

## Inhaltsverzeichnis

<b>1</b>	<b>Propensity Score Methods II - ANCOVA</b>	<b>1</b>
1.1	Datensatz . . . . .	2
1.2	$t$ -Test mit gematchten Stichproben . . . . .	2
1.3	Klassische ANCOVA: Kontrolle des PS-Logits . . . . .	2
1.4	Generalisierte ANCOVA: Interaktion der Gruppe mit PS-Logit . . . . .	3
<b>2</b>	<b>Propensity Score Stratifizierung</b>	<b>4</b>
2.1	Bewertung der Balance . . . . .	5
2.2	QQ-Plots der Kovariaten . . . . .	5
2.3	Effektschätzung . . . . .	6
2.3.1	ATE . . . . .	8
2.3.2	ATT . . . . .	8
<b>3</b>	<b>Propensity Score Gewichtung</b>	<b>8</b>
3.1	Effektschätzung . . . . .	9
3.1.1	ATE . . . . .	9
3.1.2	ATT . . . . .	10
<b>4</b>	<b>Zusammenfassung - Matching- und Propensity Score Methoden</b>	<b>11</b>

## 1 Propensity Score Methods II - ANCOVA

### Lernziele und Vorgehen

Ziel: Präzise Schätzung von ATE und ATT. Dabe gibt es verschiedene Möglichkeiten den Propensity Score zu berücksichtigen.

1. Matching über PS und anschließend einfacher  $t$ -Test
2. Matching und Kombination mit ANCOVA
  - a) Kontrolle des PS in ANCOVA
  - b) Kontrolle von PS und weiteren Kovariaten in ANCOVA
3. Falls Matching schwierig bzw. Stichprobe zu klein
  - a) Matching von PS-Strata (Schichten)
  - b) Gewichtung der Personen anhand ihrer PS



## 1.1 Datensatz

Nach der Vorbereitung und PS-Modell aus Sitzung 3 nutzen wir `match.data()`, um den gematchten Datensatz zu erhalten.

```
psdat <- match.data(ps_mi2)
head(psdat)
```

	bdi1	bdi2	age	sess	cogn	addic	group	distance	weights	subclass
1	43	27	28	24	94	0	1	0.5754179	1	1
2	33	33	41	16	73	0	1	0.5146133	1	12
3	35	25	25	13	85	0	1	0.4837395	1	23
4	27	31	46	24	74	0	1	0.4368074	1	34
5	32	33	28	19	81	0	1	0.4843960	1	45
6	21	19	35	22	90	0	1	0.2666676	1	56

- Der erzeugte Datensatz `psdat` enthält nur noch die gematchten Personen (152 von 255).
- Die Variable `distance` enthält den geschätzten Propensity Score der jeweiligen Person und ist mit unserem vorher erzeugten `Pscore` identisch.

## 1.2 t-Test mit gematchten Stichproben

```
m1 <- lm(bdi2 ~ group, data = psdat)
summary(m1)$coef
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	32.723684	0.9660954	33.87211	3.696595e-72
group	-9.684211	1.3662652	-7.08809	4.961708e-11

- $\widehat{ATE} = \widehat{ATT} = \alpha_1 = -9.68$

## 1.3 Klassische ANCOVA: Kontrolle des PS-Logits

Der Propensity Score Logit kann nach dem PS Matching noch als Kovariate in das Regressionsmodell zur Treatment-Schätzung aufgenommen werden. Hierzu erstellen wir uns in unserem gematchten Datensatz zunächst eine Variable welche den zentrierten Logit der propensity scores enthält. Der Logit des PS wird (anstelle des PS selbst) als Kovariate ins Modell aufgenommen, da der PS Logit eher linear mit dem outcome zusammenhängt als der PS.

```
# Logit als Variable erstellen und zentrieren
psdat$logit <- car::logit(psdat$distance)
psdat$logit <- scale(psdat$logit, scale = FALSE)
# Als Kovariate ins Modell aufnehmen
m2 <- lm(bdi2 ~ logit + group, data = psdat)
summary(m2)$coef
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	32.948233	0.7049150	46.74072	6.484003e-91
logit	7.423945	0.6438316	11.53088	2.177485e-22
group	-10.133309	0.9972807	-10.16094	9.357786e-19

- $\widehat{ATE} = \widehat{ATT} = \alpha_2 = -10.13$

#### 1.4 Generalisierte ANCOVA: Interaktion der Gruppe mit PS-Logit

```
m3 <- lm(bdi2 ~ logit*group, data = psdat)
summary(m3)$coef
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	32.928149	0.7048273	46.718042	1.726711e-90
logit	6.759937	0.8943457	7.558529	3.939532e-12
group	-10.134787	0.9968033	-10.167289	9.564668e-19
logit:group	1.376884	1.2878570	1.069128	2.867525e-01

- Kovariate  $Z$  hier zentrierter PS-Logit, daher  $E(Z) = 0$
- $\widehat{ATE} = \alpha_2 + \alpha_3 \cdot E(Z) = \alpha_2 = -10.13$
- $\widehat{ATT}_{X=1} = \alpha_2 + \alpha_3 \cdot E(Z|X=1) = -10.13 + 1.38 \cdot 0.03 = -10.09$
- $E(Z | X=1)$  berechnet über:

```
tapply(psdat$logit, psdat$group, mean)
```

	0	1
	-0.03024659	0.03024659



## 2 Propensity Score Stratifizierung

Die Propensity Score Stratifizierung ist eine Methode, die das Matching von Personen basierend auf der Bildung einer bestimmten Anzahl von homogenen Strata (oder Schichten)  $Q$  in Bezug auf den Propensity Score durchführt. Innerhalb jeder dieser Strata wird dann eine Schätzung des kausalen Effekts durchgeführt. Der Gesamteffekt wird dann als gewichtetes Mittel über die stratum-spezifischen kausalen Effekte berechnet. Eine häufige Wahl für die Anzahl der Strata  $Q$  ist fünf (Voreinstellung/Default: 6 Strata), obwohl es durchaus üblich ist, je nach den spezifischen Bedürfnissen der Analyse mehr oder weniger Strata zu verwenden. Mit `MatchIt` wird diese Methode als *Subclassification* bezeichnet und kann durch den Aufruf von `method = "subclass"` angewendet werden.

```
stra5 <- matchit(group ~ bdi1 + age + cogn + sess + addic,
                 data = bdi_data,
                 distance = "logit",
                 method = "subclass", # Stratifizierung
                 subclass = 5)       # Anzahl Strata
```

- Das Argument `subclass` gibt entweder an, wieviele Subklassen / Strata gebildet werden sollen (ganze Zahl eingeben), oder, einen Vektor mit Grenzwerten der Quantile der gewählten Distanzmetrik (Wahrscheinlichkeiten zwischen 0 und 1).

Anzahl Personen in den Strata:

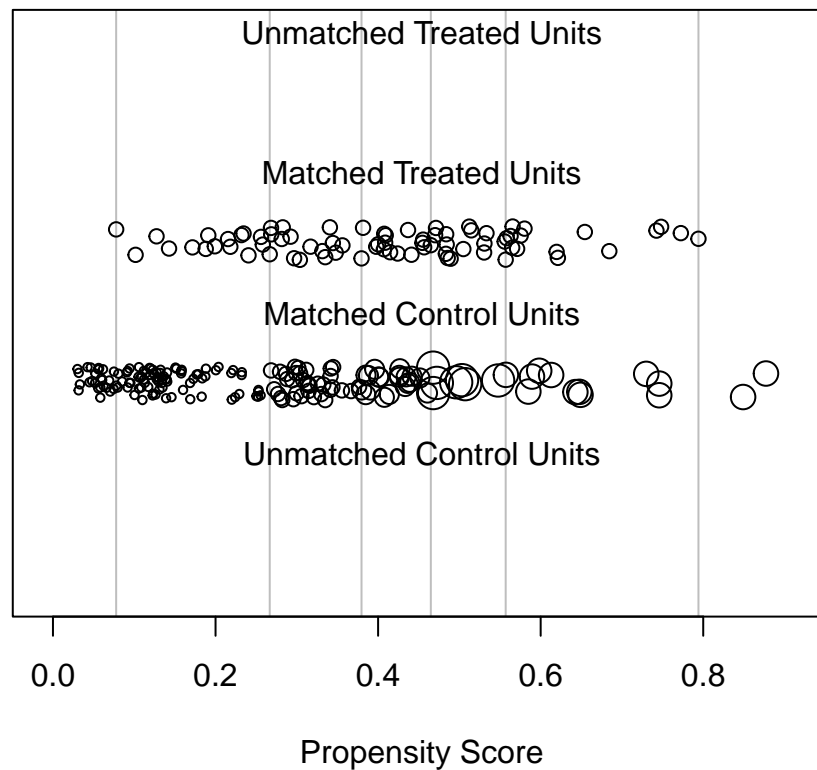
```
Nq <- t(summary(stra5)$qn)
Nq
```

	Control	Treated	Total
1	104	15	119
2	35	15	50
3	20	15	35
4	7	15	22
5	13	16	29
All	179	76	255

Gruppenspezifische PS-Verteilungen innerhalb der Strata:

```
plot(stra5, type = "jitter", interactive = FALSE)
```

## Distribution of Propensity Scores



- Verschiedene Strata durch vertikale Striche dargestellt / voneinander getrennt
- Fläche der Kreise ist proportional zur Gewichtung (`weight`) der Beobachtung

### 2.1 Bewertung der Balance

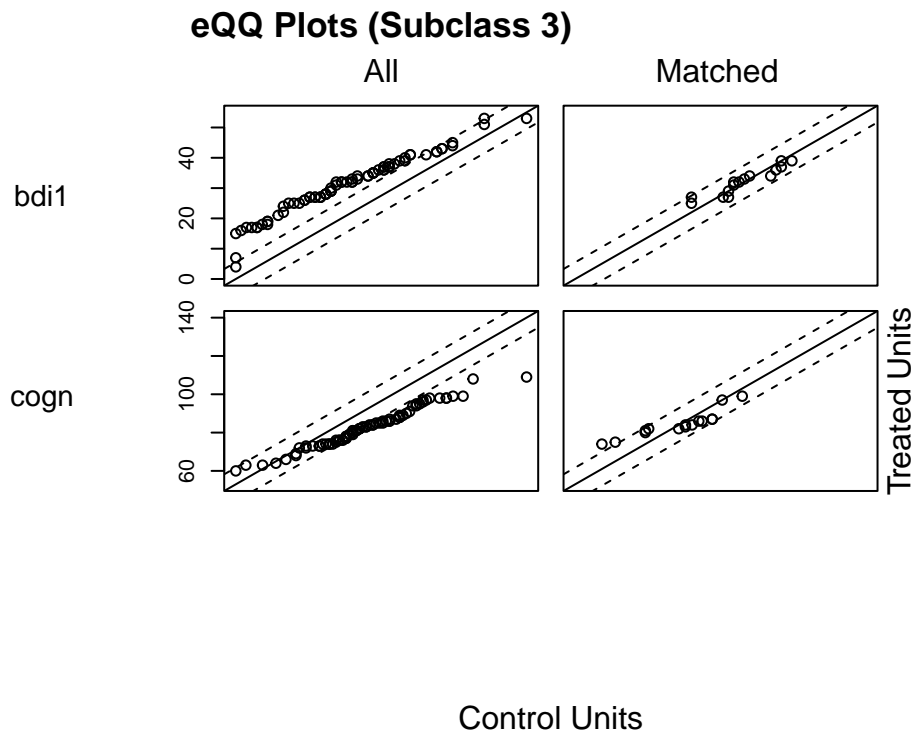
- Zusammenfassung pro Subklasse und über Subklassen hinweg

```
summary(stra5, standardize = TRUE, subclass = TRUE)
```

### 2.2 QQ-Plots der Kovariaten

- Betrachtung der Verteilung einzelner Kovariaten für jeweilige Schicht (hier: Stratum 3)

```
plot(stra5, type = "QQ",
     which.xs = c("bdi1", "cogn"),
     interactive = FALSE,
     subclass = 3)
```



## 2.3 Effektschätzung



- Bei Stratifizierung wird zunächst pro Stratum ein Treatmenteffekt geschätzt und die geschätzten Effekte der Strata werden hinterher miteinander verrechnet
- Dafür werden in einem ersten Schritt, Stratum- und gruppenspezifische Mittelwerte der Outcome-Variable erstellt (stratum-spezifische *Potential Outcomes*)
- Information über die Größe der Strata werden aus  $N_q$  (s. oben) hinzugefügt

```
dat5 <- match.data(stra5)
long <- aggregate(bdi2 ~ subclass + group,
                  data = dat5, mean)
long
```

	subclass	group	bdi2
1	1	0	17.04808
2	2	0	29.31429
3	3	0	34.25000
4	4	0	33.71429
5	5	0	40.69231
6	1	1	13.33333
7	2	1	20.13333
8	3	1	22.73333
9	4	1	27.06667
10	5	1	31.37500

```
wide <- reshape(long, direction = "wide", sep = "_",
                timevar = "group", idvar = "subclass")
wide <- cbind(wide, Nq[-6, c("Treated", "Total")])
wide
```

	subclass	bdi2_0	bdi2_1	Treated	Total
1	1	17.04808	13.33333	15	119
2	2	29.31429	20.13333	15	50
3	3	34.25000	22.73333	15	35
4	4	33.71429	27.06667	15	22
5	5	40.69231	31.37500	16	29

- In einem zweiten Schritt wird der stratum-spezifische ATE für jedes Stratum  $q$  geschätzt (Differenz der stratum-spezifischen *potential outcomes*)

```
wide$ATEq <- wide$bdi2_1 - wide$bdi2_0
wide
```

	subclass	bdi2_0	bdi2_1	Treated	Total	ATEq
1	1	17.04808	13.33333	15	119	-3.714744
2	2	29.31429	20.13333	15	50	-9.180952
3	3	34.25000	22.73333	15	35	-11.516667
4	4	33.71429	27.06667	15	22	-6.647619
5	5	40.69231	31.37500	16	29	-9.317308

- Im dritten Schritt werden ATE und ATT als gewichtete Mittelwerte über die stratum-spezifischen Treatmenteffekte berechnet
- Gesamt-Effekt als Summe der gewichteten Stratum-Effekte (gewichteter Mittelwert)

### 2.3.1 ATE

- Relative Häufigkeit des Stratum als Gewicht  $W_q$  für ATE

$$\widehat{ATE} = \sum_{q=1}^Q W_q \widehat{ATE}_q$$

```
wide$Wq <- wide$Total / sum(wide$Total)
ATE <- sum(wide$Wq * wide$ATEq)
ATE
```

```
[1] -6.747588
```

### 2.3.2 ATT

- Gewichtung mit Häufigkeit der Treatment-Gruppe innerhalb des Stratum relativ zur Treatment-Häufigkeit

$$\widehat{ATT} = \sum_{q=1}^Q W_{Tq} \widehat{ATE}_q$$

```
wide$WTq <- wide$Treated / sum(wide$Treated)
ATT <- sum(wide$WTq * wide$ATEq)
ATT
```

```
[1] -8.091798
```

## 3 Propensity Score Gewichtung

Die Propensity Score Gewichtung (PS-Gewichtung) versucht, genau wie beim Matching, den Zustand eines Zufallsexperiments erzeugen, indem Personen mit ähnlichen Eigenschaften aus der Treatment- und Kontrollgruppe auf der Grundlage ihres Propensity-Scores, also der Wahrscheinlichkeit, die Behandlung zu erhalten, gewichtet wird. Das Ziel ist auch hier eine Balance in den Kovariaten zwischen den Gruppen zu schaffen, was zu einer genaueren Schätzung des Treatmenteffekts führt.

Um die Effekte einer Intervention oder Behandlung zu berechnen, wird im Rahmen der PS-Gewichtung jeder Person ein Gewicht zugewiesen. Dieses Gewicht hängt vom Propensity Score dieser Person und ihrer Gruppenzugehörigkeit ab - also ob sie zur Kontroll- oder Treatmentgruppe gehört.



In der Kontrollgruppe erhalten Personen mit einem hohen Propensity Score ein hohes Gewicht. Das bedeutet, dass Personen, die eine hohe Wahrscheinlichkeit haben, die Behandlung zu erhalten (basierend auf ihren beobachteten Kovariaten), aber die Behandlung tatsächlich nicht erhalten haben, eine größere Bedeutung in der Analyse erhalten. Umgekehrt erhalten in der Treatmentgruppe Personen mit einem geringen Propensity Score ein hohes Gewicht. Das heißt, Personen, die eine geringe Wahrscheinlichkeit haben, die Behandlung zu erhalten, aber die Behandlung tatsächlich erhalten haben, werden in der Analyse stärker berücksichtigt.

Das Gewicht wird üblicherweise durch den Kehrwert des Propensity Scores für die Treatmentgruppe und den Kehrwert von  $1 - \hat{\pi}_i$  für die Kontrollgruppe berechnet. Diese Art der Gewichtung nennt man auch *Inverse Probability Weighting*.



$$W_i = \frac{X_i}{\hat{\pi}_i} + \frac{1 - X_i}{1 - \hat{\pi}_i}$$

$$= \frac{X_i}{P(X = 1|\mathbf{Z})} + \frac{1 - X_i}{1 - P(X = 1|\mathbf{Z})}$$

```
dat5$ATEweight <- ifelse(test = dat5$group == 1,
  yes = 1 / dat5$distance,
  no = 1 / (1 - dat5$distance))
```

- `ifelse` gibt einen Vektor aus, der für alle Personen mit einem `TRUE` auf der `test`-Variable (alle Personen mit `group==1`) den unter dem Argument `yes` angegebenen Wert enthält (hier:  $\frac{1}{\hat{\pi}_i}$ ), und für alle Personen mit einem `FALSE` auf der `test`-Variable (alle Personen mit `group==0`) den unter dem Argument `no` angegebenen Wert enthält (hier:  $\frac{1}{1-\hat{\pi}_i}$ )

## 3.1 Effektschätzung

### 3.1.1 ATE

Der ATE wird durch die Differenz der gewichteten Mittelwerte zwischen den Gruppen geschätzt. In R können wir für die Berechnung der gewichteten Mittelwerte die `weighted.mean()`-Funktion nutzen. Diese Funktion nimmt mit dem Argument `x` die Daten entgegen für die der gewichtete Mittelwert ausgerechnet werden soll. Dem Argument `w` müssen die entsprechenden Gewichte, in unserem Fall  $W_i$  übergeben werden.

```
# Kontrollgruppe
gMW_KG <- weighted.mean(x = dat5$bdi2[dat5$group == 0],
```

```

w = dat5$ATEweight[dat5$group == 0])
# Treatmentgruppe
gMW_IG <- weighted.mean(x = dat5$bdi2[dat5$group == 1],
                        w = dat5$ATEweight[dat5$group == 1])
# ATE als Differenz der gewichteten Mittelwerte der Gruppen
gMW_IG - gMW_KG

```

```
[1] -7.722647
```

- $\widehat{ATE} = -7.72$

Der ATE kann auch im Regressionsmodell über den WLS-Schätzer (Weighted Least Squares) mit `lm()` über das `weights`-Argument geschätzt werden:

```

wATE <- lm(bdi2 ~ group, weights = ATEweight, data = dat5)
coef(wATE)

```

```

(Intercept)      group
  27.257838    -7.722647

```

### 3.1.2 ATT

Bei der Berechnung des ATT erhalten alle Personen in der Treatmentgruppe, das heißt diejenigen, die tatsächlich das Treatment erhalten haben, ein Gewicht von 1. Sie werden also vollständig in der Analyse berücksichtigt.

Personen in der Kontrollgruppe erhalten jedoch ein Gewicht, das auf ihrem Propensity Score basiert. Ihnen werden die Odds ihres Propensity Scores als Gewicht zugewiesen. Das bedeutet, dass Personen in der Kontrollgruppe, die eine höhere Wahrscheinlichkeit haben, die Behandlung zu erhalten (basierend auf ihren beobachteten Eigenschaften), in der Analyse stärker berücksichtigt werden.

$$W_i = X_i + \frac{(1 - X_i) \cdot \hat{\pi}_i}{1 - \hat{\pi}_i}$$

Berechnung der Gewichte in R mit `ifelse()`:

```

dat5$ATTweight <- ifelse(test = dat5$group == 1,
                        yes = 1,
                        no = dat5$distance/(1 - dat5$distance))

```

```
gMW_KG <- weighted.mean(x = dat5$bdi2[dat5$group == 0],
                        w = dat5$ATTweight[dat5$group == 0])
gMW_IG <- weighted.mean(x = dat5$bdi2[dat5$group == 1],
                        w = dat5$ATTweight[dat5$group == 1])
# ATT als Differenz der gewichteten Mittelwerte:
gMW_IG - gMW_KG
```

```
[1] -11.49387
```

- $\widehat{ATT} = -11.49$

## 4 Zusammenfassung - Matching- und Propensity Score Methoden

1. **Exaktes Matching:** Hierbei werden Treatment- und Kontrolleinheiten anhand exakt übereinstimmender Kovariaten gematched. Diese Methode kann zwar einfach sein, ist aber oft einschränkend und bei hochdimensionalen Kovariaten oft unpraktisch, da es unwahrscheinlich ist, exakte Übereinstimmungen zu finden.
2. **Nearest Neighbor Matching:** Bei dieser Methode wird für jede Behandlungseinheit eine Kontrolleinheit (oder mehrere) ausgewählt, die der Behandlungseinheit am ähnlichsten ist. Die Ähnlichkeit wird basierend anhand einer bestimmten Distanzmetrik berechnet, die auf den Kovariaten oder den Propensity Scores basieren kann.
3. **Propensity Score Matching:** Bei dieser Methode werden Treatment- und Kontrolleinheiten anhand ihrer Propensity Scores (der geschätzten Wahrscheinlichkeit, die Behandlung zu erhalten) gematched. Zur Schätzung der Propensity Scores wird oft die logistische Regression verwendet.
4. **Propensity Score Stratification:** In dieser Methode werden die Propensity Scores verwendet, um die Daten in verschiedene Strata oder Gruppen zu unterteilen. Innerhalb dieser Strata wird angenommen, dass Behandlungs- und Kontrolleinheiten ähnliche Propensity Scores haben. Die Effektschätzung erfolgt dann innerhalb dieser Strata und die geschätzten Effekte der Strata werden aggregiert. Die Annahme hier ist, dass das Ergebnis innerhalb jedes Stratums unabhängig von der Behandlungszuordnung ist, bedingt durch den Propensity Score.
5. **Propensity Score Weighting:** Bei dieser Methode werden Gewichte basierend auf den Propensity Scores berechnet, um eine gewichtete Stichprobe zu erstellen, in der die Verteilung der Kovariaten zwischen Behandlungs- und Kontrollgruppe ausgeglichen ist.
6. **Propensity Score Regression:** Bei dieser Methode werden die Propensity Scores direkt in ein Regressionsmodell aufgenommen, um den kausalen Effekt zu schätzen. Es wird kein Matching oder eine Gewichtung durchgeführt.