

DAS SIMPSON-PARADOXON

Breno Menezes

ÜBERBLICK

- Einführung
- Beispiele
- Formelle Definition
- Abschluss

EINFÜHRUNG

EINFÜHRUNG

*Entscheidungen werden auf der Grundlage von Daten
getroffen*

*Statistik: Sammlung, Organisation, Analyse, **Interpretation**
und Präsentation von Daten*

Interpretation ist subjektiv

INTERPRETATION IST SUBJEKTIV

MEDAL TABLE

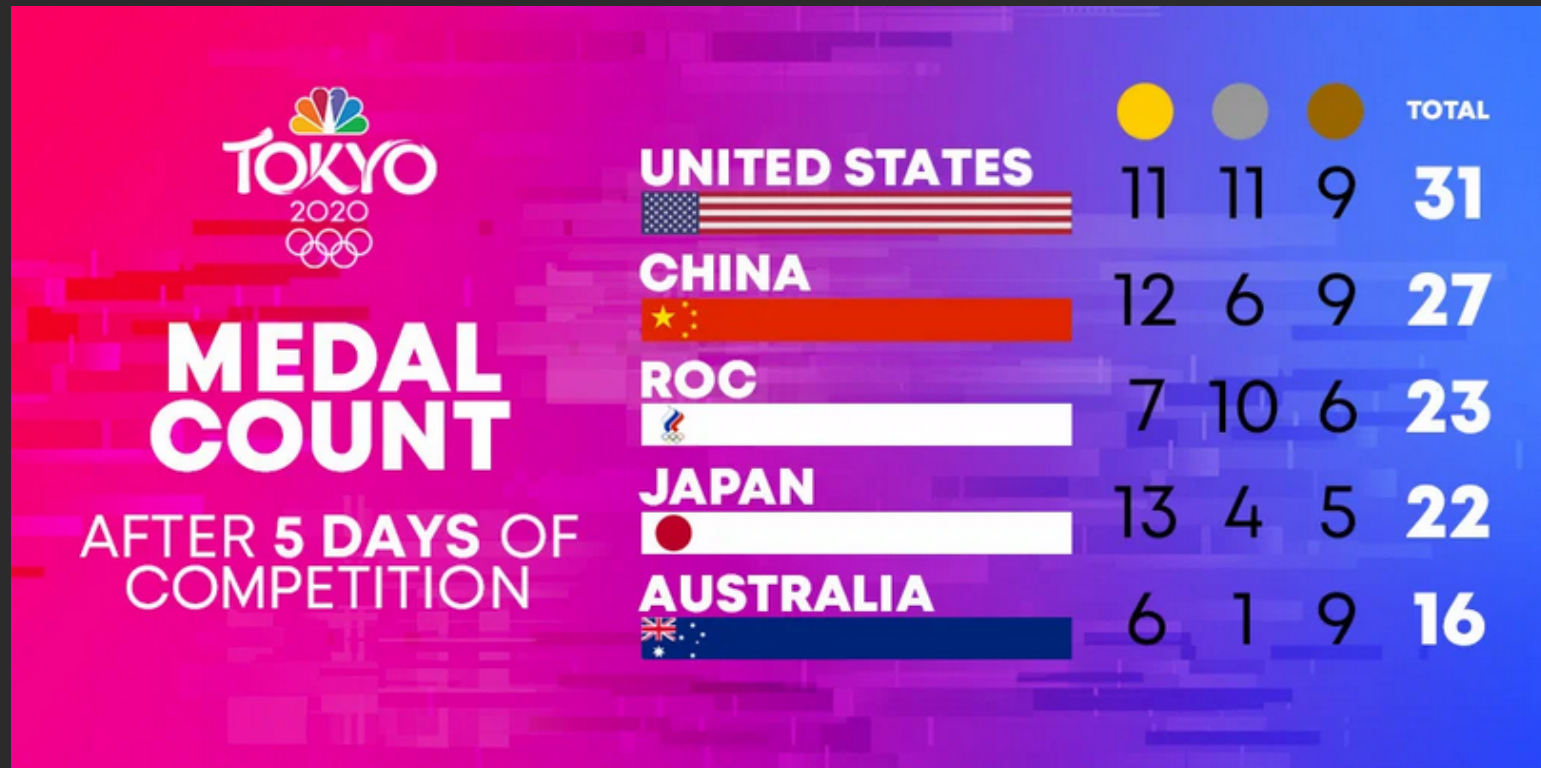
BBC
SPORT

OFFICIAL
BROADCASTER

						Total
1	Japan		13	4	5	22
2	China		12	6	9	27
3	USA		10	11	9	30
4	ROC		7	8	6	21
5	Australia		6	1	9	16
6	Great Britain		5	6	5	16

BBC

INTERPRETATION IST SUBJEKTIV



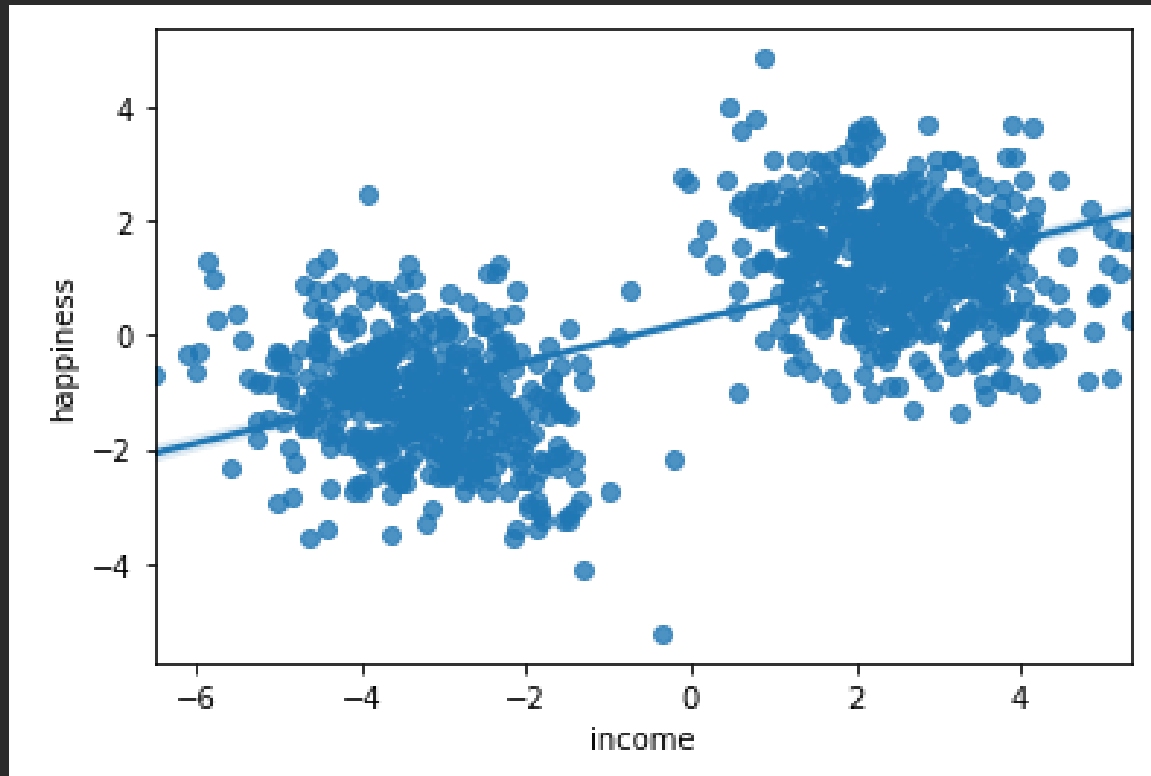
*Es besteht die Tendenz, dass die allgemeine
Schlussfolgerung auch für die spezifischen Fälle gilt*

DAS SIMPSON-PARADOXON

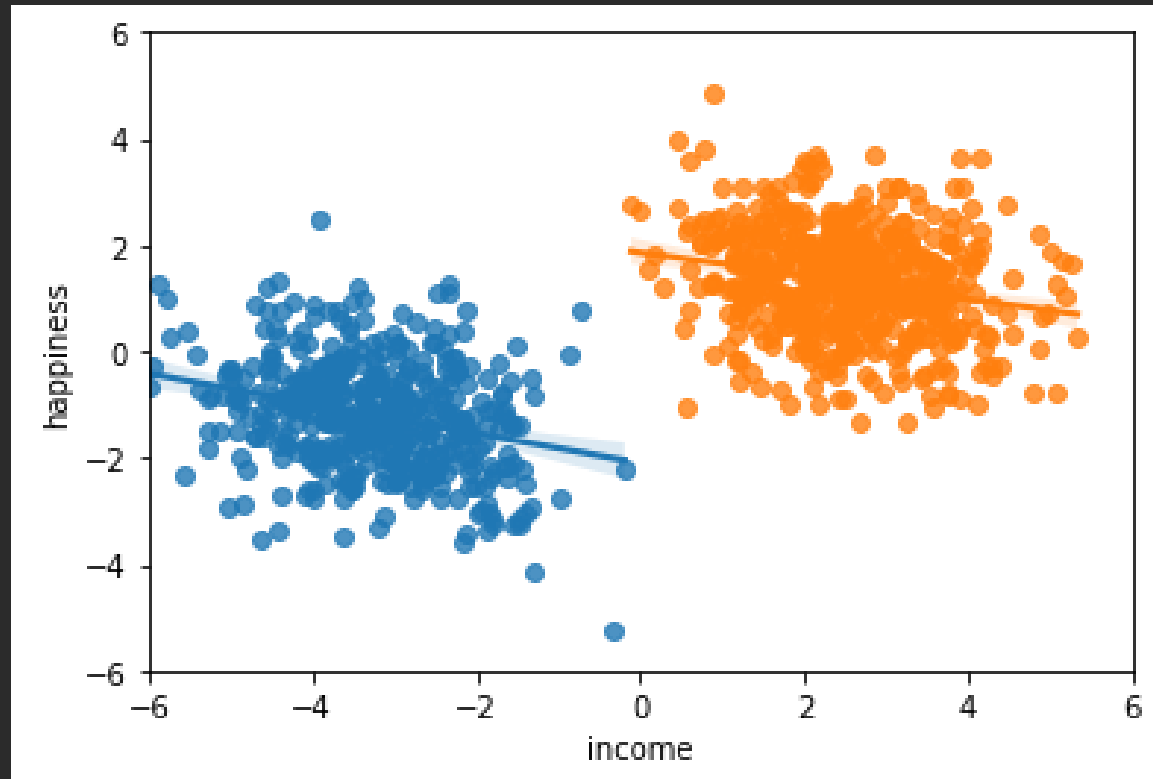
"... wird eine Gesamtstichprobe in Teilstichproben unterteilt, so können sich in allen Teilstichproben Zusammenhänge zeigen, die systematisch nicht dem Zusammenhang in der Gesamtstichprobe entsprechen oder gar konträr ausfallen."^[1]

BEISPIEL 1

MEHR GELD MACHT GLÜCKLICHER

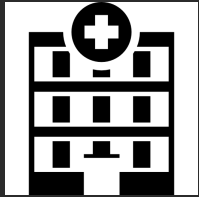


MEHR GELD MACHT GLÜCKLICHER



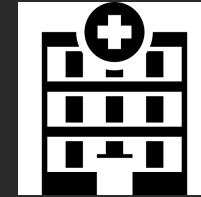
BEISPIEL 2

KRANKENHAUS A ODER B?



Krankenhaus A

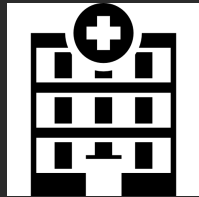
900/1000 genesen



Krankenhaus B

800/1000 genesen

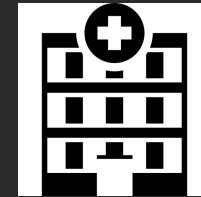
KRANKENHAUS A ODER B?



Krankenhaus A

Schwere Fälle: 30/100
Einfache Fälle: 870/900

30% & 96.7%



Krankenhaus B

Schwere Fälle: 210/400
Einfache Fälle: 590/600

52.5% & 98.3%

KRANKENHAUS A ODER B?

- Krankenhaus B hat in beiden Fällen bessere Chancen auf Genesung
- Unbeachtete Variablen (Schweregrad) können die Schlussfolgerung über einen Fakt völlig verändern

FORMELLE DEFINITION

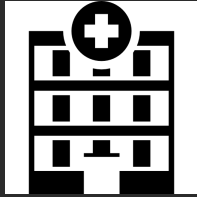
Memo: Falls $\Omega = K_1 \uplus \dots \uplus K_n$, so $\mathbb{P}(B) = \sum_{j=1}^n \mathbb{P}(K_j)\mathbb{P}(B|K_j)$.
(Formel von der totalen Wahrscheinlichkeit)

Für Ereignisse $A, B \subset \Omega$ sowie $\Omega = K_1 \uplus \dots \uplus K_n$ kann Folgendes gelten:

$$\mathbb{P}(B|A \cap K_j) > \mathbb{P}(B|A^c \cap K_j) \text{ für jedes } j = 1, \dots, n$$

$$\text{und (!) } \mathbb{P}(B|A) < \mathbb{P}(B|A^c). \quad (\text{Simpson-Paradoxon})$$

KRANKENHAUS A ODER B?



Krankenhaus A

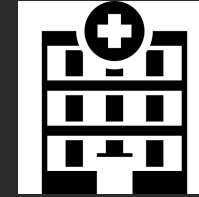
Schwere Fälle: 30/100

Einfache Fälle: 870/900

$$P(G)=0.1 * 0.3 + 0.9 * 0.967$$

$$P(G)=0.9$$

$$P(G|S)=0.3$$



Krankenhaus B

Schwere Fälle: 210/400

Einfache Fälle: 590/600

$$P(G)=0.4 * 0.525 + 0.6 * 0.983$$

$$P(G)=0.8$$

$$P(G|S)=0.52$$

FEHLER VERMEIDEN

- Daten zusammen oder separat betrachten?
- Fokus auf die richtigen Variablen
- Korrelation vs. Kausalität

ABSCHLUSS

DAS SIMPSON-PARADOXON

Statistiken ohne Kontext sind gefährlich

Vorabanalyse der Daten ist wichtig

Verschiedene Perspektiven analysieren

DAS SIMPSON-PARADOXON

Fragen?
BRENO MENEZES

breno.amenezes@gmail.com

