

Aufgabe 1) Counterfactual Thinking

Wir untersuchen den Einfluss des Kaffeekonsums (K) auf Pankreaskarzinom (PK), d.h. wir möchten wissen ob eine Änderung des Kaffeekonsums in der Population eine Änderung in der Häufigkeit von Pankreaskarzinomen zur Folge hat. Wenn man jede Person beiden Expositionen aussetzen könnte, könnte man den kausalen Zusammenhang zwischen K und PK bestimmen (z.B. durch Odds Ratio oder Relatives Risiko).

- a) Geben Sie die vier möglichen Profile an (mit Wahrscheinlichkeiten p_1, p_2, p_3 und p_4), wenn man jede Person beiden Expositionen aussetzen könnte.
- b) Berechnen Sie die Wahrscheinlichkeiten $P(PK|K)$ und $P(PK|\bar{K})$ in Abhängigkeit von p_i .
- c) Berechnen Sie die kausale Odds Ratio OR_{kausal} .
- d) Welche Wahrscheinlichkeiten müssen gleich sein, falls $OR_{kausal} = 1$?
- e) Bedeutet $OR_{kausal} = 1$, dass der Kaffeekonsum bei keiner Person das Risiko für Pankreaskarzinom beeinflusst?
- f) In einer Studie an N Personen wurden alle Personen mit Profil 1 und 3 exponiert, und alle Personen mit Profil 2 und 4 nicht. Berechnen Sie für diesen Fall die OR.
- g) In einer zweiten Studie wurden alle Personen mit Profil 3 und 4 exponiert, und alle Personen mit Profil 1 und 2 nicht. Berechnen Sie hierfür die OR.
- h) Nun wird die Zuordnung zur Expositionsgruppe zufällig durch einen Münzwurf entschieden. Berechnen Sie in diesem Fall die OR. Was stellen Sie fest?

Aufgabe 2) Statistische Interaktion

Die folgende Tabelle zeigt die Assoziation einer Exposition E mit einem Outcome D , stratifiziert nach dem Geschlecht.

<i>Geschlecht</i>			$D = 1$	$D = 0$
Frauen	E	$E = 1$	366	145
	$E = 0$	162	346	
Männer	E	$E = 1$	151	359
	$E = 0$	332	139	

- a) Berechnen Sie die OR für die beiden Strata. Was kann man über die Variable **Geschlecht** sagen?
- b) Wir wollen den Einfluss von E auf D mit logistischer Regression untersuchen. Was für ein Modell sollten wir wählen?

- c) Im Folgenden sehen Sie den R-Output für zwei verschiedene Modelle:

MODELL 1:

Call:
glm(formula = outcome ~ sex + exposition)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.5193	-0.4934	0.4807	0.4830	0.5088

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.491187	0.019758	24.861	<2e-16 ***
sex1	0.025842	0.022376	1.155	0.248
exposition1	0.002245	0.022377	0.100	0.920

MODELL 2:

glm(formula = outcome ~ sex * exposition)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.7162	-0.3189	0.2838	0.2951	0.7039

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.70488	0.02110	33.41	<2e-16 ***
sex1	-0.38599	0.02929	-13.18	<2e-16 ***
exposition1	-0.40880	0.02926	-13.97	<2e-16 ***
sex1:exposition1	0.80615	0.04098	19.67	<2e-16 ***

Interpretieren Sie die beiden Modelle.

- d) Berechnen Sie aus beiden Modellen die Odds Ratio für Männer und Frauen.
- e) Erstellen Sie für vier Untergruppen, definiert durch zwei verschiedene binäre Variablen X und E (Exposition), das *Counterfactual Profile* (insgesamt 16 Profile). Welche Profile sind relevant für die statistische Interaktion?

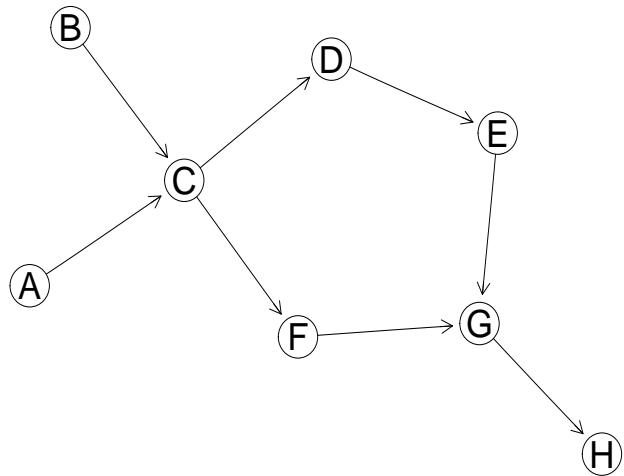
Aufgabe 3 Wir untersuchen den Einfluss der sportlichen Betätigung auf Herzkrankheiten. Es ist bekannt, dass Raucher weniger Sport treiben, und dass Personen mit einer familiären Vorbelastung sportlicher sind, um Herzkrankheiten vorzubeugen.

- a) Erstellen Sie einen DAG mit den Variablen Sport (S), Herzkrankheiten (HK), Rauchen (R) und familiäre Vorbelastung (F).
- b) Durch welche Variablen wird die Beziehung zwischen Sport und Herzkrankheiten verzerrt?

- c) Wir wissen, dass Sport das Auftreten und die Dauer einer Depression positiv beeinflusst. Außerdem ist klar, dass kranke Personen häufiger unter Depressionen leiden als Gesunde. Ergänzen Sie Ihren Graphen durch Hinzufügen der neuen Variable **Depression (D)** und erklären Sie, warum man nach D nicht adjustieren soll.

Aufgabe 4) Gegeben sei der folgende DAG:

Abbildung 1:



Welche der folgenden Aussagen ist wahr?

- i) $F \perp\!\!\!\perp H|G$
- ii) $C \perp\!\!\!\perp G|F$
- iii) $F \perp\!\!\!\perp E|C$
- iv) $A \perp\!\!\!\perp B|\emptyset$
- v) $A \perp\!\!\!\perp B|D$
- vi) $D \perp\!\!\!\perp F|\{C, G\}$
- vii) $E \perp\!\!\!\perp F|\{A, B\}$
- viii) $C \perp\!\!\!\perp E|\{D, F, H\}$