



Kausale Inferenz

Kapitel 5: Die g-Formula

Dezember 2019

Version: 10. Dezember 2019

Michael Schomaker

UNIT University, Institute of Public Health, MDM and HTA

- auch bekannt als “g-computation”
- ist nichts anderes als Standardisierung bezüglich $L = \mathbf{L} = \{L_1, L_2, \dots\}$
- Idee: viele “target quantities” sind *marginal* bezüglich der Kovariablen, also *nicht* bedingt auf \mathbf{L} .

Beispiel: Der ATE für eine binäre Zielgröße ist: $P(Y^1 = 1) - P(Y^0 = 1)$. Mich interessiert der Anteil der Patienten, die nach 1 Jahr gestorben wären wenn sie alle behandelt worden wären; und der Anteil, der gestorben wäre, wenn sie alle nicht behandelt worden wären – *unabhängig* von anderen gemessenen Kovariablen (also *nicht* unbedingt stratifiziert/bedingt bezüglich von Kovariablen).

g-formula

Konzept

Manuelle Berechnung
Berechnung mit R
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

Die g-formula [1]

Unter den Annahmen von Konsistenz, bedingter Austauschbarkeit und Positivität gilt:

$$E(Y^a) = \int_l E(Y|A = a, L = l) dF_l(l)$$

wobei $F_L(\cdot)$ die Verteilungsfunktion bezüglich L beschreibt. Im diskreten Fall gilt

$$E(Y^a) = \sum_l E(Y|A = a, L = l) \times P(L = l).$$

g-formula

Konzept

- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

i	L	A	Y
1	0	0	0
2	0	0	1
3	0	0	0
4	0	1	0
5	0	1	1
6	0	1	1
7	0	0	0
8	1	0	0
9	1	1	0
10	1	1	1
11	1	0	1
12	1	0	1

g-formula

Konzept

Manuelle Berechnung

Berechnung mit R

...schneller

Berechnung mit Stata

Konfidenzintervalle

Fallbeispiel / Übung

Daten

Kausale Frage

Strukturannahmen

Identifikation

Schätzung

Interpretation

Offene Fragen

Literaturnachweise

Einfaches Ablesen aus der Tabelle ermöglicht uns folgende Rechnung:

$$\begin{aligned} E(Y^1) &= \sum_{l=0}^1 P(Y = 1|A = 1, L = l) \times P(L = l) \\ &= P(Y = 1|A = 1, L = 0) \times P(L = 0) + \\ &\quad P(Y = 1|A = 1, L = 1) \times P(L = 1) \\ &= \frac{2}{3} \cdot \frac{7}{12} + \frac{1}{2} \cdot \frac{5}{12} = \frac{43}{72} \approx 0.5972 \end{aligned}$$

Analog erhält man

$$E(Y^0) = 0.42361$$

und damit

$$\psi_{ATE} = 0.5972 - 0.42361 = 0.17359$$

Frage: Was ist die Interpretation von ψ_{ATE} ?

g-formula

Konzept

Manuelle Berechnung

Berechnung mit R

...schneller

Berechnung mit Stata

Konfidenzintervalle

Fallbeispiel / Übung

Daten

Kausale Frage

Strukturannahmen

Identifikation

Schätzung

Interpretation

Offene Fragen

Literaturnachweise

Idee:

- 1 Schätze $P(Y = 1|A, L)$ über logistische Regression (Zusammenhang in den beobachteten Daten):
- 2 Schätze $P(L = l)$ über die empirische Verteilung (= behalte Daten der Kovariablen)
- 3 Setze $A = 1$ (interveniere! überschreibe A !)
- 4 Schätze¹ $P(Y^1 = 1) = P(Y = 1|A = 1, L = l)$ über Vorhersage (Zielgröße unter Intervention)
- 5 Berechne den ATE über $P(Y^1 = 1) - P(Y^1 = 0)$

g-formula

Konzept
Manuelle Berechnung
Berechnung mit R
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

¹ unter der Verwendung der Annahmen von Kapitel 3

Einfaches Beispiel – Schätzung mit R I

g-formula

Konzept
Manuelle Berechnung
Berechnung mit R
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

```
1 > # Daten
2 > Y <- c(0,1,0,0,1,1,0,0,0,1,1,1)
3 > A <- c(0,0,0,1,1,1,0,0,1,1,0,0)
4 > L <- c(0,0,0,0,0,0,0,0,1,1,1,1)
5 >
6 > # Schritt 1: Logistische Regression
7 > m1 <- glm(Y~A+L+A:L, family=binomial)
8 > summary(m1)
9
10 Coefficients:
11             Estimate Std. Error z value Pr(>|z|)
12 (Intercept)  -1.099      1.155  -0.951   0.341
13 A              1.792      1.683   1.064   0.287
14 L              1.792      1.683   1.064   0.287
15 A:L           -2.485      2.517  -0.987   0.323
16
17 >
18 >
19 > # Schritt 2+3: Behalte L, setze A=1 und A=0
20 > newdata1 <- as.data.frame(cbind(Y,A,L))
21 > newdata1$A <- c(rep(1,12))
22 > newdata0 <- as.data.frame(cbind(Y,A,L))
23 > newdata0$A <- c(rep(0,12))
24 >
```

```
25 > # Schritt 4. Vorhersage, gegeben die Intervention
26 > predict(m1, type="response", newdata=newdata1) # Y^1
27           1           2           3           4           5 ...
28 0.6666667 0.6666667 0.6666667 0.6666667 0.6666667 ...
29 > predict(m1, type="response", newdata=newdata0) # Y^0
30           1           2           3           4           5 ...
31 0.2500000 0.2500000 0.2500000 0.2500000 0.2500000 ...
32 >
33 > # Schritt 5: Berechne ATE ( $E(Y^1) - E(Y^0)$ )
34 > mean(predict(m1, type="response", newdata=newdata1)) -
35   mean(predict(m1, type="response", newdata=newdata0))
36 [1] 0.1736111 #korrekt!
```

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
 - ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

- 1** Schätze $P(Y = 1|A, L)$ über logistische Regression (Zusammenhang in den beobachteten Daten):

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.099	1.155	-0.951	0.341
A	1.792	1.683	1.064	0.287
L	1.792	1.683	1.064	0.287
A:L	-2.485	2.517	-0.987	0.323

- 2** ...[behalte L]

- 3** Setze $A = 1$

- 4** Berechne $P(Y^1 = 1) = P(Y = 1|A = 1, L = l)$ unter der Intervention:

$$\frac{1}{1 + \exp(-[-1.099 + 1.792 \cdot 1 + 1.792 \cdot l - 2.485 \cdot 1 \cdot l])}$$

g-formula

Konzept

Manuelle Berechnung

Berechnung mit R

...schneller

Berechnung mit Stata

Konfidenzintervalle

Fallbeispiel / Übung

Daten

Kausale Frage

Strukturannahmen

Identifikation

Schätzung

Interpretation

Offene Fragen

Literaturnachweise

...das ergibt:

i	L	A	Y	Y^0	Y^1
1	0	1	0		$\frac{2}{3}$
2	0	1	1		$\frac{2}{3}$
3	0	1	0		$\frac{2}{3}$
4	0	1	0		$\frac{2}{3}$
5	0	1	1		$\frac{2}{3}$
6	0	1	1		$\frac{2}{3}$
7	0	1	0		$\frac{2}{3}$
8	1	1	0		$\frac{1}{2}$
9	1	1	0		$\frac{1}{2}$
10	1	1	1		$\frac{1}{2}$
11	1	1	1		$\frac{1}{2}$
12	1	1	1		$\frac{1}{2}$

g-formula

Konzept

Manuelle Berechnung

Berechnung mit R

...schneller

Berechnung mit Stata

Konfidenzintervalle

Fallbeispiel / Übung

Daten

Kausale Frage

Strukturannahmen

Identifikation

Schätzung

Interpretation

Offene Fragen

Literaturnachweise

...und analog für $P(Y^0 = 1)$

i	L	A	Y	Y^0	Y^1
1	0	0	0	$\frac{1}{4}$	$\frac{2}{3}$
2	0	0	1	$\frac{1}{4}$	$\frac{2}{3}$
3	0	0	0	$\frac{1}{4}$	$\frac{2}{3}$
4	0	0	0	$\frac{1}{4}$	$\frac{2}{3}$
5	0	0	1	$\frac{1}{4}$	$\frac{2}{3}$
6	0	0	1	$\frac{1}{4}$	$\frac{2}{3}$
7	0	0	0	$\frac{1}{4}$	$\frac{2}{3}$
8	1	0	0	$\frac{2}{3}$	$\frac{1}{2}$
9	1	0	0	$\frac{2}{3}$	$\frac{1}{2}$
10	1	0	1	$\frac{2}{3}$	$\frac{1}{2}$
11	1	0	1	$\frac{2}{3}$	$\frac{1}{2}$
12	1	0	1	$\frac{2}{3}$	$\frac{1}{2}$

g-formula

Konzept

Manuelle Berechnung

Berechnung mit *R*

...schneller

Berechnung mit Stata

Konfidenzintervalle

Fallbeispiel / Übung

Daten

Kausale Frage

Strukturannahmen

Identifikation

Schätzung

Interpretation

Offene Fragen

Literaturnachweise

Ergebnis: identisch!

$$E(Y^0) = \frac{1}{4} \cdot \frac{7}{12} + \frac{2}{3} \cdot \frac{5}{12} = 0.42361$$

$$E(Y^1) = \frac{2}{3} \cdot \frac{7}{12} + \frac{1}{2} \cdot \frac{5}{12} = 0.59722$$

$$\psi_{ATE} = 0.5972 - 0.42361 = 0.17359$$

g-formula

Konzept
Manuelle Berechnung

Berechnung mit R

...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

```
1 library(ltmle)
2 daten <- as.data.frame(cbind(L,A,Y))
3 m2 <- ltmle(daten,                                # data (ordered)
4             Lnodes="L", Anodes="A", Ynodes="Y",
5             abar=list(1,0),                        # intervention(s)
6             gcomp=T,                               # method (i.e. g-comp.)
7             Qform=c(Y="Q.kplus1 ~ L + A + A:L")    # regression formula
8             )
9
10 summary(m2)
11
12 # Ergebnis
13 Additive Treatment Effect:
14   Parameter Estimate:  0.17361
15   Estimated Std Err:  0.3002
16   p-value:  0.5747
17   95% Conf Interval: (-0.48712, 0.83435)
```

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R

...schneller

- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

- g-formula manuell auch einfach in Stata² und SAS möglich

Frage: wie wird das prinzipiell in Stata gemacht?

- In *R* gibt es einige Pakete.

- Das Paket `ltmle` ist eigentlich für eine andere Methode (TMLE) konzipiert; aber `ltmle` erlaubt den Vergleich verschiedener Methoden und funktioniert auch für Longitudinaldaten und zensierte Daten (es lohnt sich also, damit zu beginnen...)

- wir bleiben bei einfachen Beispielen, aber komplexe Fragen können prinzipiell damit gelöst werden

g-formula

Konzept
Manuelle Berechnung
Berechnung mit *R*

...schneller

Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

²siehe Beispiel auf Folie 30

g-formula

Konzept
Manuelle Berechnung
Berechnung mit *R*
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

```
1 import delimited C:\Users\schomakm\Dropbox\Documents\  
   Teaching\Kausale_Inferenz\Code\cancer_data.csv  
2  
3 encode w1, generate(age)  
4 encode w2, generate(ses)  
5 encode w3, generate(comorb)  
6 encode w4, generate(stage)  
7 drop v1 id  
8  
9 //step 1: regression model  
10 logistic y i.age i.ses i.comorb i.stage i.a i.a##i.ses i.a  
   ##i.comorb  
11  
12 //step 2,3,4 combined  
13 margins a  
14  
15 display ((.1804233)/(1-.1804233))/((.3652724)/(1-.3652724))  
   )
```

- prinzipiell über Bootstrapping
- d.h. etwas aufwändiger
- Der Ansatz in `ltmle` ist nicht ganz exakt für die g-formula
- wir gehen aus Zeitgründen hier erstmal nicht ins Detail

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit *R*
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

“[...] we are looking at an example from cancer epidemiology. In this example, we are interested in the effect of dual treatment therapy (radio- and chemotherapy), compared to single therapy (chemotherapy only) on the probability of one-year survival among colorectal cancer patients. We know that there are confounders which affect both treatment assignment and the outcome, namely clinical stage, socioeconomic status, co-morbidities, and age. Evidence shows that older patients with co-morbidities have a lower probability of being offered more aggressive treatments and therefore they usually get less effective curative options. Also, colorectal cancer patients from lower socioeconomic status have a higher probability of presenting an advanced clinical stage at diagnosis, thus they usually get offered only palliative treatments.”

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit *R*
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

Daten

- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

	ID	w1	w2	w3	w4	a	y
1	1	old	2_middle	other	stage 2	0	1
2	2	old	3_lower middle	other	stage 1	0	0
3	3	old	1_upper middle	diabetes	stage 3	0	1
4	4	young	5_non-working	none	stage 1	0	0
5	5	old	3_lower middle	diabetes	stage 1	1	0
6	6	young	2_middle	diabetes	stage 2	0	1
7	7	old	3_lower middle	none	stage 3	0	0
8	8	young	2_middle	none	stage 2	1	0
9	9	old	3_lower middle	none	stage 4	1	1
10	10	old	5_non-working	none	stage 1	0	0
11	11	old	4_working	none	stage 2	1	0
12	12	young	2_middle	diabetes	stage 1	1	0
13	13	old	3_lower middle	diabetes	stage 1	1	1
14	14	young	4_working	hypertension	stage 2	0	0
15	15	young	2_middle	none	stage 1	0	0
16	16	old	3_lower middle	diabetes	stage 4	1	1
17	17	young	3_lower middle	none	stage 3	1	0
18	18	old	4_working	none	stage 4	1	0

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit *R*
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise



Aus Luque-Fernandez, Redondo Sanchez und Schomaker [3]:

“A clinician may be interested in the following: how different would the risk/chance of death have been had everyone received dual therapy compared to if everyone had experienced monotherapy? ”

Frage: Was ist eine sinnvolle target quantity?

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

Daten

Kausale Frage

- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise



Aus Luque-Fernandez, Redondo Sanchez und Schomaker [3]:
“A clinician may be interested in the following: how different would the risk/chance of death have been had everyone received dual therapy compared to if everyone had experienced monotherapy? ”

Frage: Was ist eine sinnvolle target quantity?

Zum Beispiel: $\psi_{MOR} = \frac{P(Y^1=1)/P(Y^1=0)}{P(Y^0=1)/P(Y^0=0)}$

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

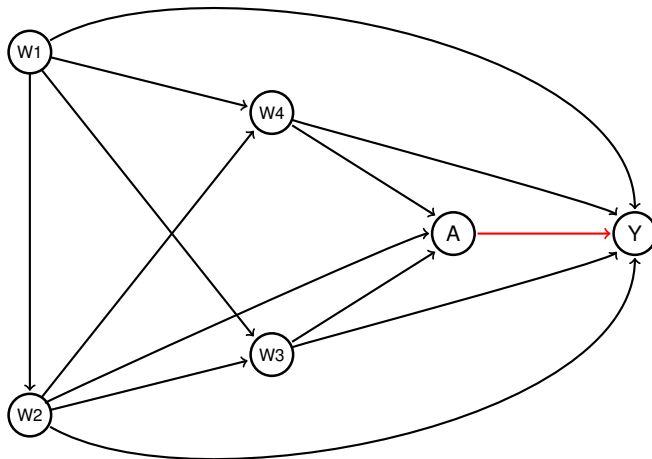
Fallbeispiel / Übung

- Daten

Kausale Frage

- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise



g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen**
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

Y = Tod; A = Chemotherapie vs. Chemotherapie & Strahlentherapie;
 W_1 = Alter; W_2 = Sozioökonomischer Status; W_3 =
Begleiterkrankungen; W_4 = Schwere der Krankheit (stage)



Frage: Ist die Annahme der Konsistenz hier erfüllt?

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise



Frage: Wenn die Strukturannahmen korrekt sind, welche Variablen müssen wir mindestens in die Analyse einschließen?

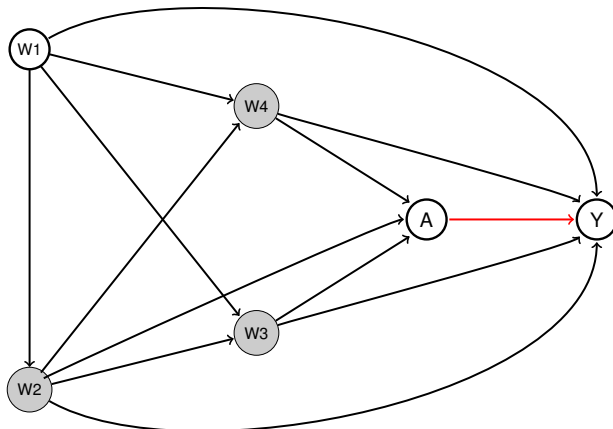
g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit *R*
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation**
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise



g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

(aber auch alle 4 Variablen können eingeschlossen werden/
keine Collider, Mediatoren)



Frage: Ist die Annahme der Positivität erfüllt?

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

```
, , w2 = 1_upper middle, w3 = none, w4 = stage 3
```

```
      w1
a    young old
0      1    3
1      5    4
```

```
, , w2 = 2_middle, w3 = none, w4 = stage 3
```

```
      w1
a    young old
0     13   17
1     25   26
```

```
, , w2 = 3_lower middle, w3 = none, w4 = stage 3
```

```
      w1
a    young old
0     17   35
1     56   72
```

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit *R*
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation**
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise

- wird benötigt, dass die bedingten Erwartungswerte in der Definition sinnvoll definiert sind
- bei mehreren (stetigen) Kovariablen oft nicht zu 100% erfüllt
- immerhin: die g-formula extrapoliert in “leere” Regionen basierend auf den Modellannahmen
- solange $E(Y|A = a, L = l)$ korrekt spezifiziert ist, ist es in Ordnung; sonst nicht.
- komplexere Methoden, die doppelt robust sind (z.B. TMLE), sind in der Regel besser wenn Positivitätsverletzungen vorliegen

g-formula

Konzept
Manuelle Berechnung
Berechnung mit R
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise

Das heißt, der kausale Effekt (also: MOR) von der Therapieform auf Mortalität ist unter den gegebenen Annahmen identifiziert und kann geschätzt werden, wenn wir die entsprechenden Daten gemessen haben. Für gute Ergebnisse benötigen wir eine gute Modellierung, auch wegen der Positivität.

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung
- Interpretation
- Offene Fragen

Literaturnachweise



g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation

Schätzung

- Interpretation
- Offene Fragen

Literaturnachweise

```
1 # Daten einlesen und laden
2 setwd("C:/Users/schomakm/Dropbox/Documents/Teaching/
   Kausale_Inferenz/Code")
3 cdata <- read.csv("cancer_data.csv")
4 attach(cdata)
5
6 # Verwende ltmle zur Anwendung der g-formula
7 library(ltmle)
8 summary(ltmle(cdata, Lnodes=c("w1", "w2", "w3", "w4"),
9                               Anodes="a", Ynodes="y",
10                               abar=list(1,0),
11                               gcomp=T,
12                               Qform=c(y="Q.kplus1 ~ w1+w2+w3+w4+a*w2+a*w3"
13                                     )
14 ))
15 # Ergebnis
16
17 Parameter Estimate: 0.38254
18 Est Std Err log(OR): 0.098287
19 p-value: <2e-16
20 95% Conf Interval: (0.31551, 0.46381)
```

Geht auch in Stata... (sogar einfacher)

```
. logistic y i.age i.ses i.comorb i.stage i.a i.a##i.ses i.a##i.comorb
```

```
Logistic regression               Number of obs   =       5,000
                                LR chi2(23)      =       749.48
                                Prob > chi2       =       0.0000
Log likelihood = -2386.4117       Pseudo R2    =       0.1357
```

	y	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age						
	young	.4675325	.0359764	-9.88	0.000	.4020797 .5436401
ses						
	2_middle	1.545988	.4636768	1.45	0.146	.8588383 2.78292
	3_lower middle	1.779709	.5140104	2.00	0.046	1.010433 3.134661
	4_working	1.758222	.5476087	1.81	0.070	.9549093 3.237319
	5_non-working	2.138082	.7144742	2.27	0.023	1.110662 4.11592

```
. margins a
```

```
Predictive margins               Number of obs   =       5,000
Model VCE      : OIM
```

```
Expression      : Pr(y), predict()
```

		Delta-method		z	P> z	[95% Conf. Interval]	
	Margin	Std. Err.					
a							
0	.3652724	.0124018	29.45	0.000	.3409654	.3895794	
1	.1804233	.0059066	30.55	0.000	.1688466	.1919999	

```
. display ((.1804233)/(1-.1804233))/((.3652724)/(1-.3652724))
.38253708
```

g-formula

Konzept
Manuelle Berechnung
Berechnung mit R
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise



Frage: was ist die Interpretation des Ergebnisses?

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit *R*
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung

Interpretation

- Offene Fragen

Literaturnachweise

Unter der Annahme, dass unsere Strukturannahmen korrekt sind, und dass wir die Modelle korrekt spezifiziert haben, können wir folgendes sagen: die Chance zu sterben ist für Patienten ca. 0.38 mal geringer wenn sie mit dualer Therapie behandelt werden im Vergleich zu Monotherapie.

Frage: für welche Patientengruppe gilt die Schlussfolgerung?

g-formula

- Konzept
- Manuelle Berechnung
- Berechnung mit R
- ...schneller
- Berechnung mit Stata
- Konfidenzintervalle

Fallbeispiel / Übung

- Daten
- Kausale Frage
- Strukturannahmen
- Identifikation
- Schätzung

- Interpretation**
- Offene Fragen

Literaturnachweise



- Modellierung: wir nehmen an, dass unsere Regressionsmodelle korrekt spezifiziert werden.

Frage: was ist das Beste, das wir hier tun können?

- Was tun wir bei fehlenden Daten? Wir intervenieren und zwingen jeden Patienten zu kommen!

Frage: was machen wir dann praktisch? Welche Annahmen brauchen wir?

g-formula

Konzept
Manuelle Berechnung
Berechnung mit R
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation

Offene Fragen

Literaturnachweise



- [1] J. Robins.
A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect.
Mathematical Modelling, 7(9-12):1393–1512, 1986.

- [2] M. A. Luque-Fernandez, M. Schomaker, B. Rachet, and M. E. Schnitzer.
Targeted maximum likelihood estimation for a binary treatment: A tutorial.
Statistics in Medicine, 37(16):2530–2546, 2018.

- [3] M. A. Luque-Fernandez, D. Redondo-Sanchez, and M. Schomaker.
Effect modification and collapsibility in evaluations of public health interventions.
American Journal of Public Health, 109(3):e12–e13, 2019.

- [4] J. M. Snowden, S. Rose, and K. M. Mortimer.
Implementation of g-computation on a simulated data set: demonstration of a causal inference technique.
American Journal of Epidemiology, 173(7):731–8, 2011.

g-formula

Konzept
Manuelle Berechnung
Berechnung mit *R*
...schneller
Berechnung mit Stata
Konfidenzintervalle

Fallbeispiel / Übung

Daten
Kausale Frage
Strukturannahmen
Identifikation
Schätzung
Interpretation
Offene Fragen

Literaturnachweise