

Mathematik III – Wahrscheinlichkeit & Stochastik

Das simpsonsche Paradoxon

Bedingte Wahrscheinlichkeiten / Statistik

Fakultät II – Management und Informationssysteme - Informatik

Vorgelegt von: Christopher Stieglitz

Dozentin: Prof. Dr. Petram

Inhaltsverzeichnis

- Einleitung
 - Definition – Was ist das simpsonsche Paradoxon?
 - Bedingte Wahrscheinlichkeiten
- Aggregierte vs. Unterteilte Werte
 - Beispiel anhand von Medizinischen Testgruppen
 - Erklärung anhand von bedingter Wahrscheinlichkeit
- Das Paradoxon visualisiert
- Schlussfolgerung

Einleitung - Definition

Das simpsonsche Paradoxon beschreibt eine Situation der Statistik bei der eine aggregierte Sicht der erhobenen Zahlen zu einem anderen, oder entgegengesetzten Ergebnis führt, wie auf einer detaillierten Betrachtung der Daten.

Eine Korrelation kann in diesem Fall durch andere Betrachtung derselben Zahlen umgekehrt werden, wenn sie anders gruppiert sind.

Einleitung – Bedingte Wahrscheinlichkeit

- Berechnen einer Wahrscheinlichkeit unter der Bedingung, dass ein Ereignis bereits eingetreten ist
- $P(A|B) = P(A \cap B) / P(B)$
- Anwendungsbereiche:
 - Bayes'sches Inferenzverfahren
 - Modellierung von Abhängigkeiten zwischen Ergebnissen
 - Statistische Analyse und Entscheidungsfindungen

Aggregierte Werte - Beispiel

Was sehen wir:

- Medikament X schneidet besser ab als das Placebo
 - => Das Medikament wirkt
- In jeder Testgruppe sind jeweils 80 Personen

Medikament	Wirksam	Unwirksam	Wirksamkeit
X	40	40	50%
Y (Placebo)	32	48	40%

$$P(\text{Wirksam} | \text{Placebo}) = 32 / 80 = 0.4$$

Unterteilte Werte - Beispiel

- Das Medikament schneidet nun in beiden Unterteilungen schlechter ab als zuvor
 - => Medikament wirkt nicht
 - Die Verteilung der Geschlechter auf die Medikamenten sind ungleich verteilt
- $P(\text{Wirksam} | \text{Placebo}) = P(\text{Wirksam} \cap \text{Placebo}) / P(\text{Placebo}) = (14+18)/(20+60) = 0.4$

	Männer			Frauen		
Medikament	Wirksam	Unwirksam	Wirksamkeit	Wirksam	Unwirksam	Wirksamkeit
X	36	24	60%	4	16	20%
Y (Placebo)	14	6	70%	18	42	30%

=> Wenn man das Geschlecht weiß, verschreibt man das Medikament nicht?

Aggregierte & Unterteilte Werte - Erklärt

- $P(\text{Wirksam} | \text{Placebo})$ berechnet sich aus der gewichteten Summe der Wahrscheinlichkeiten $P(\text{Wirksam} | \text{Placebo} \& \text{Männlich})$ und $P(\text{Wirksam} | \text{Placebo} \& \text{Weiblich})$
- Aus der Verteilung ist erkenntlich:
 - Frauen sind beim Medikament unterrepräsentiert
 - Männer sind beim Placebo unterrepräsentiert
 - Bei Männern wirken beide Proben besser, zugunsten vom Medikament

$$P(M | \text{Med}) = 60 / (60 + 20) = 0.75$$

$$P(W | \text{Med}) = 20 / (20 + 60) = 0.25$$

$$P(M | \text{Plac}) = 20 / (20 + 60) = 0.25$$

$$P(W | \text{Plac}) = 60 / (60 + 20) = 0.75$$

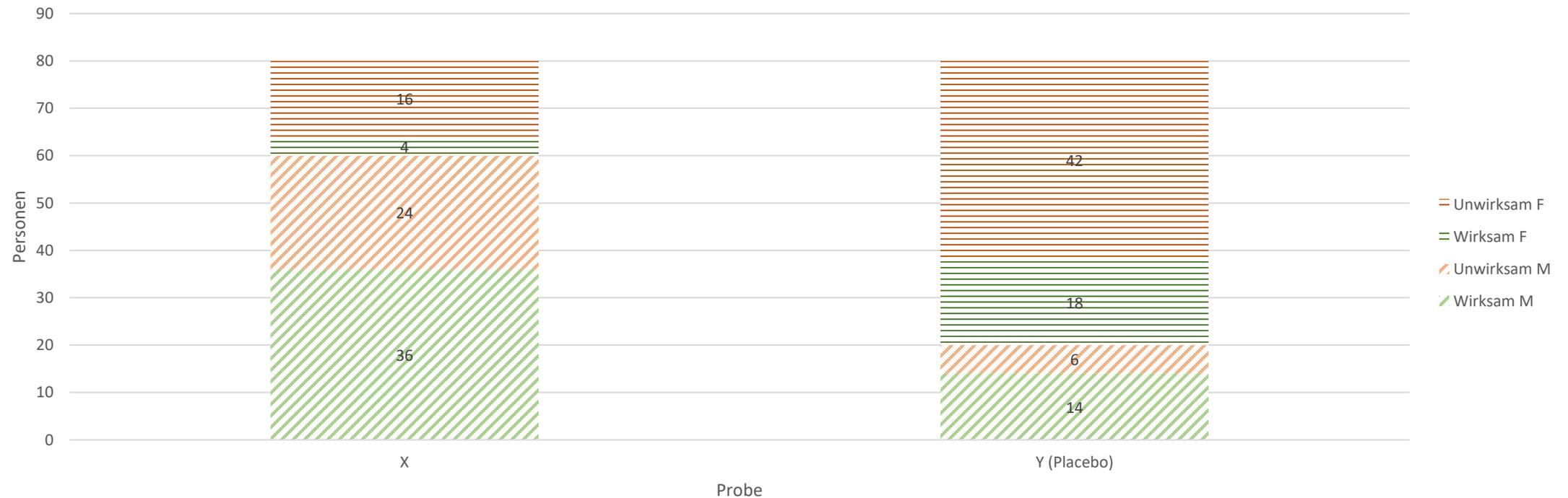
Zurückrechnen:

$$P(\text{Wirksam} | \text{Med}) = 0.75 * 0.6 + 0.25 * 0.2 = 0.5$$

$$P(\text{Wirksam} | \text{Plac}) = 0.75 * 0.3 + 0.25 * 0.7 = 0.4$$

Das Paradoxon visualisiert

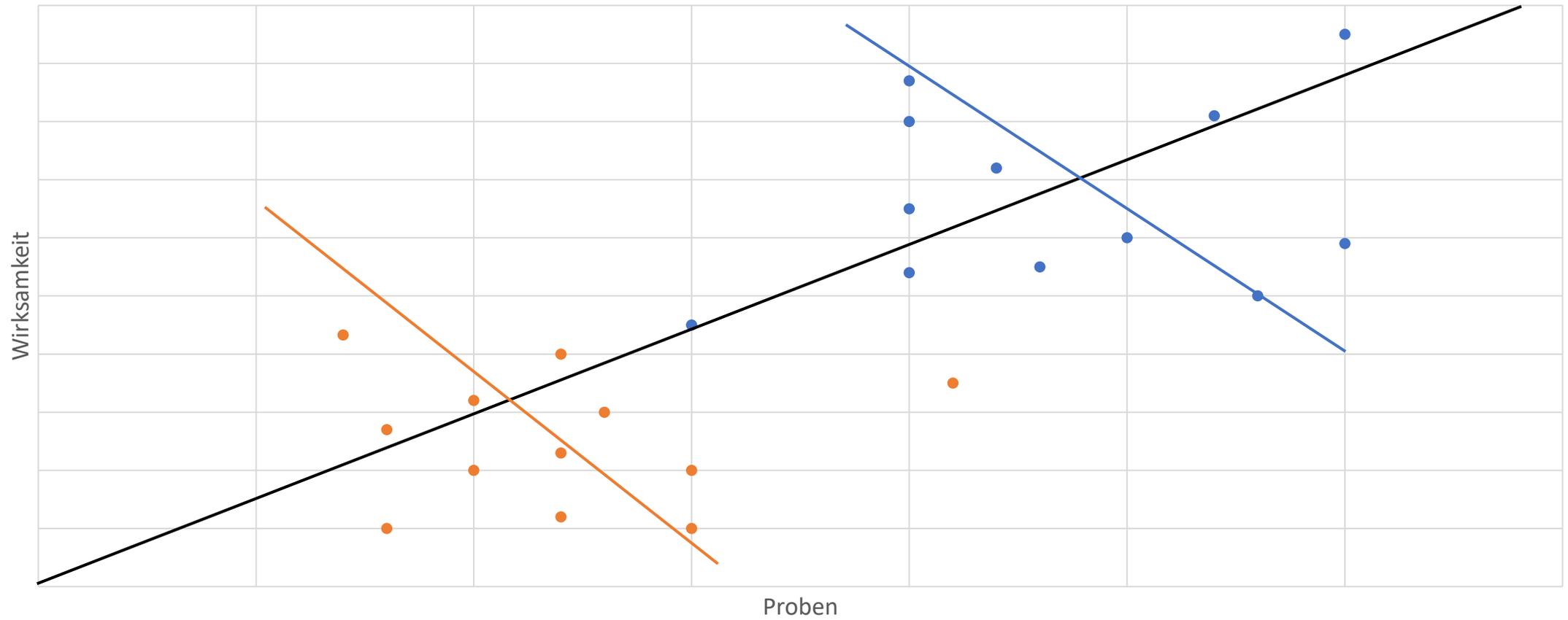
- Querstriche = Männer, Grünfarben = wirksam
- Die Gewichtung des Geschlechts, pro Probe ist augenscheinlich nicht ausbalanciert
- Auch hier ist das Paradoxon aufgrund des Aggregationsfehlers nicht direkt erkennbar (Man würde X besser bewerten)



Schlussfolgerung

- Das Paradoxon tritt auch auf, wenn die Probenverteilung nicht gleich ist
- Statistiken immer hinterfragen
- Einfluss- oder Störfaktoren immer mit einbeziehen (Hier: Abhängigkeit des Geschlechts)
 - Bzw. bei unklarem Verhalten der Daten weiter analysieren
 - Geschlecht beeinflusst das Ergebnis, sowie die Ursache
 - Kausalität liegt darin, dass Männer überwiegend das Medikament und Frauen überwiegend das Placebo genommen haben

Schlussfolgerung - visualisiert



Quellen

- **Das Simpson-Paradoxon | Mathewelten | ARTE**
 - https://www.youtube.com/watch?v=I-FTa_fwB0
- **DEWIKI - Simpson-Paradoxon**
 - <https://dewiki.de/Lexikon/Simpson-Paradoxon>
- **Das simpsonsche Paradoxon - So kommt Licht ins Dunkel**
 - Timm Grams, Fulda, 13.07.2010
 - <https://www2.hs-fulda.de/~grams/Denkfallen/Simpson.pdf>