DAS SIMPSON-PARADOXON

Breno Menezes

ÜBERBLICK / GLIEDERUNG

- Einführung
- Beispiele
- Formelle Definition
- Abschluss

EINFÜHRUNG

EINFÜHRUNG

Entscheidungen werden auf der Grundlage von Daten getroffen

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

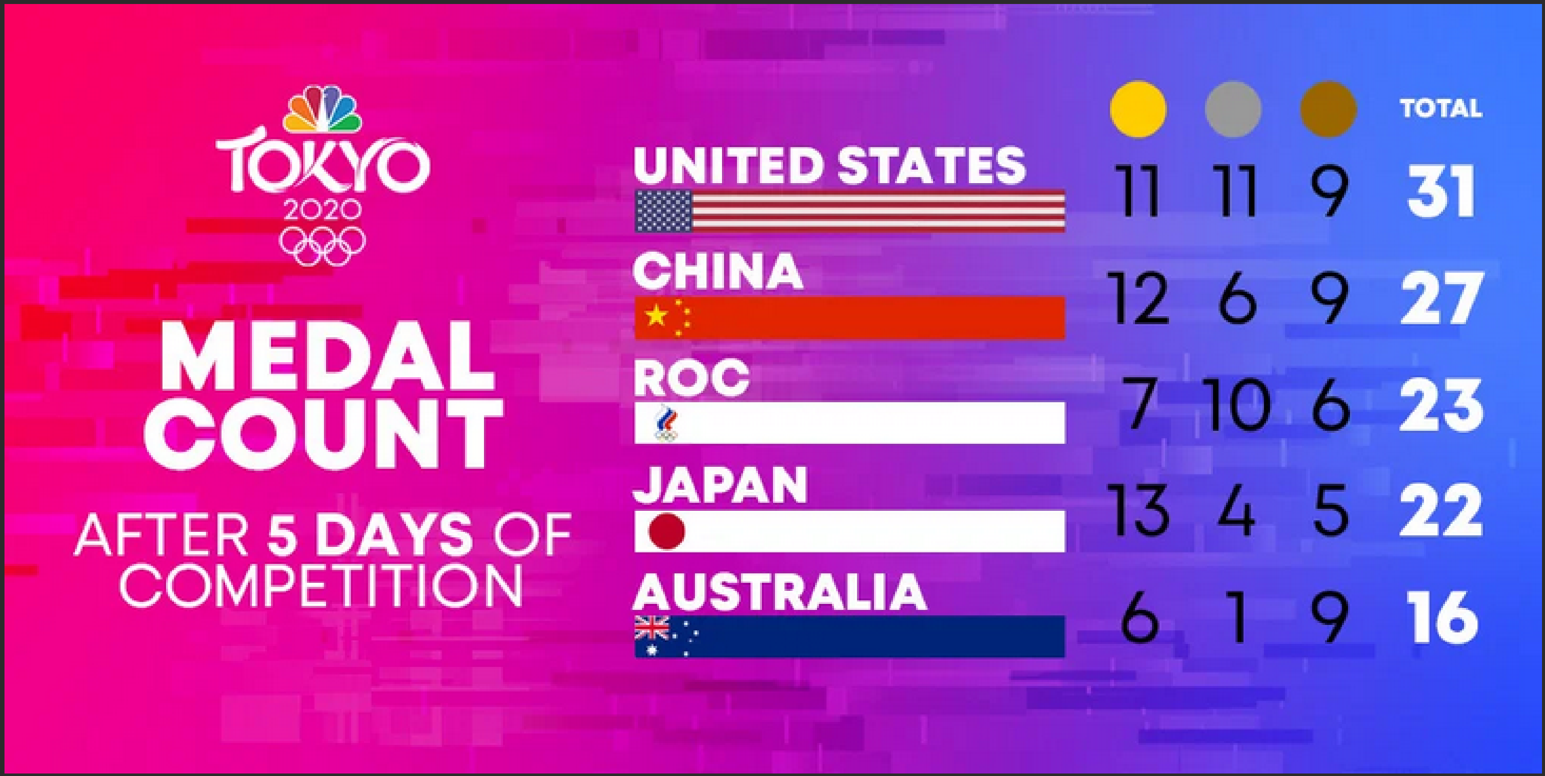
***Interpretation** ist subjektiv*

INTERPRETATION IST SUBJEKTIV

MEDAL TABLE

		G	S	B	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

INTERPRETATION IST SUBJEKTIV



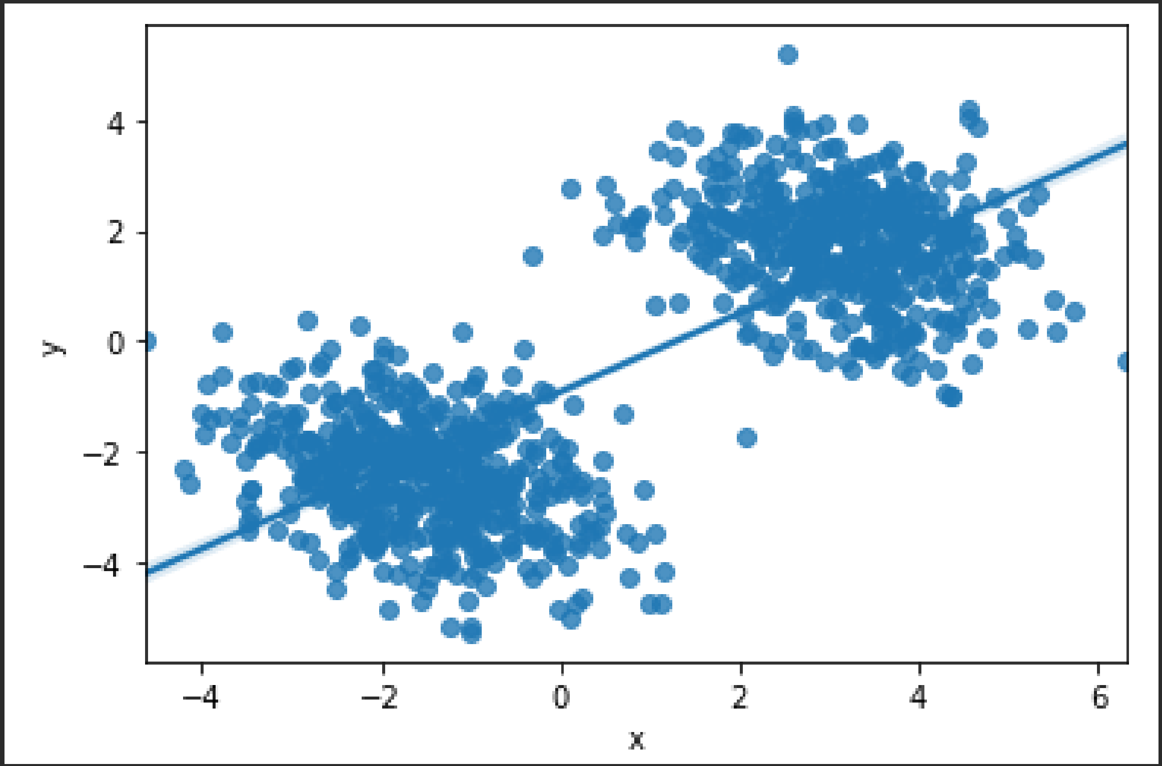
Es besteht die Tendenz zu glauben, dass die allgemeine Regel auch für besondere Fälle gilt

DAS SIMPSON-PARADOXON

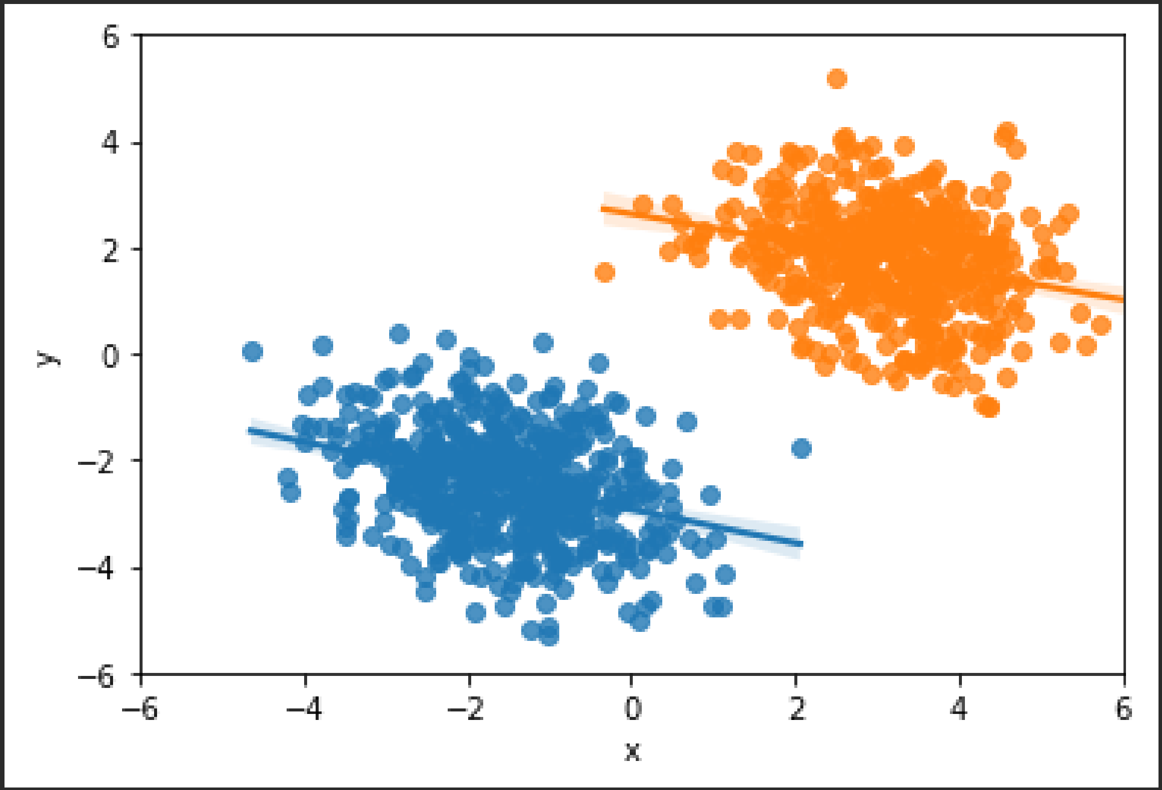
"... wird eine Gesamtstichprobe in Teilstichproben unterteilt, so können sich in allen Teilstichproben Zusammenhänge zeigen, die systematisch nicht dem Zusammenhang in der Gesamtstichprobe entsprechen oder gar konträr ausfallen."[1]

BEISPIEL 1

MEHR GELD MACHT GLÜCKLICHER

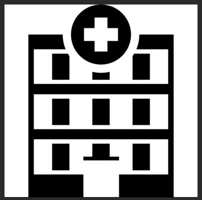


MEHR GELD MACHT GLÜCKLICHER



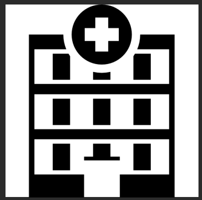
BEISPIEL 2

KRANKENHAUS A ODER B?



Krankenhaus A

900/1000 geheilt

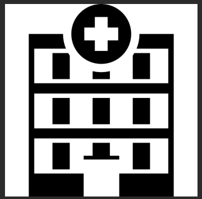


Krankenhaus B

800/1000 geheilt

BESSER INS KRANKENHAUS A GEHEN, ODER?

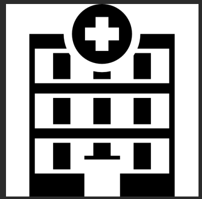
KRANKENHAUS A ODER B?



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 Überlebensraten
- Unbeachtete Variablen können die Schlussfolgerung über einen Fakt völlig verändern (Schweregrad)

FORMELLE DEFINITION

Memo: Falls $\Omega = K_1 \uplus \ldots \uplus K_n,$ so $\mathbb{P}(B) = \sum_{j=1}^{n} \mathbb{P}(K_j)\mathbb{P}(B|K_j)$.

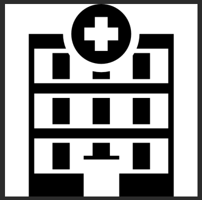
(Formel von der totalen Wahrscheinlichkeit)

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

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

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

KRANKENHAUS A ODER B?



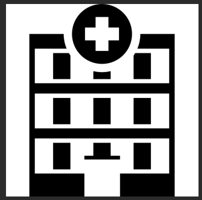
Krankenhaus A

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

30% & 96.7%

*P(U)=0.1 * 0.3 + 0.9 * 0.967*

P(U)=0.9



Krankenhaus B

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

52.5% & 98.3%

*P(U)=0.4 * 0.525 + 0.6 * 0.983*

P(U)=0.8

ABSCHLUSS

DAS SIMPSON-PARADOXON

Statistiken ohne Kontext sind gefährlich

Fehler vermeiden:

- Daten Aufteilung
- Fokus auf die richtigen Variablen

DAS SIMPSON-PARADOXON

Fragen?

BRENO MENEZES

breno.amenezes@gmail.com