Berkson's Paradoxon

Teamprojekt

Niman Deskaj

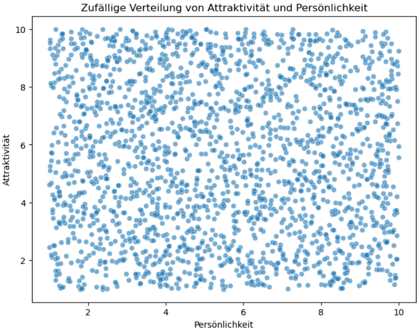
Eberhard Karls Universität Tübingen

26.04.2024

Inhalte

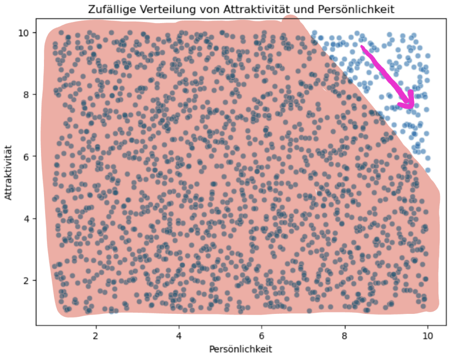
Motivation

Szenario: Frau möchte Mann daten.
Annahme: Variablen *Persönlichkeit* und *Attraktivität* unabhängig.



- keine unattraktiven und unfreundlichen
- unfreundlich, aber attraktiv, oder freundlich und unattraktiv möglich

Motivation



Somit ist durch ihre Präferenzen möglich, dass viele der attraktiven weniger freundlich sind und viele der freundlichen weniger attraktiv
$\Rightarrow$ **negative Korrelation**

Definition

Berkson's Paradoxon: Korrelation zweier (unabhängiger) Größen, zwischen denen kein Kausalzusammenhang besteht, sondern nur eine zufällige oder indirekte Beziehung

- *veridical paradox* (kontraintuitive Einschätzung)
- *selection bias* (Stichprobenverzerrung)

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Historischer Hintergrund

- Joseph Burkson (1946 in *Limitations of the Application of Fourfold Table Analysis to Hospital Data*)
- Fehlinterpretation von statistischen Daten durch fehlerhafte Auswahl von Probanden für Studie
- Risikofaktor: Diabetes, Krankheit: Choleszytitis
- Krankenhauspatienten ohne Diabetes haben viel wahrscheinlicher Choleszytitis, als jemand aus der Bevölkerung, da der Patient aus einem anderen Grund als Diabetes (möglich Choleszytitis-Verursacher) ins Krankenhaus gekommen sein muss
  $\Rightarrow$ **negative Korrelation, ABER kein kausaler Zusammenhang**

Mathematische Formulierung

Seien X und Y zwei unabhängige Variablen, dann gilt:

$$P(X|Y) = P(X)$$

Wir wollen, wie in den Beispielen, dass mind. ein Ereignis auftritt oder beide, aber nicht keins:

$$X \cup Y = X + Y - (X \cap Y)$$

mit $P(X \cup Y) \leq 1$. **Für uns $P(X \cup Y) < 1$, da $(\neg X \cap \neg Y)$ nicht erwünscht**

Mathematische Formulierung

Wir erhalten

$$P(X) < \frac{P(X)}{P(X \cup Y)}$$

wobei

$$\frac{P(X)}{P(X \cup Y)} = \frac{P(X \cap (X \cup Y))}{P(X \cup Y)} = P(X|X \cup Y)$$

und folglich

$$P(X) < P(X|X \cup Y)$$

Mathematische Formulierung

Und da

$$P(X) = P(X|Y) = P(X|Y, X \cup Y)$$

gilt

$$\boxed{P(X|Y, X \cup Y) < P(X|X \cup Y)}$$

Mathematische Formulierung

Was, wenn wir nicht *(X or Y)* sondern *(X and Y)* als Auswahlbedingung wollen?

Wir erwarten, dass X und Y immer gleichzeitig auftreten also

$$P(X|Y,(X \cap Y)) = P(Y|X,(X \cap Y)) = 1$$

$\Rightarrow$ perfekte positive Korrelation

Vermeidung von Stichprobenverzerrung

- zufällige Stichprobenmenge
- Umfragen/Studien: auf Gemeinsamkeiten/wenige Unterschiede zwischen Untergruppe und Population achten
- Auf große Stichprobenmengen fokussieren

Fazit

Wichtige Erinnerung an die Wissenschaft zur gezielten Auswahl der Stichproben und Populationen um ungenaue oder falsche Schlussfolgerungen zu vermeiden!

Quellen

Vielen Dank für Eure Aufmerksamkeit!

- https://en.wikipedia.org/wiki/Berkson
- https://www.youtube.com/watch?v=0kpFcdKwwvM
- https://link.springer.com/chapter/10.1007/978-3-662-64776-9_10
- https://towardsdatascience.com/berksons-paradox-in-machine-learning-113818ac7657
- http://corysimon.github.io/articles/berksons-paradox-are-handsome-men-really-jerks/