

Kausale Inferenz

Kapitel 8: Matching

Sommersemester 2017

Version: 16. Juni 2017

Michael Schomaker

University of Cape Town, CIDER, South Africa

Ludwig-Maximilians-Universität München, Institut für Statistik



- Hunderdtausende Publikationen zu Matching!
- Entwickelt unter anderem von Donald Rubin
- Sehr bekannt in den Sozialwissenschaften
- Soll Abhängigkeit von Modellspezifikation verringern

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Unter Konsistenz haben wir folgenden Datensatz:

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



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Finde die kontrafaktischen Werte in den Daten
- Nehme die Behandelten ($A=1$) und finde Kontrollen ($A=0$) mit gleichen Kovariablen
 - $L_{A=1} = L_{A=0} \rightarrow$ exaktes matchen
 - $L_{A=1} \approx L_{A=0} \rightarrow$ matchen
- Y_i^0 für Person i mit $A = 1$ ist Y_j von Person j mit (ungefähr) denselbsen Kovariablen
- Lösche Beobachtungen, die nicht gematcht wurden
- Dies stellt Austauschbarkeit her, da die Verteilung bezüglich L für $A = 1$ und $A = 0$ identisch ist
- Schätzt den ATT!

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching



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

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Idee: Exaktes Matching (Finaler Datensatz)

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			
8	1	0	0
9	1	1	0
10	1	1	1
11	1	0	1
12			



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

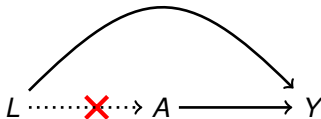
Literaturnachweise

- Der Datensatz reduziert sich von 12 auf 10 Beobachtungen
- Der Datensatz ist bezüglich L “ausbalanciert”:

- $P(L|A = 1) = \frac{2}{5}; \quad P(L|A = 0) = \frac{2}{5}$

- $P(A|L = 1) = \frac{1}{2}; \quad P(A|L = 0) = \frac{1}{2}$

- Damit ist Austauschbarkeit erreicht:



- “Experiment” in den Daten: Assoziation = Kausalität



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

- Lösung hängt von der konkreten Zuordnung der Gruppen ab (wir hätten auch andere Gruppen bilden können!)

- Hier: $ATT = \frac{3}{5} - \frac{2}{5} = 20\%$ (korrekt wäre 18.3%)



$$\begin{aligned}\psi_{\text{ATT}} &= E(Y^1|A=1) - E(Y^0|A=1) \\ &\stackrel{\text{Kons.}}{=} E(Y|A=1) - E(Y^0|A=1) \\ &= E(Y|A=1) - \sum_l E(Y^0|L=l, A=1)P(L=l|A=1) \\ &\stackrel{\text{Aust.}}{=} E(Y|A=1) - \sum_l E(Y^0|L=l, A=0)P(L=l|A=1) \\ &\stackrel{\text{Kons.}}{=} E(Y|A=1) - \sum_l E(Y|L=l, A=0)P(L=l|A=1)\end{aligned}$$

Exaktes Matching heißt, dass die Verteilung der Kovariablen der Behandelten ($A=1$), der Verteilung der ausgewählten Kontrollen (I_c) entspricht.

$$P(L=l|A=1) \stackrel{!}{=} P(L=l|A=0, I_c)$$

Damit ergibt sich:

$$\psi_{\text{ATT}} = E(Y|A=1) - \sum_l E(Y|L=l, A=0)P(L=l|A=0, I_c)$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Um den ATE zu schätzen, müssen wir folgendes tun:

- ψ_{ATT} berechnen, wie oben beschrieben
- ψ_{ATC} berechnen:
 - Matche Kontrollen ($A=0$) zu Behandelten ($A=1$)
 - Vertausche Nuller und Einser im Datensatz (für Software)
 - Schätzt den Effekt für $A = 0$ mit Codierung $A = 1$, d.h. setze Minuszeichen vor Effektschätzung
- Dann ist es einfach den ATE zu schätzen:

$$\psi_{ATE} = P(A = 1) \cdot \psi_{ATT} + P(A = 0) \cdot \psi_{ATC}$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

- Exaktes Matching
- Vergrößertes Exaktes Matching
- Nearest Neighbour
- Andere Methoden

Überprüfung des Matching

- Selektionsbias
- Balance
- Propensity Scores

Weiteres

- Konfidenzintervalle
- Modellspezifikation
- Sonstiges

Literaturnachweise

- 1 Wähle aus, welche Beobachtungen gematcht werden sollen (Behandelte oder Kontrollen)
- 2 Wähle ein Abstandsmaß
- 3 Finde “matches” auf Basis des Abstandsmaßes zu den Beobachtungen aus 1)
- 4 Lösche die ungematchten Beobachtungen
- 5 Überprüfe, ob Verteilung der Kovariablen ausbalanciert (bezüglich A)
- 6 Evtl: wiederhole 1-5 um Balance zu verbessern
- 7 Führe Analyse auf reduziertem Datensatz durch (Assoziation = Kausalität bei perfekter Balancierung)



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

- Exaktes Matching
- Vergrößertes Exaktes Matching
- Nearest Neighbour
- Andere Methoden

Überprüfung des Matching

- Selektionsbias
- Balance
- Propensity Scores

Weiteres

- Konfidenzintervalle
- Modellspezifikation
- Sonstiges

Literaturnachweise

- matche nicht $1 : 1$, sondern $1 : k$ – weise also jeder Beobachtung mit $A = 1$ potentiell mehrere Kontrollen zu (und verwende mittleres Y)
- verwende Kontrollen potentiell für mehrere Beobachtungen mit $A = 1$ (→ Gewichtung, siehe unten)
- matche exakt $k : k$, definiere die Gruppen also als all diejenigen Beobachtungen mit $A = 0/1$, die die gleichen Kovariablenausprägungen haben
- verwende ein match nur dann, wenn er wirklich gut ist (geringer Abstand); und reduziere dann den Datensatz damit weiter falls nötig (ATT ist nicht mehr Zielparameter!)



Sei $\mathbf{L} = \{L_1, \dots, L_k\} = L$ der Kovariablenvektor. Exaktes Matching zweier Beobachtungen i und j basiert auf dem Abstandsmaß:

$$D_{1,ij} = \begin{cases} 0 & \text{wenn } L_i = L_j \\ \infty & \text{wenn } L_i \neq L_j \end{cases}$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Sei $\mathbf{L} = \{L_1, \dots, L_k\} = L$ der Kovariablenvektor.
Mahalanobis-Matching zweier Beobachtungen i und j basiert auf dem Abstandsmaß:

$$D_{2,ij} = (L_i - L_j)' \Sigma_c^{-1} (L_i - L_j)$$

Σ_c bezeichnet dabei die Kovarianzmatrix in der Kontrollgruppe ($A=0$).

(Hochkorrelierte Variablen sollen geringer gewichtet in die Distanzberechnung einfließen)

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

- Exaktes Matching
- Vergrößertes Exaktes Matching
- Nearest Neighbour
- Andere Methoden

Überprüfung des Matching

- Selektionsbias
- Balance
- Propensity Scores

Weiteres

- Konfidenzintervalle
- Modellspezifikation
- Sonstiges

Literaturnachweise



Sei $\mathbf{L} = \{L_1, \dots, L_k\} = L$ der Kovariablenvektor.
Propensity-Score-Matching zweier Beobachtungen i und j
basiert auf dem Abstandsmaß:

$$D_{3,ij} = P(A_i = 1|L_i) - P(A_j = 1|L_j)$$

bzw.

$$D_{4,ij} = \log \left(\frac{P(A_i = 1|L_i)}{1 - P(A_i = 1|L_i)} \right) - \log \left(\frac{P(A_j = 1|L_j)}{1 - P(A_j = 1|L_j)} \right)$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- exaktes 1:1 Matching wie oben beschrieben
- k:k Matching: verschmelze alle Behandelten und Kontrollen mit gleicher Kovariablenkombination in einer Klasse s , wobei $s = 1, \dots, S$.

- Adjustiere dann mit folgenden Gewichten [1]:

$$w_i = \begin{cases} 1 & \text{wenn } A_i = 1 \\ \frac{n_c}{n_t} \times \frac{n_{t,s}}{n_{c,s}} & \text{wenn } A_i = 0 \end{cases}$$

(n_c = Anzahl der Kontrollen; n_t = Anzahl der Behandelten;
 s = Klasse s)

- k:k matching ist “exaktes matchen” im *R*-Paket `MatchIt` [2]

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Beispiel:

i	Y	A	L	Gewichte	Klasse
1	0	0	0	1.05	1
2	1	0	0	1.05	1
3	0	0	0	1.05	1
4	0	1	0	1.00	1
5	1	1	0	1.00	1
6	1	1	0	1.00	1
7	0	0	0	1.05	1
8	0	0	1	0.93	2
9	0	1	1	1.00	2
10	1	1	1	1.00	2
11	1	0	1	0.93	2
12	1	0	1	0.93	2

$$w_1 = \begin{cases} 1 & \text{wenn } A_i = 1 \\ \frac{7}{5} \times \frac{3}{4} = 1.05 & \text{wenn } A_i = 0 \end{cases}$$



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



```
library(MatchIt)

# ATT
matchm <- matchit(A ~ L, data = sdata1, method = "exact")
mam <- glm(Y~A, data=match.data(matchm), weights=weights)

# ATC
matchm2 <- matchit(A~L, data = sdata2, method = "exact")
mam2 <- glm(Y~A, data=match.data(matchm2), weights=weights)

p1*coefficients(mam)[2] + p0*-coefficients(mam2)[2]
```

Liefert korrektes Ergebnis: 0.1736 (!)

Beispiel: Krebs (kategoriale Daten!)

$$\psi_{\text{OR}} = 0.373$$

$$\hat{\psi}_{\text{OR, exaktes matching}} = 0.376$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Teile stetige Variablen (oder kategoriale Variablen mit vielen Klassen) in grobe Klassen auf
- *matche exakt* mit Bezug auf diese Klassen
- durch Kategorisierung geht Information verloren, aber (bei guter Datenlage) verbesserte Balance bezüglich L möglich: → wichtig für kausale Interpretation
- Nach dem Matching wird die Vergrößerung wieder aufgehoben
- beim Krebsbeispiel identische Resultate, da kategorial

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

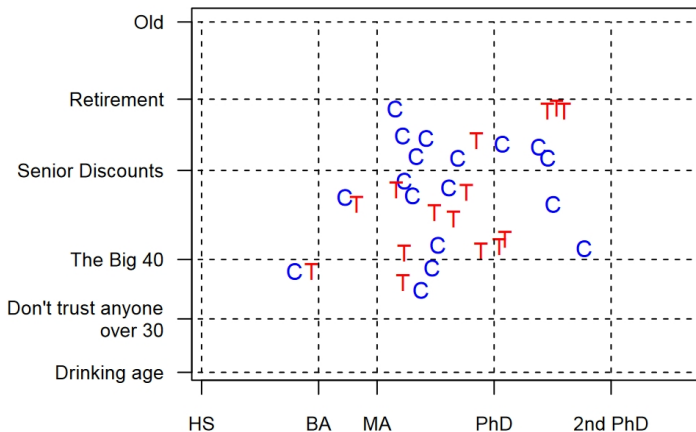
Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Matche so, dass $f(L_{A=1}) \approx f(L_{A=0})$
- Das heisst, die Annahme die zur Identifizierung des ATT (und damit ATE) kann verletzt sein
- Wie gut die Kovariablenverteilung nach NN-Matching passen, muss auf alle Fälle überprüft werden (siehe unten)
- Es wird argumentiert, dass bei kleineren Unperfektheiten einfach per Regression nach“adjustiert” werden kann
- diese Argumentation ist nicht hieb- und stichfest (siehe Kapitel 7)

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Die Kontrollbeobachtung j , die zur Beobachtung i gematcht wird, ist diejenige, für die das Distanzmaß $D_{2,ij}$ (oder $D_{3,ij}$, $D_{4,ij}$ etc.) minimiert wird
- Gibt es mehrere Beobachtungen, für die der Abstand identisch ist, kann entweder zufällig eine der konkurrierenden Beobachtungen ausgewählt werden, oder die nächste
- Je nach Entscheidung können die Ergebnisse variieren!

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Man kann eine Grenze c ("caliper") definieren, die festlegt, ab welchem Abstand nicht mehr gematcht werden soll; z.B.:

$$D_{1,ij}^* = \begin{cases} (L_i - L_j)' \Sigma_c^{-1} (L_i - L_j) & \text{wenn } |\text{logit}(P(A_i = 1|L_i)) - \text{logit}(P(A_j = 1|L_j))| \leq c \\ \infty & \text{wenn } |\text{logit}(P(A_i = 1|L_i)) - \text{logit}(P(A_j = 1|L_j))| > c \end{cases}$$

Problem: Beobachtungen werden ausgeschlossen, damit Analyse einer schwer zu definierenden Subpopulation \rightarrow ATT wird evtl. nicht korrekt geschätzt

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Sollen wir nur strikt die Beobachtungen matchen, die sehr nahe zusammenliegen (gemessen mit dem Distanzmaß)?
 - Gute Idee, um Annahmen der Identifizierung zu erfüllen (→ kein Bias)
 - ...aber eventuell werden keine passenden matches gefunden (→ Selektionsbias)
- Sollen wir die Kriterien lockerer definieren?
 - Schlechte Idee, um Annahmen der Identifizierung zu erfüllen (→ Bias)
 - ...aber eventuell werden alle Beobachtungen gematched (→ kein Selektionsbias)

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Matchen mit Zurücklegen

- Bei Schätzung des ATEs wird es fast immer entweder mehr Kontrollen oder mehr Behandelte geben ($n_t \neq n_c$)
- Möglichkeit 1: matche nicht alle Beobachtungen

Beispiel: Krebsdaten

	Control	Treated
All	1832	3168
Matched	1832	1832
Unmatched	0	1336
Discarded	0	0

→ ATT nicht unbedingt identifiziert!

- Möglichkeit 2: Ziehe mit Zurücklegen, matche also Kontrollbeobachtungen ggf. häufiger als einmal

Beispiel: Krebsdaten

	Control	Treated
All	1832	3168
Matched	1107	3168
Unmatched	725	0
Discarded	0	0



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Interpretation als 1:1 Matching, bei der Beobachtungen öfters verwendet werden (k_i mal). Adjustierung mit allgemeinen Gewichten wie oben beschrieben:

$$w_i = \begin{cases} 1 & \text{wenn } A_i = 1 \\ \frac{n_c}{n_t} \times \frac{n_{t,s}}{n_{c,s}} & \text{wenn } A_i = 0 \end{cases}$$

Bei 1:1 matching ist $\frac{n_{t,s}}{n_{c,s}} = 1$, aber Beobachtung i wird k_i mal verwendet; damit ist müssen

$$w_i = \begin{cases} 1 & \text{wenn } A_i = 1 \\ \frac{n_c}{n_t} \times k_i & \text{wenn } A_i = 0 \end{cases}$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Sample sizes:

	Control	Treated
All	7	5
Matched	4	5
Unmatched	3	0
Discarded	0	0

```
> match.data(matchm)
  Y A L distance weights
1  0 0 0 0.4285714      0.8
2  1 0 0 0.4285714      1.6
4  0 1 0 0.4285714      1.0
5  1 1 0 0.4285714      1.0
6  1 1 0 0.4285714      1.0
9  0 1 1 0.4000000      1.0
10 1 1 1 0.4000000      1.0
11 1 0 1 0.4000000      0.8
12 1 0 1 0.4000000      0.8
```

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



i	Y	A	L	weights
1	0	0	0	0.8
2	1	0	0	1.6
3	0	0	0	0
4	0	1	0	1
5	1	1	0	1
6	1	1	0	1
7	0	0	0	0
8	0	0	1	0
9	0	1	1	1
10	1	1	1	1
11	1	0	1	0.8
12	1	0	1	0.8

$$w_1 = \frac{4}{5} \times 1 = 0.8$$

$$w_2 = \frac{4}{5} \times 2 = 1.6$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- führt im Optimalfall zu Quasi-Randomisierung
- Interessant: fast alle anderen Methoden basieren auf bedingter Randomisierung (→ Kapitel 4), da innerhalb von Blöcken (Kovariablenschichten) gematcht wird!
- In der Regel ist bedingte Randomisierung effizienter
- Wir kennen auch die Probleme aus Kapitel 6, und es gibt noch mehr...siehe Video auf nächster Seite...
- Problem: unnötige Dimensionsreduktion
- Theorie baut auf wahrem propensity score auf, nicht geschätztem → teilweise Unklarheit

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

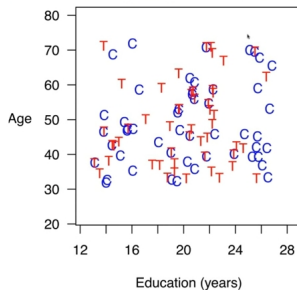
Literaturnachweise

“Why propensity score matching should not be used”

Hausaufgabe: Video schauen



Best Case: Propensity Score Matching is Suboptimal



15/23

35:31



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

<https://www.youtube.com/watch?v=rBv39pK1iEs>



- Es gibt eine Vielzahl weiterer Matching-Methoden [3]
- Wir haben nur eine kleine Auswahl behandelt
- “Optimal Matching”
 - matche nicht auf Basis der naheliegendsten Beobachtung (“greedy matching”)
 - versuche stattdessen *global* das Distanzmaß zu optimieren
- “Genetic Matching”
 - Methode für bessere Balancierung der Kovariablen
- ...

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Bei exaktem Matching können Fälle ausgeschlossen werden, wenn es in einer Gruppe nur Behandelte, oder nur Kontrollen gibt

	Control	Treated
All	1832	3168
Matched	1826	2535
Discarded	6	633

- Bei NN-Matching können Fälle ausgeschlossen werden, wenn ohne Zurücklegen gematcht wird und es weniger Kontrollen ($A = 0$) als Behandelte ($A = 1$) gibt

	Control	Treated
All	1832	3168
Matched	1832	1832
Unmatched	0	1336
Discarded	0	0

- Ein Ausschluß von Fällen kann immer zu Selektionsbias führen

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Grundidee war, dass $f(L|A = 1) \approx f(L|A = 0, I_c)$, damit ATT identifiziert ist
- idealerweise würde man die multivariate Verteilung anschauen; praktisch sind es jedoch einfach die empirischen Verteilungen jedes L_i
- Mittelwertsvergleich ist ein guter Start (sollte besser sein nach Matching)
- Median, Quantile, Maximum ect können auch angeschaut werden
- Vergleiche per Tests sind nicht optimal, deskriptiver Vergleich genügt

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Verggrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Summary of balance for all data:

	Means Treated	Means Control	Std. Mean Diff.
wlyoung	0.4031	0.4050	-0.0039
wlold	0.5969	0.5950	0.0039
w22_middle	0.2083	0.2385	-0.0744
w23_lower middle	0.4886	0.4864	0.0046
w24_working	0.1660	0.1468	0.0516
w25_non-working	0.1045	0.0775	0.0882
w3hypertension	0.1225	0.1397	-0.0527
w3diabetes	0.1622	0.1474	0.0403
w3other	0.2052	0.1463	0.1458
w3alzheimer	0.0991	0.0688	0.1015
w3heart attack	0.1177	0.0540	0.1976
w4stage 2	0.1847	0.3079	-0.3175
w4stage 3	0.2696	0.1201	0.3368
w4stage 4	0.2743	0.0076	0.5976

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Summary of balance for matched data:

	Means Treated	Means Control	Std. Mean Diff.
wlyoung	0.4031	0.4549	-0.1055
wlold	0.5969	0.5451	0.1055
w22_middle	0.2083	0.2254	-0.0420
w23_lower middle	0.4886	0.5038	-0.0303
w24_working	0.1660	0.1405	0.0687
w25_non-working	0.1045	0.1026	0.0062
w3hypertension	0.1225	0.0934	0.0886
w3diabetes	0.1622	0.1503	0.0325
w3other	0.2052	0.1976	0.0188
w3alzheimer	0.0991	0.1439	-0.1500
w3heart attack	0.1177	0.0919	0.0803
w4stage 2	0.1847	0.1840	0.0016
w4stage 3	0.2696	0.2715	-0.0043
w4stage 4	0.2743	0.2743	0.0000

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Percent Balance Improvement:

	Std. M. Diff.	eCDF Med	eCDF Mean	eCDF Max
wlyoung	-2584.4889	-843.7177	-843.7177	-843.7177
wlold	-2584.4889	-843.7177	-843.7177	-843.7177
w22_middle	43.5652	12.5092	12.5092	12.5092
w23_lower middle	-563.7681	-418.6632	-418.6632	-418.6632
w24_working	-33.1587	97.5019	97.5019	97.5019
w25_non-working	92.9780	22.9930	22.9930	22.9930
w3hypertension	-68.2210	45.1819	45.1819	45.1819
w3diabetes	19.3213	18.5409	18.5409	18.5409
w3other	87.1354	74.5216	74.5216	74.5216
w3alzheimer	-47.7419	40.2303	40.2303	40.2303
w3heart attack	59.3664	40.8685	40.8685	40.8685
w4stage 2	99.4876	5.4199	5.4199	5.4199
w4stage 3	98.7330	40.7981	40.7981	40.7981
w4stage 4	100.0000	1.9114	1.9114	1.9114

Hier: teilweise Verbesserung, teilweise nicht!

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Schaue statt allen L_i den niedrigdimensionaleren propensity score an (allgemein: das entsprechende Distanzmaß)
- Im gematchten Datensatz sollten sich die propensity score Verteilungen überlappen (“region of common support”)
- Bei Matchen mit Zurücklegen kann der Kreisradius das Gewicht der Beobachtung repräsentieren

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

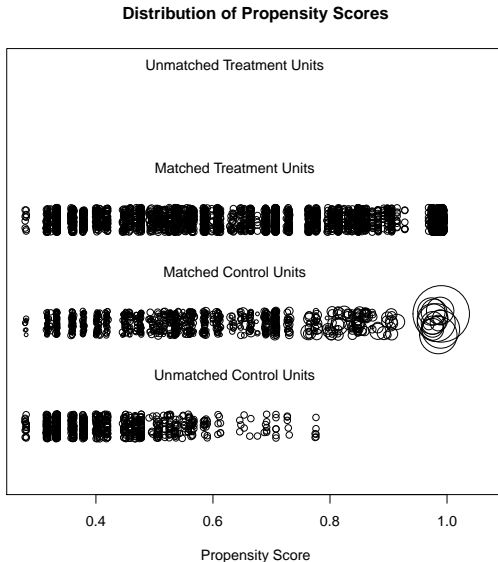
Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

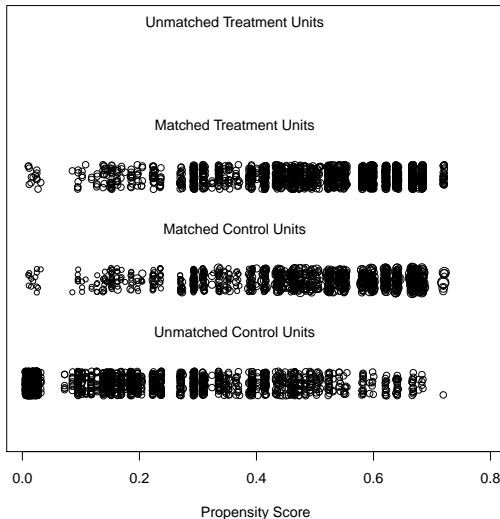
Konfidenzintervalle

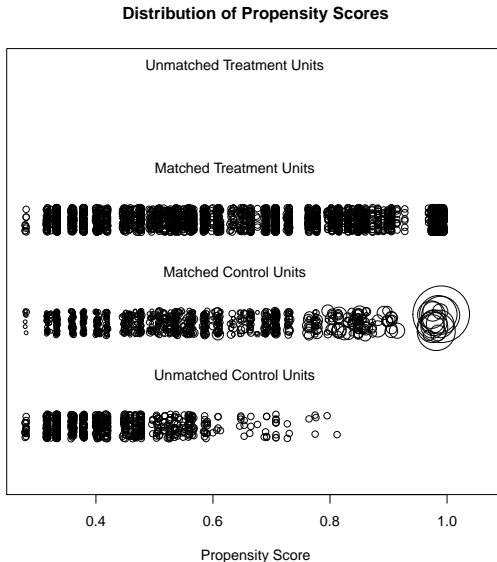
Modellspezifikation

Sonstiges

Literaturnachweise

Distribution of Propensity Scores





Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

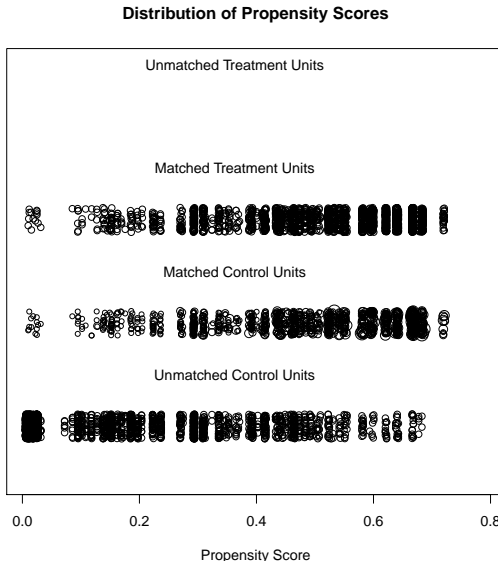
Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

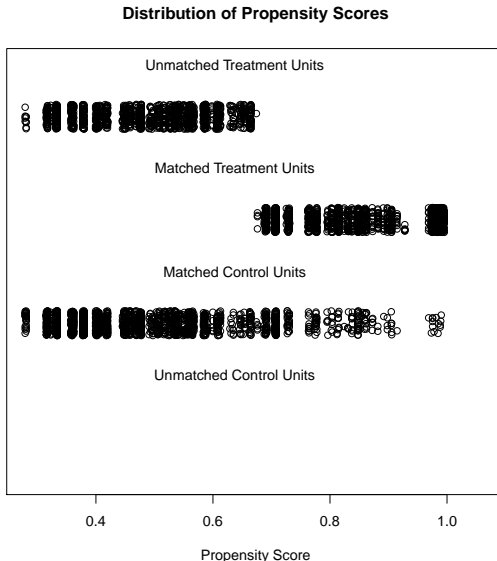
Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Propensity Score (ATT) – ohne Zurücklegen



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

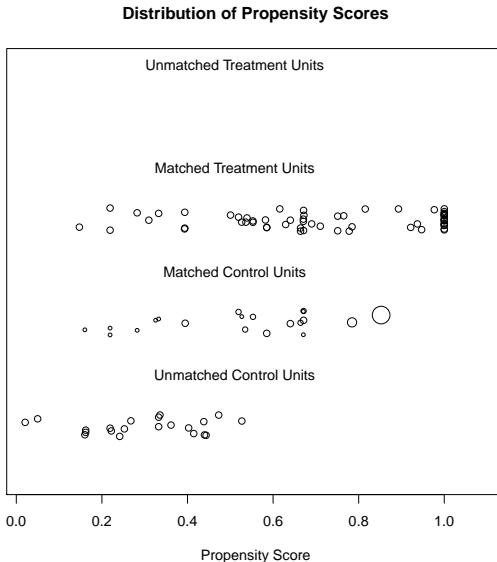
Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Propensity Score (ATT) – kleine Stichprobe ($n = 100$)



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Verwende Histogramme um Verteilung des propensity scores vor und nach dem Matchen zu vergleichen
- nach dem Matchen sollte die Verteilung von L , und damit auch vom propensity score, bei den Behandelten und den Kontrollen ähnlich sein

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

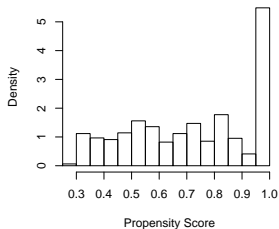
Sonstiges

Literaturnachweise

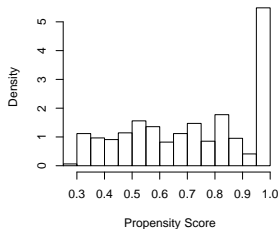
Propensity Score (ATT)



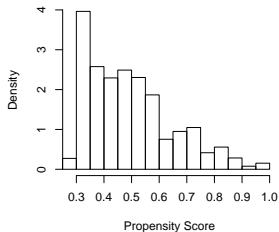
Raw Treated



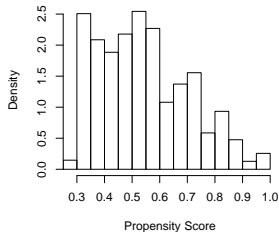
Matched Treated



Raw Control



Matched Control



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

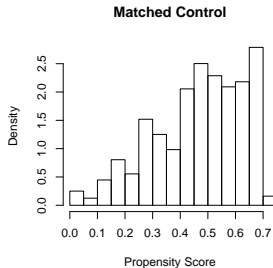
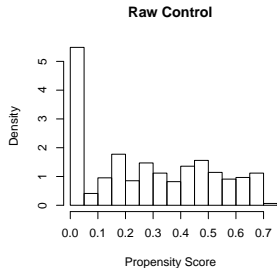
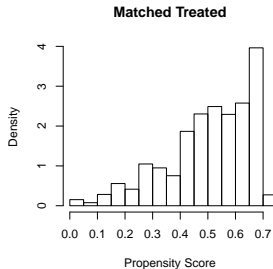
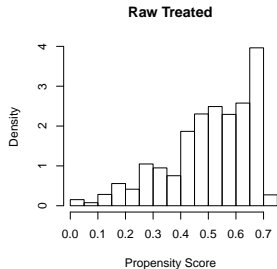
Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Propensity Score (ATC)



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Bisher keine Übereinkunft über bestes Vorgehen!
- Großer Nachteil von Matching
- Probleme:
 - Unsicherheit bezüglich der Schätzung des propensity scores
 - Unsicherheit bezüglich der (möglicherweise rekursiven) Wahl der Matching-Konfigurierung

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

- einfachste Option:
 - verwende gematchte Daten
 - verwende normale Analysemethode mit entsprechenden Standardfehlern
 - Beispiel: gewichtete Regression
- man betrachtet die gesamte Matching-Prozedur also als fix
- einfach, praktisch, allgemein anwendbar, aber nicht unbedingt korrekt



Varianz bei 1:1 exaktem Matching

Sei $Y_{i(k)}$ und $Y_{j(k)}$ die Zielgrößen zweier gematchten Beobachtungen i und j , die zusammen in der Klasse k gruppiert werden; und sei n_t die Anzahl der Beobachtungen mit $A = 1$; dann ist die Varianzschätzung des ATT als

$$\text{Var}(\hat{\psi}_{\text{ATT}, 1:1\text{EM}}) = \frac{1}{n_t - 1} \sum_{i=1}^{n_t} (Y_{i(k)} - Y_{j(k)} - \hat{\psi}_{\text{ATT}, 1:1\text{EM}})^2$$

gegeben.

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Beispiel: ...von oben

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

$$Y_{1(1)} - Y_{4(1)} - \hat{\psi} = 0 - (-0.2)$$

$$Y_{2(2)} - Y_{5(2)} - \hat{\psi} = 0 - (-0.2)$$

$$Y_{3(3)} - Y_{6(3)} - \hat{\psi} = -1 - (-0.2)$$

$$Y_{8(4)} - Y_{9(4)} - \hat{\psi} = 0 - (-0.2)$$

$$Y_{11(5)} - Y_{10(5)} - \hat{\psi} = 0 - (-0.2)$$

$$\begin{aligned} \text{Var}(\hat{\psi}) &= \frac{1}{4}(4 \cdot 0.2^2 + 1 \cdot [-0.8]^2) \\ &= 0.2 \end{aligned}$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Greift die Idee auf, dass kontrafaktische Werte über den gematchten Partner “imputiert” werden können
- Damit können die individuellen Effektunterschiede, sowie deren Variation, berechnet werden
- plausibel, nicht zu komplex, trotzdem eher pragmatisch
- Vorteil: relativ allgemein anwendbar
- Berechne, oder schätze, bedingte Erwartungswerte $E(Y|A, L)$ um kontrafaktische Werte “vorherzusagen” (möglicherweise stochastisch)
- Varianz des Schätzers = Varianz der individuellen Effekte

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



Beispiel: ...von oben

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

$$E(Y|A=0, L=0) = \frac{1}{3}$$

$$E(Y|A=0, L=1) = \frac{1}{2}$$

$$E(Y|A=1, L=0) = \frac{2}{3}$$

$$E(Y|A=1, L=1) = \frac{1}{2}$$

$$\psi_{ATT} = E(\psi_i) = 20\%$$

$$\begin{aligned} \text{Var}(\psi_{ATT}) &= \text{Var}(\psi_i) \\ &= 0.325 \end{aligned}$$

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

besser: aus Verteilungen ziehen, z.B. Binomial mit $p = \frac{1}{3}$ etc.



■ Bootstrapping

- Problem: wie wird das manuelle rekursive Verfahren der Überprüfung der Matching-Qualität aufgenommen?

■ Für NN-Matching gibt es erste komplexe Schätzmethoden, die asymptotisch korrekt sind [5]

- für endlichen Stichprobenumfang wenig getestet bisher, prospektiv aber vielversprechend

■ ...

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- Matching wurde entwickelt um Sensitivität bezüglich Modellspezifikation zu verringern – was heißt das?
- Unter perfekter Erfüllung der Annahmen (siehe oben) erreichen wir Austauschbarkeit
 - Damit müssen wir den Einfluss von L gar nicht mehr modellieren
 - wird teilweise trotzdem gemacht, siehe potentielle Probleme damit in Kapitel 7
- Außerdem: Vergleich auf die “Region gemeinsamer Unterstützung”

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes

Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

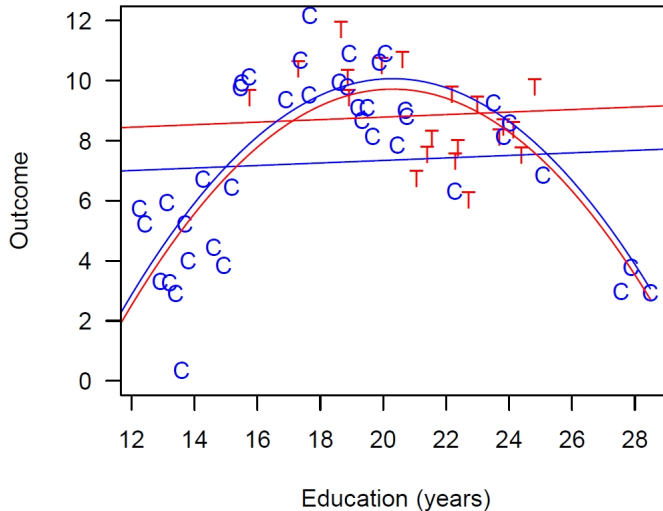
Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Graphisches Beispiel [6] – ohne Matching



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

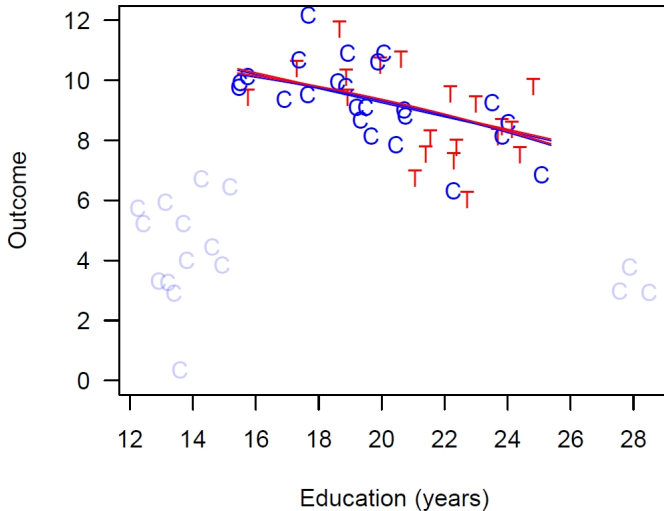
Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise

Graphisches Beispiel [6] – ohne Matching



Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- In der Regel ist $n_c \neq n_t$: d.h. viele Beobachtungen werden bei NN-Matching ausgeschlossen
 - Positivitätsverletzungen, die stärker bei den ausgeschlossenen Beobachtungen ausgeprägt sind, können unentdeckt bleiben
 - Effektmodifizierung wird evtl. nicht korrekt erfaßt wenn zu viele relevante Beobachtungen ausgeschlossen werden
- “Doctor, according to our study, this drug is beneficial for patients who have a propensity score between 0.11 and 0.93 when they arrive at your office, but it may kill those with propensity scores below 0.11” [7]

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching

Vergrößertes Exaktes
Matching

Nearest Neighbour

Andere Methoden

Überprüfung des
Matching

Selektionsbias

Balance

Propensity Scores

Weiteres

Konfidenzintervalle

Modellspezifikation

Sonstiges

Literaturnachweise



- [1] S. M. Iacus, G. King, and G. Porro.
Causal inference without balance checking: Coarsened exact matching.
Political Analysis, 20(1):1–24, 2012.
- [2] Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart.
MatchIt: Nonparametric preprocessing for parametric causal inference.
Journal of Statistical Software, 42(8):1–28, 2011.
- [3] E. A. Stuart.
Matching methods for causal inference: A review and a look forward.
Statistical Science, 25(1):1–21, 2010.
- [4] D. E. Ho, K. Imai, G. King, and E. A. Stuart.
Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference.
Political Analysis, 15(3):199–236, 2007.
- [5] A. Abadie and G. W. Imbens.
Large sample properties of matching estimators for average treatment effects.
Econometrica, 74(1):235–267, 2006.
- [6] G. King.
Matching methods for causal inference.
<https://qipsr.as.uky.edu/gary-king-matching-methods-causal-inference>.
- [7] M. Hernan and J. Robins.
Causal inference.
Chapman & Hall/CRC, Boca Raton, 2017.

Grundidee

Allgemeines Vorgehen

Abstandsmaße

Matchmethoden

Exaktes Matching
Vergrößertes Exaktes
Matching
Nearest Neighbour
Andere Methoden

Überprüfung des Matching

Selektionsbias
Balance
Propensity Scores

Weiteres

Konfidenzintervalle
Modellspezifikation
Sonstiges

Literaturnachweise