



Lösungsvorschläge zum 1. Übungsblatt zur „Mathematik und Statistik für Biologen“

Lösung zur Aufgabe 1

(3 Punkte)

Bei allen drei Sätzen kann man nicht die Kausalität folgern, da hier eine „Gleichzeitigkeit“ beobachtet wird. Aufgrund der Gleichzeitigkeit können immer konfundierte Faktoren eine Rolle gespielt haben, die hier nicht mit berücksichtigt wurden, und deshalb können kausale Zusammenhänge nicht gefolgert werden.

Zum 2-ten Satz: Um hier eine Aussage machen zu können, fehlt insbesondere auch noch ein Vergleich! Es müsste verglichen werden, wie viele Personen ohne Vater aufgewachsen sind und nicht diese kriminellen Wege eingeschlagen haben.

Weitere mögliche Aspekte:

Zu Satz 1:

- Keine Angaben, ob die Daten nur aus z.B. einer Region stammen oder nicht. Falls sie nur aus einer Region oder bestimmten Regionen stammen, erfolgt dadurch eine Überrepräsentation. Z.B. könnte der Fall auftreten, dass nur sozial schwache Ortschaften berücksichtigt wurden.
- Auch kann nicht gefolgert werden, dass das Aufwachsen ohne Vater zu Schulversagen, Drogenabhängigkeit usw. führt. Dazu müsste ein Vergleich erfolgen, wieviele ohne Vater aufwachsen und trotzdem gut in der Schule waren oder keine kriminelle Wege gegangen sind.

Somit bleibt nur die Vermutung übrig, dass Schulversager, Drogenabhängige, usw. wahrscheinlich ohne Vater aufgewachsen sind - folgern kann man das aber nicht!

Zu Satz 2:

Ähnliche Begründung wie oben. Unter „allen Insassen“ ist nun nicht sicher, ob wirklich „alle Insassen“ aus dem ganzen Land oder doch nur aus Gefängnissen aus bestimmten Regionen gemeint werden.

Wieder ist zu argumentieren: Es liegt die Vermutung nahe, dass Aufwachsen ohne Vater zu kriminellen Handlungen führen könnte. Aber eine Folgerung kann aus statistischer Sicht nicht folgen.

Zu Satz 3:

Hier stellt sich wieder die Frage, wie wurde dieser Test durchgeführt! Es kann durchaus wieder sampling bias auftreten, falls nur bestimmte Regionen betrachtet wurden. Auch wäre denkbar, dass non response bias auftreten könnte, da evtl. nicht alle Befragten eine Antwort lieferten (aus Gründen der Scham, Peinlichkeit,...).

Schlussfolgernd ist zu bemerken, dass eine Folgerung von kausalen Zusammenhängen unter statistischer Sicht sehr schwierig ist, falls nicht explizit angegeben wurde:

- wie genau eine Datenerhebung zustande gekommen ist
- ob alle konfundierten Faktoren mitberücksichtigt worden sind.

Lösung zur Aufgabe 2

(3 Punkte)

- Ja: Vergleiche Todesfälle in der gesamten Studiengruppe (Alle) mit der Kontrollgruppe. Beide Gruppen unterscheiden sich nur dadurch, dass bei einem Teil der Studiengruppe die Vorsorgeuntersuchung durchgeführt wurde.
- Weil die Vorsorgeuntersuchung keinen bzw. kaum Einfluss auf Krankheiten verschieden von Brustkrebs hatte.
- Weil:
 - ärmere Frauen häufiger die Vorsorgeuntersuchung verweigerten als reichere Frauen,
 - die meisten Krankheiten stärker bei ärmeren Frauen als bei reicheren Frauen auftraten.
- Vergleiche Todesrate durch Brustkrebs in der KG (2.0) mit Todesrate bei Verweigert (1.5). Keine dieser Beiden Gruppen wird durch die Vorsorgeuntersuchung beeinflusst, allerdings besteht die KG aus im Schnitt reicheren Frauen als die Gruppe Verweigert (da vor allem ärmere Frauen die Vorsorgeuntersuchung verweigert haben).
- Nicht sinnvoll. Die betrachteten Gruppen unterscheiden sich nicht nur durch die Untersuchung, sondern auch noch durch andere Faktoren (z.B. Einkommen). Der Einfluss dieser Faktoren vermengt sich hier mit dem der Untersuchung, so dass man nicht auf die Untersuchung zurückschließen kann.
- Widerlegen: Todesrate in der gesamten SG (Alle) bzgl. sonstiger Todesursachen ist ungefähr so groß wie die entsprechende Rate in der KG.
- Nein, vielmehr wurde durch die Untersuchung nur die Zahl der im ersten Jahr der Studie **entdeckten** Fälle von Brustkrebs beeinflusst. Diese ist aber (insbesondere in der KG!) nicht gleich der im ersten Jahr vorhandenen Erkrankungen an Brustkrebs.

Lösung zur Aufgabe 3

(3 Punkte)

- **sampling bias:** Tritt bei 2) auf, da die Institutsdirektoren in der Regel nicht die Nutzer und auch nicht repräsentativ für die Nutzer sind.
- **non response bias:** tritt bei 1) und 2) auf, da davon auszugehen ist, dass nicht jeder Mitarbeiter die Zeit hat diese Umfrage auszufüllen. Sogesehen tritt eine Verzerrung durch Nichtbeantworten bei 2) wahrscheinlich sogar noch mehr auf, da die Institutsdirektoren in der Regel noch weniger Zeit haben als deren Mitarbeiter.