

Kausale Effektschätzung - Matching

Aufgabe 1: Datensatz und Deskriptivstatistiken

```
str(dat)
```

```
'data.frame':  103 obs. of  5 variables:
 $ sex      : num  1 1 1 1 1 1 1 1 1 1 ...
 $ matVL    : num  0 1 1 1 1 1 1 1 1 1 ...
 $ eng_pre  : num  0.74 0.59 0.73 0.65 0.63 0.77 0.61 0.64 0.63 0.65 ...
 $ eng_post : num  0.84 0.84 0.9 0.25 0.69 0.9 0.35 0.89 0.67 0.74 ...
 $ group    : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
summary(dat)
```

sex	matVL	eng_pre	eng_post
Min. :0.0000	Min. :0.0000	Min. :0.1200	Min. :0.2000
1st Qu.:1.0000	1st Qu.:0.0000	1st Qu.:0.5500	1st Qu.:0.5350
Median :1.0000	Median :0.0000	Median :0.6500	Median :0.7300
Mean :0.8544	Mean :0.2621	Mean :0.6595	Mean :0.6683
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:0.7750	3rd Qu.:0.8300
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :0.9800

group
Min. :0.000
1st Qu.:0.000
Median :0.000
Mean :0.466
3rd Qu.:1.000
Max. :1.000

```
# Gruppenspezifische Mittelwerte
library(psych)
describeBy(dat, group = dat$group)
```

```
Descriptive statistics by group
group: 0
      vars  n mean   sd median trimmed  mad  min  max range  skew kurtosis
```

sex	1	55	0.89	0.31	1.00	0.98	0.00	0.00	1.00	1.00	-2.44	4.03
matVL	2	55	0.18	0.39	0.00	0.11	0.00	0.00	1.00	1.00	1.61	0.59
eng_pre	3	55	0.71	0.11	0.70	0.71	0.10	0.45	0.95	0.50	-0.02	-0.59
eng_post	4	55	0.71	0.23	0.79	0.74	0.12	0.20	0.98	0.78	-1.17	-0.06
group	5	55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NaN	NaN

```

se
sex      0.04
matVL    0.05
eng_pre  0.02
eng_post 0.03
group    0.00

```

group: 1

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis
sex	1	48	0.81	0.39	1.00	0.88	0.00	0.00	1.00	1.00	-1.55	0.42
matVL	2	48	0.35	0.48	0.00	0.32	0.00	0.00	1.00	1.00	0.59	-1.68
eng_pre	3	48	0.61	0.19	0.59	0.61	0.17	0.12	1.00	0.88	0.05	-0.15
eng_post	4	48	0.62	0.16	0.60	0.62	0.16	0.28	0.97	0.69	0.00	-0.52
group	5	48	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	NaN	NaN

```

se
sex      0.06
matVL    0.07
eng_pre  0.03
eng_post 0.02
group    0.00

```

Aufgabe 2: Prima Facie Effekt

```

# t-Test/einfache Regression
m1 <- lm(eng_post ~ group, data = dat)
summary(m1)$coef

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.70800000	0.02735778	25.879297	2.280386e-46
group	-0.08529167	0.04007548	-2.128275	3.574717e-02

Negativer Effekt des Englischkurses: Teilnehmende des Englischtrainings haben im Mittel ein geringeres Ergebnis (PFE = -0.085) im Englisch-Test als die Teilnehmenden des Matheskurses.

Aufgabe 3: Exaktes Matching

```
library(MatchIt)
exakt <- matchit(group ~ sex + eng_pre + matVL,
                 method = "exact",
                 data = dat)
```

Aufgabe 4: Personen und Subklassen

```
exakt
```

A matchit object

- method: Exact matching
- number of obs.: 103 (original), 47 (matched)
- target estimand: ATT
- covariates: sex, eng_pre, matVL

```
exdat <- match.data(exakt)
str(exdat)
```

Classes 'matchdata' and 'data.frame': 47 obs. of 7 variables:

```
$ sex      : num  1 1 1 1 1 0 1 1 1 1 ...
$ matVL    : num  1 1 1 1 1 1 0 0 0 0 ...
$ eng_pre  : num  0.73 0.63 0.61 0.64 0.63 0.69 0.82 0.45 0.49 0.76 ...
$ eng_post : num  0.9 0.69 0.35 0.89 0.67 0.74 0.84 0.4 0.87 0.76 ...
$ group    : num  0 0 0 0 0 0 0 0 0 0 ...
$ weights  : num  1.136 0.568 1.136 1.136 0.568 ...
$ subclass: Factor w/ 16 levels "1","2","3","4",...: 1 2 3 4 2 5 6 7 8 9 ...
- attr(*, "weights")= chr "weights"
- attr(*, "subclass")= chr "subclass"
```

```
summary(exakt, un = F)
```

Call:

```
matchit(formula = group ~ sex + eng_pre + matVL, data = dat,
        method = "exact")
```

Summary of Balance for Matched Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean
sex	0.9545	0.9545	0	.	0
eng_pre	0.6168	0.6168	0	0.9726	0
matVL	0.2273	0.2273	-0	.	0
	eCDF Max	Std. Pair	Dist.		
sex	0	0			
eng_pre	0	0			
matVL	0	0			

Sample Sizes:

	Control	Treated
All	55.	48
Matched (ESS)	13.96	22
Matched	25.	22
Unmatched	30.	26
Discarded	0.	0

Es wurden 16 Subklassen mit gleichen Ausprägungen auf den Kovariaten gebildet. In diesen wurden insgesamt 25 Personen der Kontrollbedingung (hier: Mathe) und 22 aus der Treatmentbedingung (Englisch) gematched. Nicht gematched wurden 30 bzw. 26 Personen.

Die Frage nach der Verringerung der Unterschiedlichkeit ist beim exakten Matching in gewisser Weise überflüssig: Die Unterschiedlichkeit in den Kovariaten ist natürlich komplett eliminiert, da exakt gematcht wurde. Das heißt, die Gruppen unterscheiden sich nicht mehr auf den Kovariaten.

Aufgabe 5: Effektschätzung

```
exdat <- match.data(exakt)
agg <- aggregate(eng_post ~ subclass + group, data = exdat, mean)
wide <- reshape(agg, direction = "wide", sep = "_", idvar = "subclass", timevar = "group")
exdat <- merge(exdat, wide)
# Schätzung des ATE
mean(exdat$eng_post_1 - exdat$eng_post_0)
```

```
[1] -0.09698582
```

```
# Schätzung des ATT
tg <- exdat[exdat$group==1, ] # Teildatensatz Treatmentgruppe
mean(tg$eng_post_1 - tg$eng_post_0)
```

```
[1] -0.1037879
```

$$\widehat{ATE} = -0.097$$

$$\widehat{ATT} = -0.104$$

Aufgabe 6: Nearest Neighbor Matching

```
set.seed(10)
near <- matchit(group ~ sex + eng_pre + matVL, data = dat,
  method = "nearest", distance = "mahalanobis",
  ratio = 1, replace = FALSE)
```

Aufgabe 7: Beurteilung der Balance

```
summary(near, standardize = TRUE)
```

Call:

```
matchit(formula = group ~ sex + eng_pre + matVL, data = dat,
  method = "nearest", distance = "mahalanobis", replace = FALSE,
  ratio = 1)
```

Summary of Balance for All Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean
sex	0.8125	0.8909	-0.2009	.	0.0784
eng_pre	0.6058	0.7064	-0.5295	2.7903	0.1704
matVL	0.3542	0.1818	0.3604	.	0.1723

	eCDF Max
sex	0.0784
eng_pre	0.4042
matVL	0.1723

Summary of Balance for Matched Data:

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean
sex	0.8125	0.8750	-0.1601	.	0.0625
eng_pre	0.6058	0.6948	-0.4685	2.6229	0.1431
matVL	0.3542	0.2083	0.3049	.	0.1458
	eCDF Max	Std. Pair Dist.			
sex	0.0625	0.1601			
eng_pre	0.3750	0.6726			
matVL	0.1458	0.3049			

Sample Sizes:

	Control	Treated
All	55	48
Matched	48	48
Unmatched	7	0
Discarded	0	0

Die Unterschiede der Besetzung auf der Geschlechtsvariable sind um rund 20% reduziert worden.

Unterschiede auf dem Englisch-Vortest sind um rund 11% reduziert worden und betragen im Mittel noch 0.47 Standardabweichungen (wobei die Kontrollgruppe höhere Vortestwerte hat). Die ungleiche Verteilung in der Vorliebe für Mathe wurde um rund 15% reduziert, die Vorliebe ist aber nach wie vor höher in der Treatmentgruppe: Rund 35% der Treatmentgruppen-Personen geben eine Vorliebe an, während es in der Kontrollgruppe nur ca. 21% sind.

Aufgabe 8: Effektschätzung

```
neardat <- match.data(near)
# Durch 1:1 Matching ohne Zurücklegen entspricht der Unterschied in
# Mittelwerten dem mittleren Unterschied über Personen. Daher einfache
# Mittelwertsdifferenz z.B. über `lm()` berechenbar:
lm(eng_post ~ group, data = neardat)
```

Call:

```
lm(formula = eng_post ~ group, data = neardat)
```

Coefficients:

```
(Intercept)      group
  0.70000      -0.07729
```

$$\widehat{ATE} = \widehat{ATT} = -0.077$$