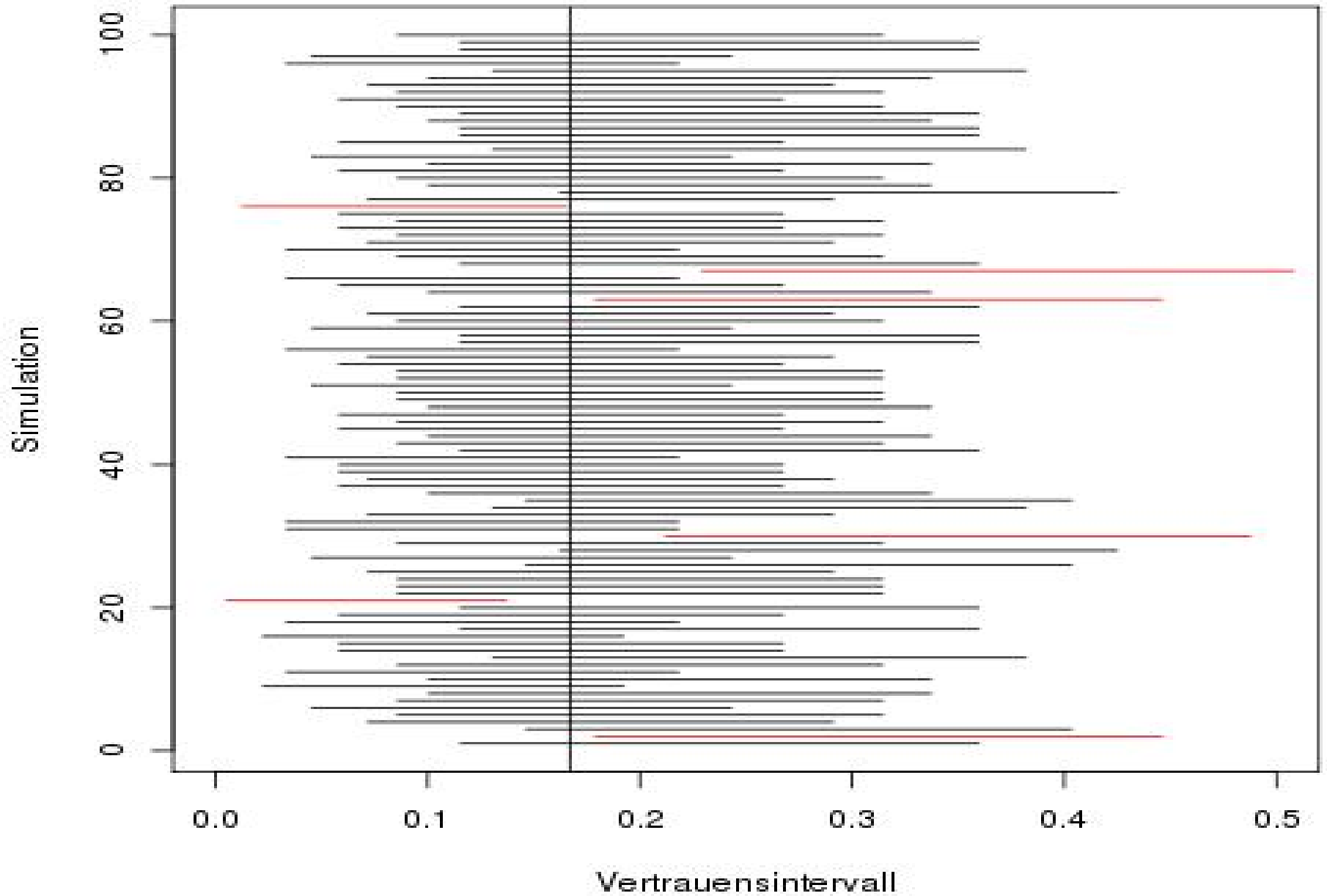


Hypothesentest

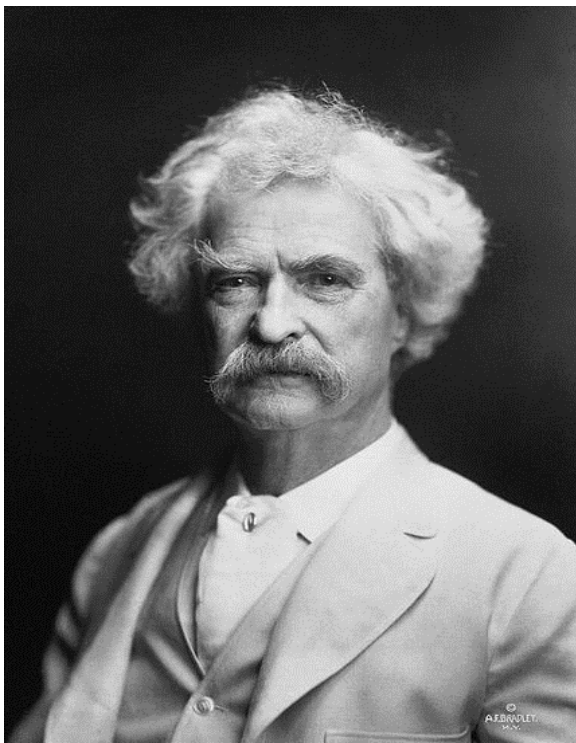
95% - Vertrauensintervall (Input: Sign.niveau, Daten)

- Enthält wahren Parameter mit Wa. 95%
 - Enthält alle Parameter, bei denen ein Test auf dem 5% Niveau NICHT verwirft (= plausible Werte)
- + Klar, ob und wie deutlich verworfen wird
 - + Klar, wie gross der wahre Parameter etwa ist
 - Keine klare Aussage über Fehler 1./2. Art

Coverage = 0.94



Mark Twain zitiert britischen Premierminister Benjamin Disraeli

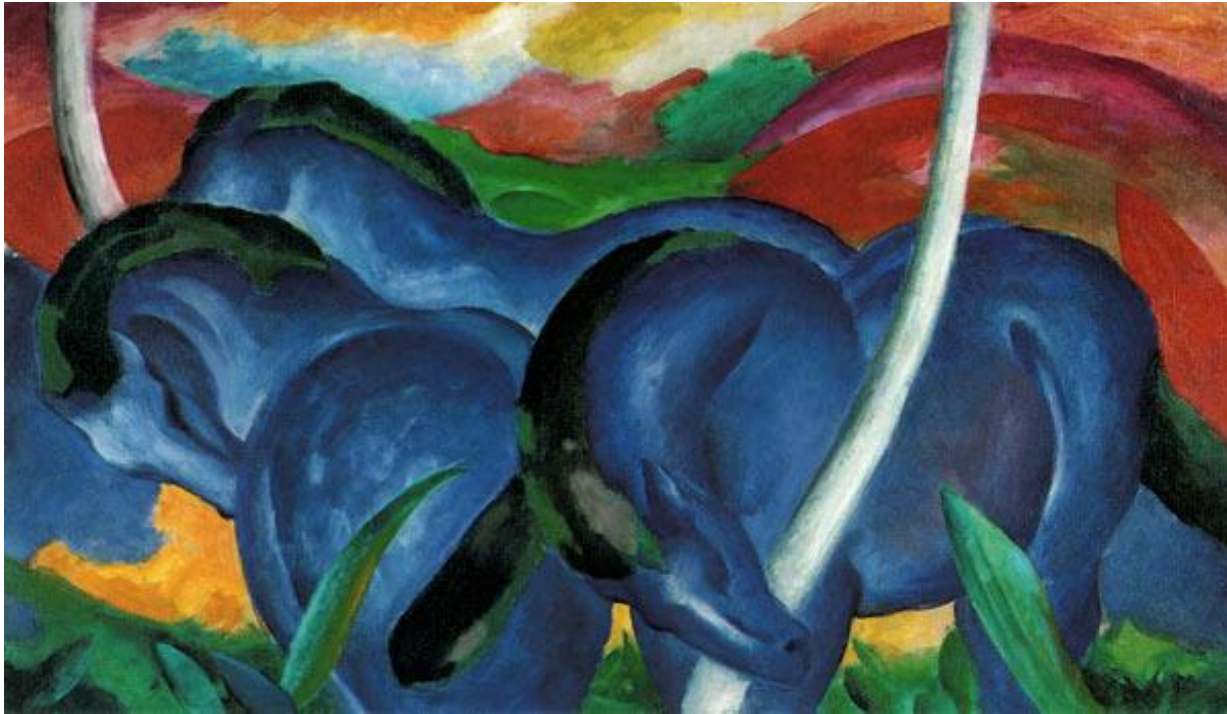


„There are three kinds of lies:
lies,
damned lies,
and statistics.“

Frederick Mosteller
Gründer Statistik-Department, Harvard

„It's easy to lie with statistics;
it is **easier to lie without them.**“

Beschreibe mit einem Wort!



Pferde ?

Marc ?

blau ?

Ohne Reduktion ist
kein Überblick möglich

Bei jeder Datenreduktion
kann wichtige Information
verloren gehen

Simpson's Paradox

(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

	Bewerber	Aufgenommen
Männer	8442	44%
Frauen	4321	35%

Werden Frauen benachteiligt?

Simpson's Paradox

(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

Dept	Männer		Frauen	
	Bew.	Akz.	Bew.	Akz.
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%

Simpson's Paradox

(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

Dept	Männer		Frauen	
	Bew.	Akz.	Bew.	Akz.
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%

Simpson's Paradox

(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

Dept	Männer		Frauen	
	Bew.	Akz.	Bew.	Akz.
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%

Nein: Frauen bewerben sich mehr bei „schwierigen“ Departments!

Arrow's Paradox: Demokratie

Einzelmeinungen



→ Wahlergebnis

Arrow's Paradox: A, B oder C?

- 17 Personen müssen zw. A, B oder C entscheiden.
- Präferenzen:
 - 5 $A > B > C$
 - 2 $A > C > B$
 - 2 $C > A > B$
 - 4 $C > B > A$
 - 4 $B > C > A$
- Mehrheit (1,0,0):
A gewinnt
- Platz 1 oder 2 (1,1,0):
B gewinnt
- Borda Count (2,1,0):
C gewinnt
- Paarweise:
 $A > B$, $B > C$, $C > A$
Unentschieden



Wahlmethode entscheidend !

Finde gute Wahlmethode: Anforderungen

- **Transitiv:**
Wenn $A > B$, $B > C$ dann $A > C$
- **Pareto effizient:**
Wenn $A > B$ für *alle* Wähler, dann muss $A > B$ auch im Wahlergebnis sein
- **Independence of Irrelevant Alternative (IIA):**
Wenn $A > B > C$ und C stirbt, soll $A > B$ bleiben

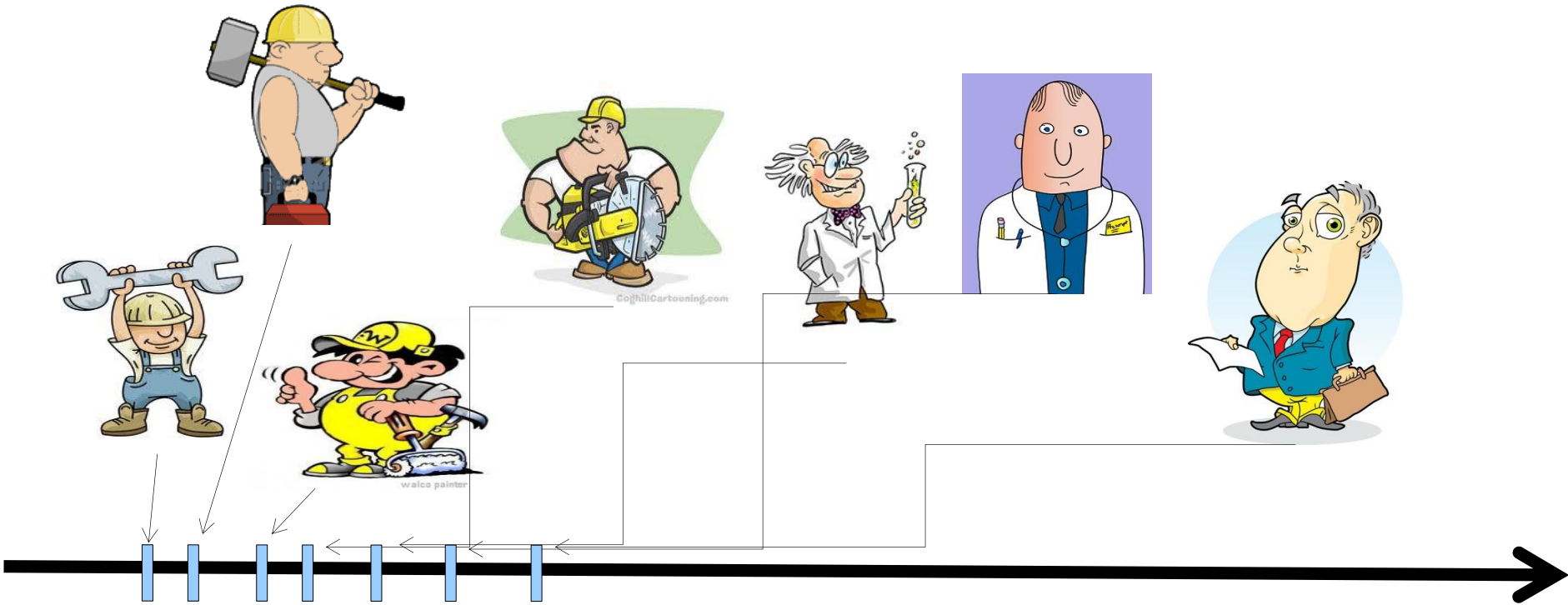
Arrow's Impossibility Theorem

Falls **3 oder mehr** Optionen:

Es gibt kein Wahlsystem, das alle drei Anforderungen erfüllt.

Median und arithmetisches Mittel

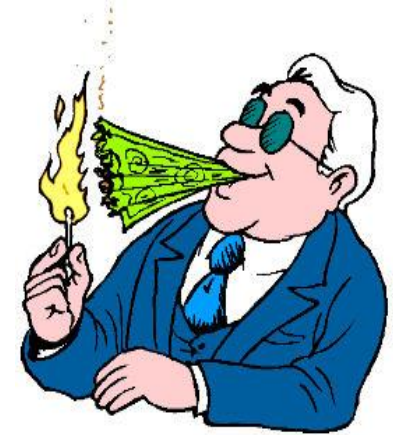
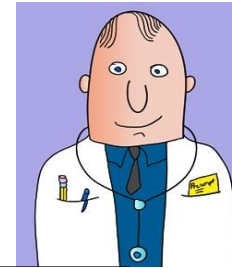
Einkommen von 7 Personen



Median

arithm. Mittel

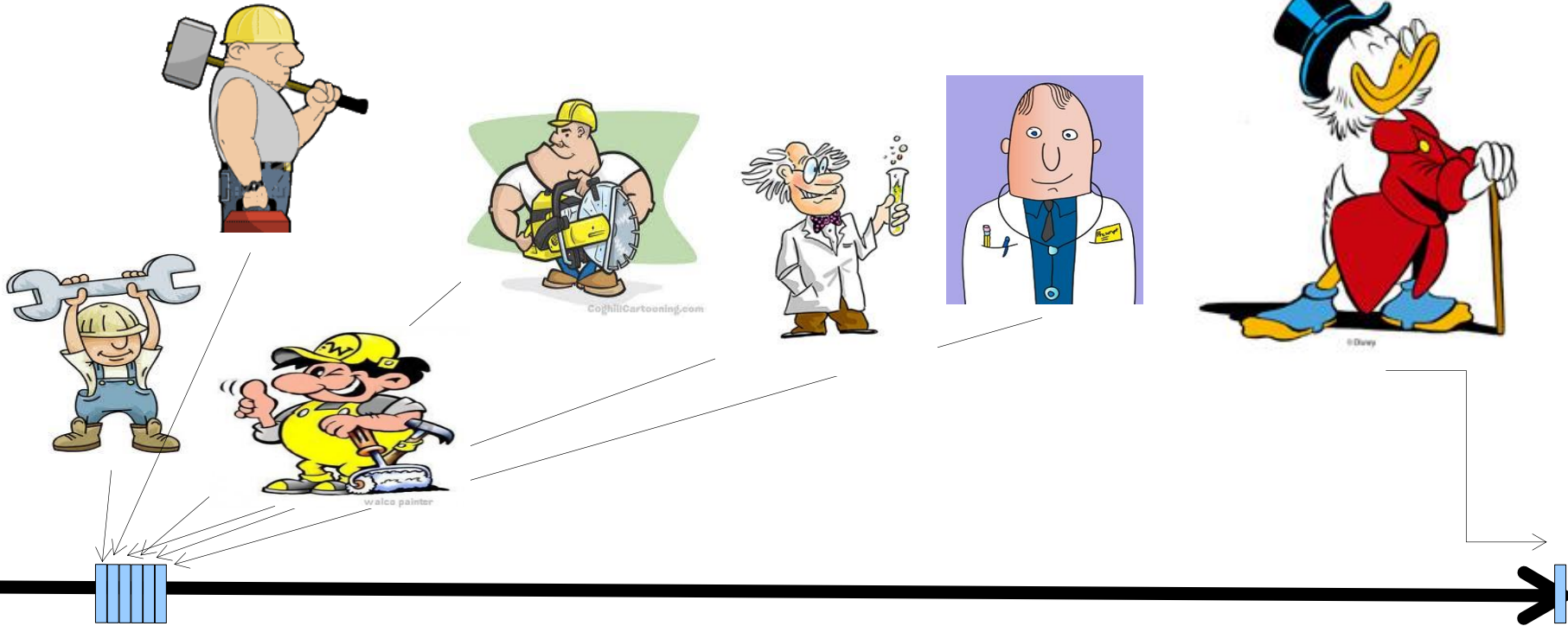
Einkommen von 7 Personen



| Median

| arithm. Mittel

Einkommen von 7 Personen

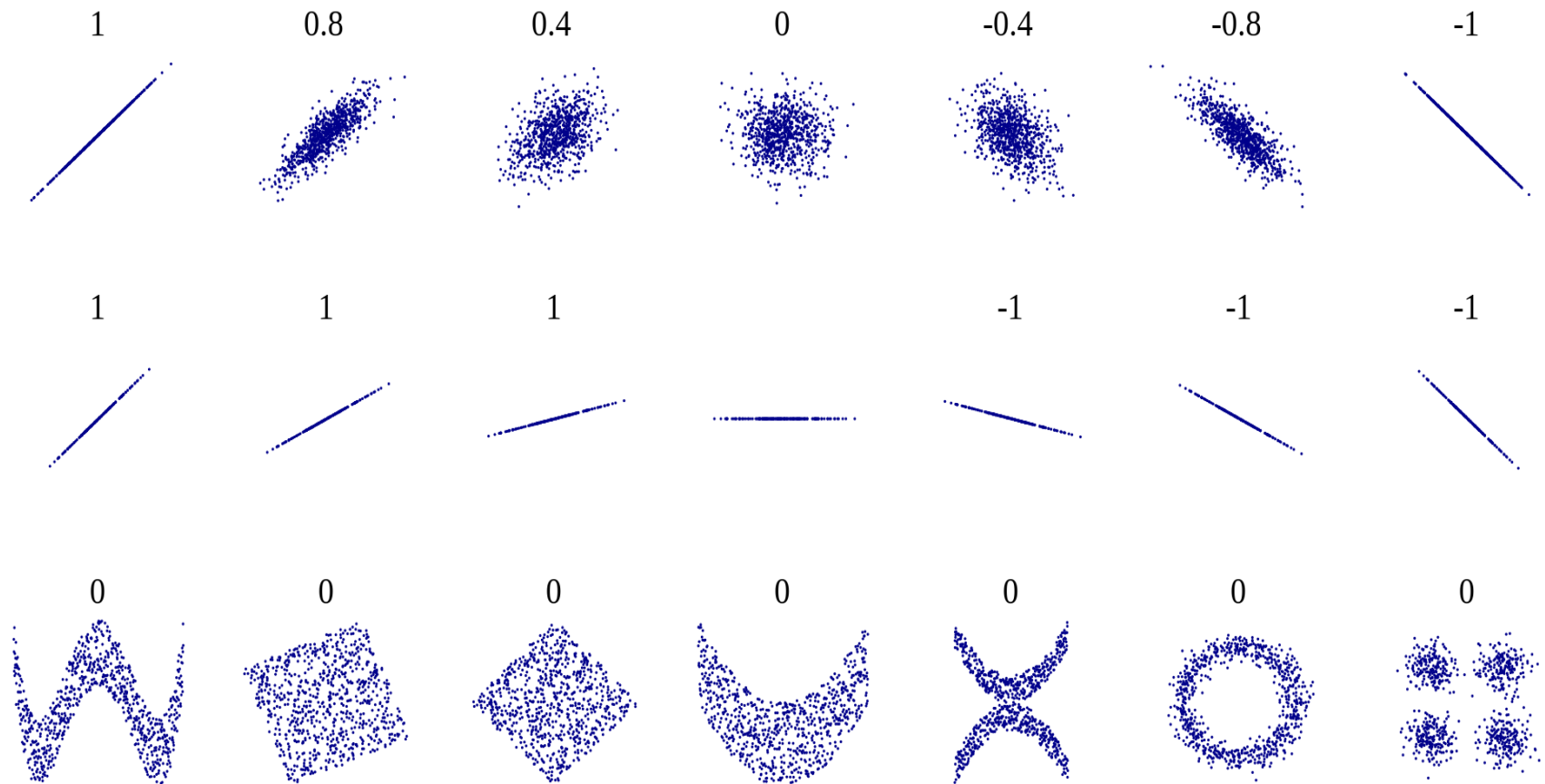


| Median

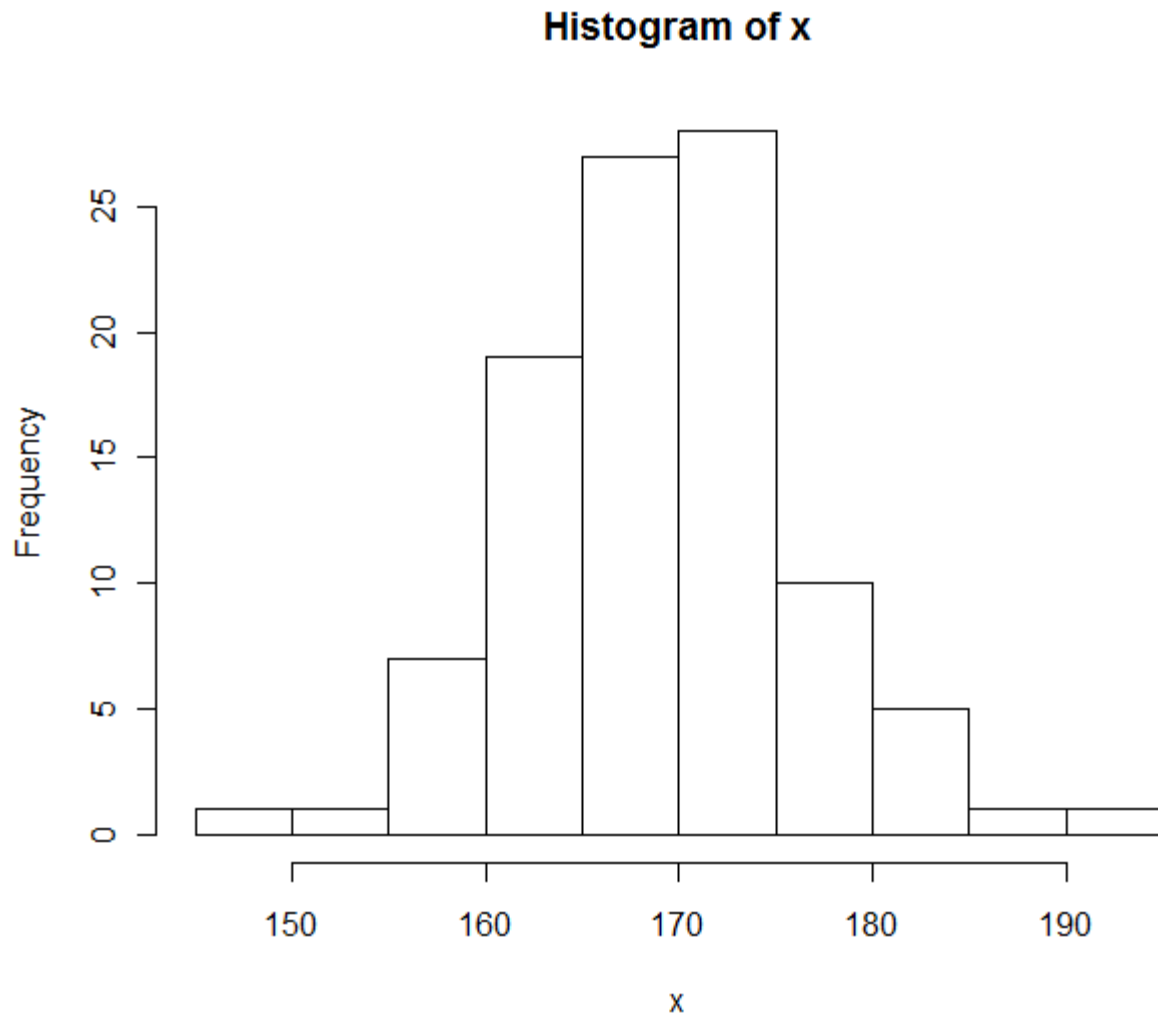
Robust gegen Ausreisser

| arithm. Mittel

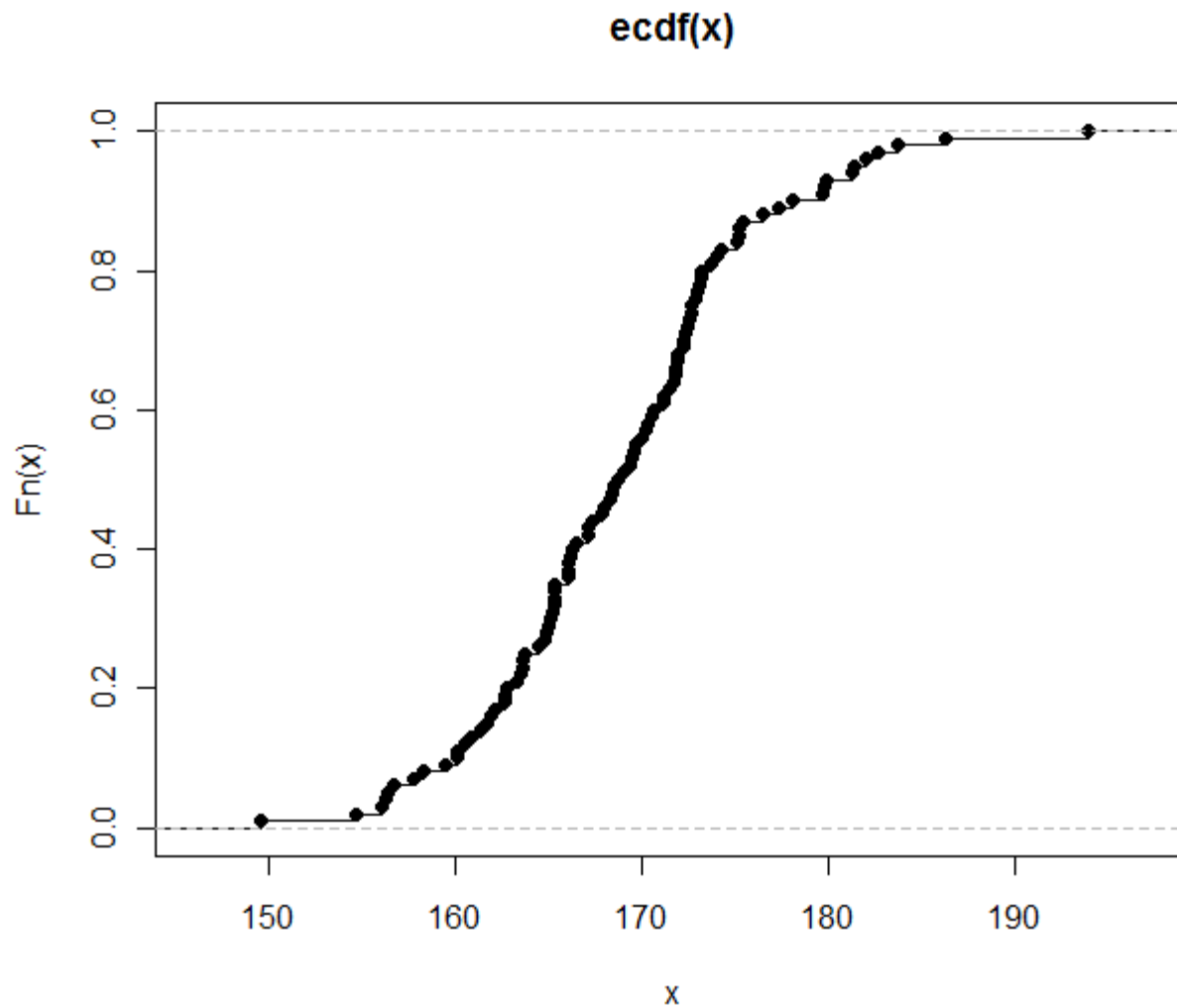
Korrelation



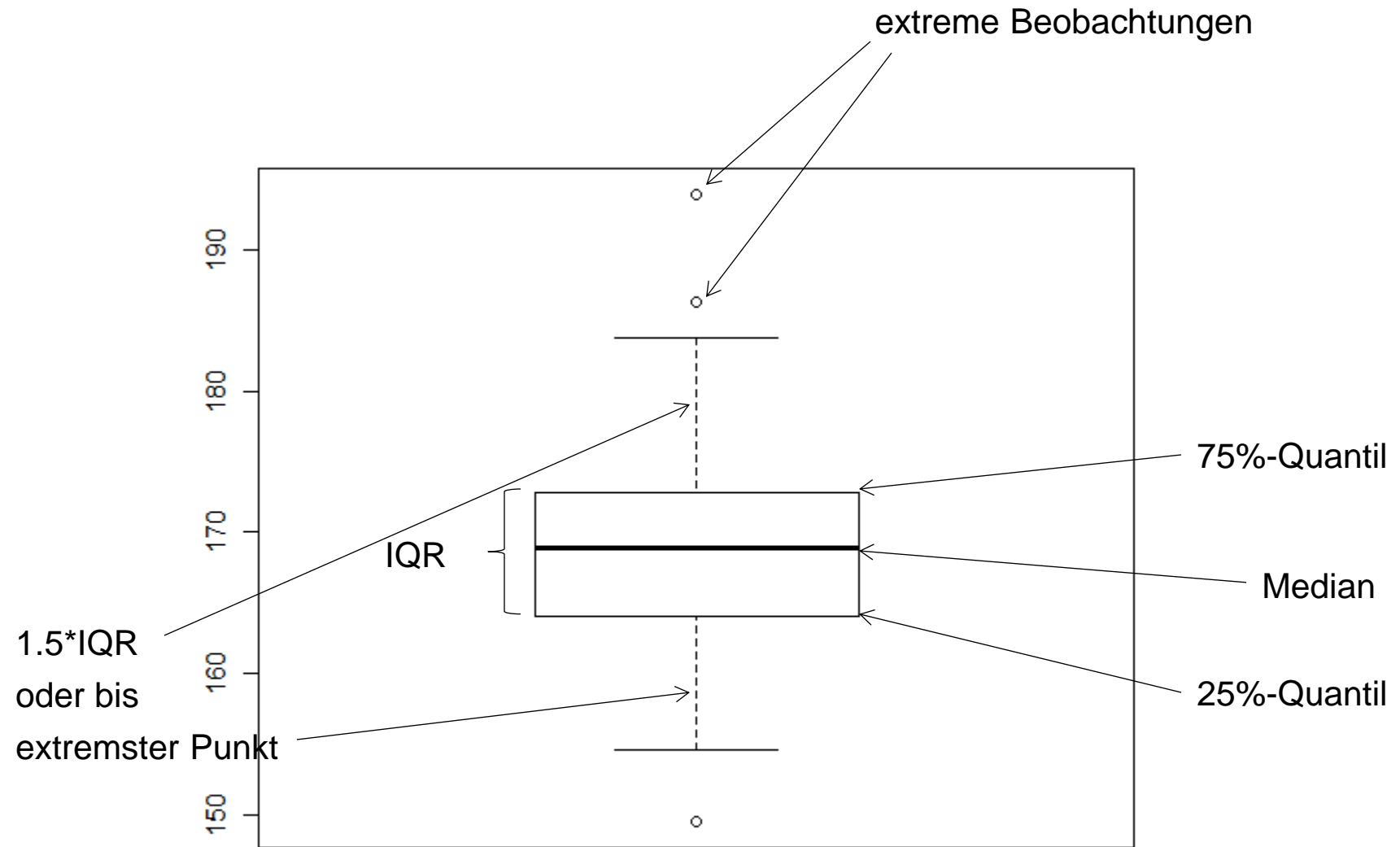
Histogramm



Empirische Verteilungsfunktion (Empirical Cumulative Distribution Function – ECDF)



Boxplot



Prüfung Wi10, A 4.2: Richtige Zuordnung?



- a) A3, B2, C1, D5, E4
- c) A3, B4, C2, D5, E1
- e) A1, B5, C3, D4, E2

- b) A3, B4, C1, D5, E2
- d) A1, B4, C3, D5, E2
- f) A2, B4, C1, D5, E3

