Inhaltsverzeichnis

1 Vorwort

Bevor Sie mit dem Studium des Inhalts beginnen, möchte ich einige wichtige Punkte hervorheben.

Das Skript dient als umfassende Quelle, die tiefergehende Informationen und zusätzliche Details zu den in den Videos behandelten Themen bietet. Ich ermutige Sie dazu, sowohl das Skript als auch die Videos zu nutzen, um ein ganzheitliches Verständnis des Lernstoffs zu gewinnen.

Außerdem finden Sie in den *Margins* dieses Skripts öfter ein *Videosymbol* (siehe rechts). Dieses Symbol verweist stets auf ein zum Thema zugehöriges Video. Klicken Sie daher einfach auf das Symbol und Sie werden direkt auf das entsprechende Video weitergeleitet (das Symbol hier rechts führt Sie nur zu Moodle). Es dient dazu, die wichtigsten (nicht alle) Inhalte und Abläufe in R zu veranschaulichen.

2 Kovarianzanalyse

Im folgenden werden wir die Schätzung kausaler Effekte anhand der (generalisierten) Kovarianzanalyse[1] betrachten. Dafür greifen wir auf die Theorie kausaler Effekte zurück und spezifizieren die g_0- und g_1-Funktionen unter verschiedenen Modellannahmen.

Lernziele und Vorgehen

Sie können kausale Effekte in einfachen Regressionsmodellen, in der klassischen und in der generalisierten ANCOVA bestimmen. Dafür werden wir wie folgt vorgehen:

0. Wiederholung der ANCOVA im Kontext kausaler Effekte
1. Deskriptive Datenanalyse
2. Bestimmung kausaler Effekte in der …
   a) … einfachen Regression
   b) … klassischen ANCOVA
   c) … generalisierten ANCOVA anhand von `lm`-Modellen
3. Verwendung von `EffectLiteR` zur Bestimmung der Effekte in den gleichen Modellen a), b) und c)

2.1 Vorbereitung

```
setwd("C:/Users/me/myworkingdirectory")
load("spf1.rda")
```

Außerdem benötigen wir die folgenden Pakete:

```
if (!require("car")) install.packages("car")
if (!require("jtools")) install.packages("jtools")
if (!require("sjPlot")) install.packages("sjPlot")
if (!require("EffectLiteR")) install.packages("EffectLiteR")
```

Der obige Code prüft zuerst, ob das gewünschte Paket installiert ist. Ist dies nicht der Fall, wird das Paket mit `install.packages()` installiert.

[1] g-ANCOVA

Paket	Beschreibung
`car`	Das "Companion to Applied Regression"-Paket bietet Funktionen für die Regression und die Varianzanalyse.
`jtools`	Dieses Paket bietet Hilfsfunktionen für die Datenanalyse.
`sjPlot`	Dieses Paket bietet Funktionen zum Erstellen von Plots für verschiedene Arten von Daten und statistischen Modellen.
`EffectLiteR`	Dieses Paket kann verwendet werden, um kausale Effekte zu schätzen und zu visualisieren.

2.1.1 Erinnerung: ANCOVA im Kontext kausaler Effekte

Die ANCOVA ermöglicht bei der Berechnung des Average Treatment Effects (ATE), des durchschnittlichen kausalen Effekts, für eine Kovariate Z oder für einen Vektor aus Kovariaten $\mathbf{Z}$ zu kontrollieren. Im einfachsten Fall gibt es eine Outcomevariable Y, eine dichotome Treatmentvariable X und eine Kovariate Z. Die Kontrollgruppe wird standardmäßig mit $X = 0$ und die Treatmentgruppe mit $X = 1$ kodiert.
Eine allgemein Notation, welche auch die generalisierte ANCOVA miteinschließt ist die Zerlegung in die Intercept- und die Effektfunktion:

$$E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X$$

Der ATE kann dann anhand des Erwartungswerts der g_1-Funktion geschätzt werden:

$$\widehat{ATE} = E[g_1(Z)] = E(\alpha_2 + \alpha_3 Z) = \alpha_2 + \alpha_3 E(Z)$$

Der ATE ist der durchschnittliche Effekt der Behandlung über die gesamte Population, und er wird durch den Erwartungswert von $g_1(Z)$ geschätzt. In Worten bedeutet dies, dass wir den durchschnittlichen Unterschied zwischen Behandlung und Kontrolle über alle möglichen Werte von Z betrachten. Der Average Treatment Effect of the Treated (ATT, durchschnittlicher kausaler Effekt der *Behandelten*) kann anhand des Erwartungswerts der g_1-Funktion für die Behandelten geschätzt werden:

$$\widehat{ATT}_{X=1} = E[g_1(Z)|X = 1]$$

Der ATT ist der durchschnittliche Effekt der Behandlung auf diejenigen, die tatsächlich behandelt wurden ($X = 1$). Dies ist besonders interessant, wenn die Behandlung nicht zufällig vergeben wurde und wir daher mehr über die Wirkung der Behandlung auf diejenigen erfahren wollen, die sie tatsächlich erhalten haben.
Außerdem können sogenannte ($Z = z$)-bedingte Treatmenteffekte anhand von $g_1(z)$ geschätzt werden. Diese erlauben gezieltere Aussagen über die Wirksamkeit der Behandlung, da man nun auf eine Subpopulation $Z = z$ bedingt.

2.2 Datensatz

Der Datensatz `spf1.rda` beinhaltet Daten einer Schulleistungsvergleichsstudie für Kinder mit sonderpädagogischem Förderbedarf. Der Datensatz enthält folgende Variablen:

- Schulform: Inklusion an Regelschulen (1) versus Förderschulen (0) (Variable `schule`)
- Kompetenzwerte im Lesen (Variable `lesen`), wobei die Verteilung in der Gesamt-Population einen Mittelwert (μ) von 500 und eine Standardabweichung (σ) von 50 aufweist
- Die Kinder unterscheiden sich vor der Schulzuweisung in ihrer kognitiven Leistungsfähigkeit (Variable `KFT`)

Unsere Zielsetzung ist es, zu untersuchen, ob sich die durchschnittliche Leseleistung zwischen Regelschulen und Förderschulen unterscheidet. Dabei wollen wir die kognitive Leistungsfähigkeit als kontinuierliche Kovariate berücksichtigen.

```r
str(spf1)
```

```
'data.frame':	289 obs. of  3 variables:
 $ schule: Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 $ KFT   : num  7 10 13 8 12 20 15 6 16 10 ...
 $ lesen : num  208 243 309 309 533 ...
```

```r
summary(spf1)
```

```
 schule       KFT             lesen       
 0: 77   Min.   : 0.000   Min.   : 47.28  
 1:212   1st Qu.: 5.000   1st Qu.:221.92  
         Median : 9.000   Median :311.11  
         Mean   : 9.419   Mean   :308.40  
         3rd Qu.:13.000   3rd Qu.:378.51  
         Max.   :25.000   Max.   :616.58  
```

Der Datensatz enthält keine fehlenden Werte.

2.3 Deskriptive Datenanalyse

Zunächst möchten wir die Verteilung der SchülerInnen sowie der Lesekompetenzwerte (`lesen`) und der Kovariate (`KFT`) über die beiden Gruppen betrachten.

- Verteilung der SchülerInnen auf die Schulformen:

```
table(spf1$schule)
```

```
 0   1
77 212
```

- $n_{\text{Förderschule};X=0} = 77$ und $n_{\text{Regelschule};X=1} = 212$

- Unbedingter Mittelwert und bedingte Mittelwerte der Lesekompetenz:

```
# Unbedingter Mittelwert
mean(spf1$lesen)
```

```
[1] 308.4032
```

```
# Gruppenspezifische Lesekompetenzwerte
library(car)
Tapply(lesen ~ schule, data = spf1, fun = mean)
```

```
       0        1
245.2624 331.3364
```

- $\bar{x} = 308,40$, $\bar{x}_{\text{Förderschule}} = 245,26$ und $\bar{x}_{\text{Regelschule}} = 331,34$

- Gruppenspezifische Werte auf der Kovariate `KFT`:

```
# Zentrierung der Kovariate am Gesamtmittelwert
library(jtools)
spf1$KFT_z <- center(data = spf1$KFT)
# Gruppenspezifische Kompetenzwerte
Tapply(KFT_z ~ schule, data = spf1, fun = mean)
```

```
         0          1 
-1.6134903  0.5860319 
```

SchülerInnen an der Förderschule haben eine durchschnittliche kognitive Leistungsfähigkeit, die um 1,61 Punkte unter dem Gesamtdurchschnitt der kognitiven Leistungsfähigkeit liegt. SchülerInnen an der Regelschule liegen um 0,59 Punkte über dem Gesamtdurchschnitt.

2.4 Bestimmung kausaler Effekte

2.4.1 Einfache Regression ohne Kovariaten

- Y sei eine kontinuierliche, normalverteilte Variable.
- X sei ein dichotomer Prädiktor.

Unser Ziel ist es, die kontinuierliche Variable Y durch den dichotomen Prädiktor X zu erklären:

$$\widehat{Y} = \alpha_0 + \alpha_1 X$$

Dieses Modell entspricht dem t-Test für unabhängige Stichproben und kann in R mit `lm(Y ~ 1 + X, data)` geschätzt werden.

2.4.1.1 Übertragung auf die Notation kausaler Effekte

Die einfache Regression ohne Kovariaten impliziert folgende Modellparameter:

- $E(Y|X,Z) = E(Y|X) = \widehat{Y} = \alpha_0 + \alpha_1 X$
- $E(Y|X=0) = \alpha_0$
- $E(Y|X=1) = \alpha_0 + \alpha_1$
- $g_0(Z) = \alpha_0$
- $g_1(Z) = \alpha_1$
- $\widehat{ATE} = E[g_1(Z)] = \alpha_1$
- $\widehat{ATT}_{X=0} = E[g_1(Z)|X=0] = \alpha_1$
- $\widehat{ATT}_{X=1} = E[g_1(Z)|X=1] = \alpha_1$

In diesem Modell ist also der Average Treatment Effect ($\widehat{ATE}$) gleich dem Average Treatment Effect of the Treated ($\widehat{ATT}_{X=1}$) und gleich dem Average Treatment Effect of the Untreated ($\widehat{ATT}_{X=0}$). Sie sind alle gleich α_1. Das entspricht im Regressionsmodell dem Regressionsgewicht. Inhaltlich stellt dieser Wert den Mittelwertsunterschied der beiden Gruppen auf Y dar.
Wichtig zu beachten ist, dass in diesem Modell der kausale Effekt des Treatments nur unter

der Annahme einer vollständigen Randomisierung unverzerrt geschätzt wird. Mit anderen Worten, die Gültigkeit dieser Schätzung hängt davon ab, dass die Zuteilung zum Treatment ($X = 1$) vollständig zufällig erfolgt ist.

2.4.1.2 Schätzung in R

Wir erstellen ein lineares Modell, um die Leseleistung auf Basis der Schulform zu schätzen:

```r
m1 <- lm(lesen ~ schule, spf1)
summary(m1)$coef
```

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 245.26238   11.66955 21.017300 5.109259e-60
schule1      86.07403   13.62495  6.317382 1.011166e-09
```

- $E(Y|X,Z) = E(Y|X) = \widehat{Y} = 245.26 + 86.07 \cdot schule$
- $\widehat{ATE} = \widehat{ATT}_{X=1} = \widehat{ATT}_{X=0} = \alpha_1 = 86.07.$

Der $\widehat{ATE}$, $\widehat{ATT}_{X=1}$ und $\widehat{ATT}_{X=0}$ sind alle gleich $\alpha_1 = 86.07$. Dieser Wert entspricht der Differenz der Gruppenmittelwerte.
Es gibt einen bedeutsamen Effekt der Schulform (unter Annahme vollständiger Randomisierung), $t(287) = 6,32, p < 0,001$. Dies bedeutet, dass die Leseleistung signifikant höher ist für SchülerInnen an Regelschulen im Vergleich zu SchülerInnen an Förderschulen, unter der Annahme, dass die SchülerInnen zufällig diesen Schulformen zugewiesen wurden.

2.4.2 Traditionelle ANCOVA

- Ein kontinuierlicher Prädiktor Z und ein dichotomer Prädiktor X sind vorhanden.
- Es wird angenommen, dass Z und X additive Effekte haben:
  - $\hat{Y} = \alpha_0 + \alpha_1 Z + \alpha_2 X$
  - In R würde dies durch `lm(Y ~ 1 + Z + X, data)` geschätzt.

2.4.2.1 Übertragung auf die Notation kausaler Effekte

Die traditionelle ANCOVA kann in Bezug auf kausale Effekte wie folgt interpretiert werden:

- $E(Y|X,Z) = \widehat{Y} = \alpha_0 + \alpha_1 Z + \alpha_2 X$
- $E(Y|X=0, Z=0) = \alpha_0$
- $E(Y|X=1, Z=0) = \alpha_0 + \alpha_2$
- $g_0(Z) = \alpha_0 + \alpha_1 Z$
- $g_1(Z) = \alpha_2$
- $\widehat{ATE} = E[g_1(Z)] = \alpha_2$
- $\widehat{ATT}_{X=1} = E[g_1(Z)|X=1] = \alpha_2$
- $\widehat{ATT}_{X=0} = E[g_1(Z)|X=0] = \alpha_2$
- Durchschnittlicher geschätzter kausaler Effekt ($\widehat{ATE}$) entspricht der Differenz adjustierter Mittelwerte α_2.
- Keine Interaktion zwischen Treatment und Kovariate
- Die Effektfunktion g_1 nimmt bei allen Ausprägungen der Kovariaten Z den gleichen Wert (α_2) an, d.h. der Effekt ist für alle Ausprägungen von Z konstant
- ATE und ATT sind identisch und werden beide anhand von α_2 geschätzt.

2.4.2.2 Schätzung in R

```
m2 <- lm(lesen ~ KFT_z + schule, spf1)
summary(m2)$coef
```

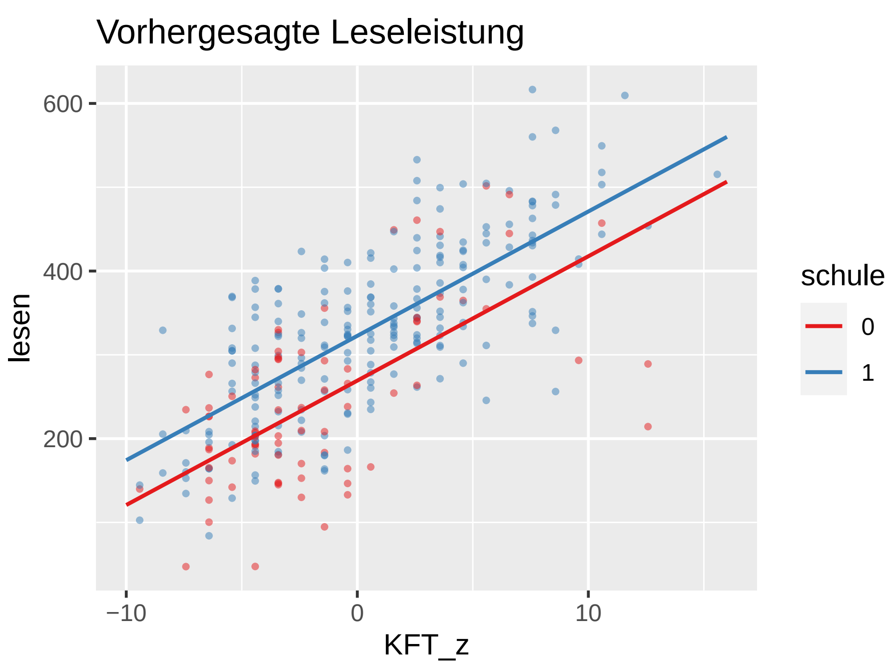
```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 269.20726  8.4186139 31.977623 1.927362e-96
KFT_z        14.84043  0.8841188 16.785559 1.735117e-44
schule1      53.43218  9.8803973  5.407898 1.347654e-07
```

- $E(Y|X,Z) = 269.21 + 14.84 *$ `KFT_z` $+ 53.43 *$ `schule`
- $g_0(Z) = 269.21 + 14.84 *$ `KFT_z`
- $g_1(Z) = 53.43$
- $\widehat{ATE} = \widehat{ATT}_{X=1} = \alpha_2 = 53.43$
- Es gibt einen bedeutsamen Treatmenteffekt nach Kontrolle der Gruppenunterschiede in der kognitiven Leistungsfähigkeit. Er beträgt 53,43 ($t(286) = 5{,}41, p < 0{,}001$) und ist damit kleiner, als wir es aufgrund der einfachen Regression geschätzt haben.

- Es werden adjustierte Mittelwerte verglichen, d.h. es wird beim Vergleich der mittleren Lesekompetenz zwischen den Schulformen für Unterschiede in der kognitiven Leistungsfähigkeit kontrolliert.
- Da kein Interaktionsterm zwischen Kovariate Z und Treatment X berücksichtigt wird, liegt hier die Annahme zugrunde, dass das Treatment unabhängig von Z wirkt, also dass das Treatment bei allen Kovariatenausprägungen in gleicher Weise wirkt. Diese Annahme ist oft nicht zutreffend. Wir werden sie daher später aufheben.

2.4.2.3 Visualisierung

```r
library(sjPlot)
plot_model(m2, type = "pred",
           terms = c("KFT_z", "schule"),
           show.data = TRUE,  # anzeigen der Datenpunkte auf dem Plot
           dot.size = 1,  # definieren der Größe der Datenpunkte
           title = "Vorhergesagte Leseleistung",
           ci.lvl = NA)  # Konfidenzintervall nicht anzeigen
```



In diesem Plot zeigt die X-Achse die zentrierten Werte von `KFT_z`, die Y-Achse zeigt die vorhergesagten Werte der Leseleistung `lesen`, und die Farbe der Punkte und Linien zeigen die Zugehörigkeit zur Schulform an. Beachten Sie, dass das Argument `ci.lvl = NA` das Zeichnen des Konfidenzintervalls unterdrückt. Wenn Sie ein Konfidenzintervall zeichnen möchten, können Sie `ci.lvl` auf das gewünschte Konfidenzniveau setzen (z.B. `ci.lvl = 0.95` für ein 95% Konfidenzintervall).

2.4.3 g-ANCOVA

- Ein kontinuierlicher Prädiktor Z und ein dichotomer Prädiktor X
- Im Gegensatz zur traditionellen ANCOVA wird nun ein Interaktionseffekt zwischen Treatment und Kovariate angenommen.
  - $\hat{Y} = \alpha_0 + \alpha_1 Z + \alpha_2 X + \alpha_3 ZX$
  - `lm(Y ~ 1 + Z * X, data)`

2.4.3.1 Übertragung auf die Notation kausaler Effekte

- $E(Y|X, Z) = \hat{Y} = \alpha_0 + \alpha_1 Z + \alpha_2 X + \alpha_3 ZX$
- $E(Y|X = 0, Z = 0) = \alpha_0$
- $E(Y|X = 1, Z = 0) = \alpha_0 + \alpha_2$
- $g_0(Z) = \alpha_0 + \alpha_1 Z$
- $g_1(Z) = \alpha_2 + \alpha_3 Z$
- $\widehat{ATE} = E(g_1(Z)) = \alpha_2 + \alpha_3 E(Z)$
- Wenn die Kovariate am Gesamtmittelwert zentriert wurde, gilt $E(Z) = 0$. Unter dieser Voraussetzung folgt für den ATE: $\widehat{ATE} = \alpha_2 + \alpha_3 E(Z) = \alpha_2$
- $\widehat{ATT}_{X=1} = E[g_1(Z)|X = 1] = \alpha_2 + \alpha_3 \cdot E(Z|X = 1)$
- $\widehat{ATT}_{X=0} = E[g_1(Z)|X = 0] = \alpha_2 + \alpha_3 \cdot E(Z|X = 0)$
- Im Gegensatz zur klassischen ANCOVA hängt der Effekt hier von Z ab, denn $g_1(Z) = \alpha_2 + \alpha_3 Z$.

2.4.3.2 Schätzung in R

```
m3 <- lm(lesen ~ KFT_z * schule, spf1)
summary(m3)$coef
```

```
                Estimate Std. Error   t value     Pr(>|t|)
(Intercept)   266.163593   8.776199 30.327889 3.298191e-91
KFT_z          12.954039   1.785389  7.255585 3.801220e-12
schule1        56.117487  10.116192  5.547294 6.626204e-08
KFT_z:schule1   2.497901   2.054494  1.215823 2.250584e-01
```

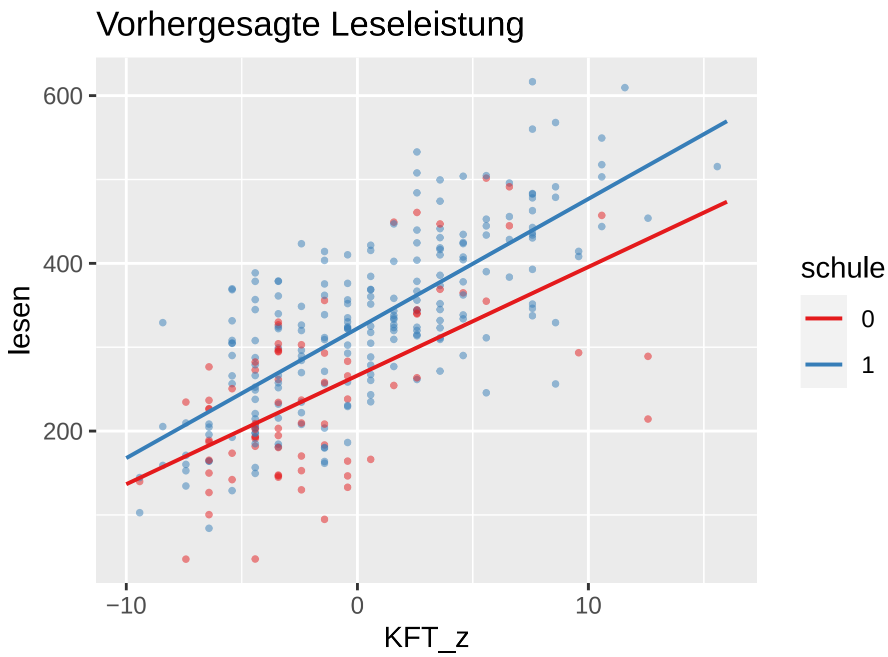
- $E(Y|X,Z) = 266.16 + 12.95$ * `KFT_z` + 56.12 * `schule` + 2.5 * `schule` * `KFT_z`
- $g_0(Z) = 266.16 + 12.95$ * `KFT_z`
- $g_1(Z) = 56.12 + 2.5$ * `KFT_z`
- $\widehat{ATE} = \alpha_2 = 56.12$
- $\widehat{ATT}_{X=1} = \alpha_2 + \alpha_3 * E(Z|X=1) = 56.12 + 2.5 * 0.59 = 57.58$
- $\widehat{ATT}_{X=0} = \alpha_2 + \alpha_3 \cdot E(Z|X=0) = 56.12 + 2.5 * \text{-}1.61 = 52.09$
- Es gibt einen bedeutsamen mittleren Treatmenteffekt von 56,12 ($t(285) = 5,55, p < 0,001$).
- Für Kinder auf der Regelschule ($X = 1$) wird der durchschnittliche Effekt auf 57,58 geschätzt, für Kinder auf der Förderschule beträgt die durchschnittliche Effektschätzung 52,09.
- Der Interaktionseffekt zwischen `schule` und `KFT_z` ist nicht signifikant ($t(285) = 1,21, p = 0,225$).
- Im Rahmen der Regressionsanalyse wie Sie sie im vergangegen Semester kennen gelernt haben würden wir daher aus statistischer Sicht das vorherige Modell der klassischen ANCOVA (ohne Interaktion) bevorzugen.
- Im Rahmen der Schätzung kausaler Effekte verfolgen wir das Ziel, die Effekte möglichst genau zu schätzen und die Strong Ignorability Assumption zu erfüllen. Daher kann es hier sinnvoll sein, auch den (nicht-signfikanten) Interaktionsterm weiterhin zu berücksichtigen.

Strong Ignorability Assumption

Die Strong Ignorability Assumption ist eine zentrale Annahme bei der Schätzung kausaler Effekte. Sie besagt, dass die Treatmentzuweisung unabhängig von den potential Outcomes ist, wenn wir alle relevanten Kovariaten berücksichtigen. Wenn diese Annahme erfüllt ist, können wir kausale Effekte aus beobachteten Daten schätzen, indem wir Unterschiede in den Kovariaten zwischen den Gruppen kontrollieren.

2.4.3.3 Visualisierung

```
plot_model(m3, type = "pred",
           terms = c("KFT_z", "schule"),
           show.data = TRUE, dot.size = 1,
           title = "Vorhergesagte Leseleistung",
           ci.lvl = NA)
```



2.5 EffectLiteR

Mit dem `EffectLiteR` Paket kann man alle oben besprochenen Effekte berechnen, mit dem Vorteil, dass die Effekte direkt ausgegeben werden und nicht selbst berechnet werden müssen. Darüber hinaus bietet es Funktionen zur Visualisierung dieser Effekte, was die Interpretation und Kommunikation der Ergebnisse erleichtert.

Mit dem Ausführen der folgenden Befehle wird das Paket installiert, geladen und die dazugehörige Dokumnentation aufgerufen:

```
install.packages("EffectLiteR")
library(EffectLiteR)
?effectLite()
```

2.5.1 Befehl mit Argumenten

Funktion	Bedeutung
`effectLite(` `y,`	Bestimmt die kausalen Effekte für abhängige Variable,

Funktion	Bedeutung
`x,`	Treatmentvariable mit
`control,`	Kontrollgruppe,
`z,`	kontinuierliche Kovariate(n)
`k,`	kategoriale Kovariate(n)
`data,`	Datensatz.
`interactions)`	`"all"` (Default) -> Zulassung aller möglichen Interaktionen (generalisierte ANCOVA) `"none"` -> klassisches ANCOVA-Modell, in dem Interaktion aus kontinuierlichen und kategorialen Kovariaten zugelassen werden `"no"` -> Modell ohne Interaktionen Weitere Definition von Interaktionen siehe `?effectLite`

- Mit dem Argument `control` wird angegeben, welche der Gruppen auf der Treatmentvariable X als Kontrollgruppe verwendet werden soll. In unserem Beispiel ist die Regelschule ($X = 1$) die Interventionsgruppe und die Förderschule ($X = 0$) die Kontrollgruppe. Wir schreiben daher `control = "0"`.
- Da wir hier maximal mit einer kontinuierlichen Kovariaten arbeiten, macht es keinen Unterschied, ob wir `interactions = "none"` oder `interactions = "no"` verwenden. Beide Schreibweisen führen dazu, dass keine Interaktion zwischen Kovariate Z und Treatmentvariable X zugelassen werden. Einen Unterschied würden die beiden Einstellungen erst machen, wenn wir zusätzlich zur kontinuierlichen Kovariaten Z noch eine kategoriale Kovariate K berücksichtigen würden. Mit `interactions = "none"` würden wir eine Interaktion der beiden Kovariaten zulassen, mit `interactions = "no"` würden wir das nicht tun. In beiden Fällen ist jedoch die Interaktion zwischen Kovariate(n) und Treatment unterbunden. Wollen wir diese Interaktion zulassen (generalisierte ANCOVA), verwenden wir den Default `interactions = "all"`.
- Mit `method` kann außerdem definiert werden, ob Fitfunktionen aus dem Regressionskontext (`method = "lm"`) oder aus dem Strukturgleichungskontext (`method = "sem"`) verwendet werden. Die Voreinstellung, die wir so übernehmen werden, ist `"sem"`. Die inferenzstatistischen Ergebnisse können daher leicht von denen abweichen, die wir oben mit den `lm`-Befehlen bestimmt hatten und es werden χ^2-Tests (statt F-Tests) für die Hypothesentests ausgegeben. Wenn Sie die Ergebnisse von oben exakt reproduzieren möchten, verwenden Sie probeweise `method = "lm"`.
- Alle weiteren Argumente, die Sie ggf. in der grafischen Benutzeroberfläche schon gesehen haben (z.B. `homoscedasticity`), brauchen wir nicht anzusprechen, weil wir sie auf ihrem Default belassen.

2.6 Übertragung der Modelle in EffectLiteR

Im folgenden ist die `effectLite()` Funktion mit den entsprechenden Einstellungen für die Argumente zur Reproduktion der oben eingeführten Modelle angegeben. Wir erläutern die einzelnen Abschnitte des Outputs der `effectLite()` Funktion anhand des letzten Modells (generalisierte ANCOVA).

2.6.1 t-Test ohne Kovariaten

```
effectLite(y = "lesen", x = "schule",
           control = "0",
           data = spf1)
```

2.6.2 Traditionelle ANCOVA (kontinuierliche Kovariate)

```
effectLite(y = "lesen", x = "schule",
           control = "0", data = spf1,
           z = "KFT_z", interactions = "none")
```

2.6.3 g-ANCOVA mit Interaktion

```
effectLite(y = "lesen", x = "schule",
           control = "0", data = spf1,
           z = "KFT_z")
```

2.6.3.1 Abschnitt: Regressionsmodel

```
 --------------------- Regression Model ---------------------

 E(Y|X,Z) = g0(Z) + g1(Z)*I_X=1
  g0(Z) = g000 + g001 * Z1
  g1(Z) = g100 + g101 * Z1

 Intercept Function g0(Z)  [Reference group: 0]
```

```
  Coefficient   Estimate      SE   Est./SE   p-value
         g000    266.164   9.668    27.531         0
         g001     12.954   1.967     6.586         0

Effect Function g1(Z)   [schule: 1 vs. 0]

  Coefficient   Estimate       SE   Est./SE   p-value
         g100     56.117   10.786     5.203     0.000
         g101      2.498    2.191     1.140     0.254
```

- Diese Tabelle ist zu lesen, wie die Koeffiziententabellen von `lm`-Modellen.
- Die Spalte `Coefficient` gibt an, um welchen Parameter es sich handelt:

  `g000`$= \alpha_0$, `g001`$= \alpha_1$, `g100`$= \alpha_2$, `g101`$= \alpha_3$
- Die Spalte `Estimate` enthält das geschätzte Regressionsgewicht.
- In der Spalte `SE` folgt der Standardfehler (standard error) des geschätzten Parameters
- dann das Verhältnis aus Parameterschätzung zu Standardfehler (`Est./SE`) und zuletzt
- der p-Wert, anhand dessen die Signifikanz abgelesen werden kann.

- Die Regresionsgleichung lautet also:

$$\widehat{lesen} = 266,16 + 12,95 \cdot KFT_z + 56,12 \cdot schule + 2,50 \cdot KFT_z \cdot schule$$

- Dabei entsprechen die Intercept- und Effektfunktion:

$$g_0(Z) = 266,16 + 12,95 * KFT_z$$
$$g_1(Z) = 56,12 + 2,50 * KFT_z$$

2.6.3.2 Abschnitt: Hypothesentests

```
--------------------- Main Hypotheses ---------------------

H0: No average effects: E[g1(Z)] = 0
H0: No covariate effects in control group: g0(Z) = constant
H0: No treatment*covariate interaction: g1(Z) = constant
H0: No treatment effects: g1(Z) = 0

                                        Wald Chi-Square   df    p-value
No average effects                                 26.9    1   2.09e-07
No covariate effects in control group              43.4    1   4.51e-11
No treatment*covariate interaction                  1.3    1   2.54e-01
No treatment effects                               27.1    2   1.32e-06
```

- In diesem Modell werden vier Hypothesen getestet, deren Nullhypothese zunächst genannt wird. Im zweiten Abschnitt folgt für jede der Hypothesen der χ^2-Test.

- Für den durchschnittlichen kausalen Effekt lesen wir zum Beispiel ab, dass er signifikant von Null verschieden ist: $\chi^2(1) = 26,9; p < 0,001$.

- Die Interaktion von Treatment und Kovariate ist nicht signifikant, $\chi^2(1) = 1,30; p = 0,254$.

2.6.3.3 Exkurs

Die Hypothesentests und die Tests für die Regressionsparameter sind in unserem Beispiel mit einer Kovariaten teilweise ineinander überführbar. Der zweite Hypothesentest (`No covariate effects in control group`) ist identisch zum Test, dass der Regressionsparameter `g0001` gleich Null ist. Der dritte Hypothesentest (`No treatment*covariate interaction`) ist identisch zum Test, dass der Regressionsparameter `g101` gleich Null ist. Der vierte Hypothesentest (`No treatment effects`) ist identisch zum Test, dass der Regressionsparameter `g100` gleich Null ist.
Die p-Werte sind jeweils identisch, denn die χ^2-Prüfgrößen ergeben sich durch Quadrierung der z-Werte.

2.6.3.4 Abschnitt: Kausale Effekte

```
--------------------- Average Effects ---------------------
```

```
           Estimate     SE   Est./SE    p-value   Effect Size
E[g1(Z)]       56.1   10.8      5.19   2.09e-07          0.56


 --------------------- Effects given a Treatment Condition ---------------------

               Estimate     SE   Est./SE    p-value   Effect Size
E[g1(Z)|X=0]       52.1   10.6      4.92   8.79e-07         0.520
E[g1(Z)|X=1]       57.6   11.2      5.14   2.71e-07         0.575
```

- Hier stehen die Schätzungen für die kausalen Effekte (`Estimate`) mit Standardfehler (`SE`), z-Wert (`Est./SE`), p-Wert (`p-value`) und Effektgröße (`Effect Size`; hier Cohen's d).

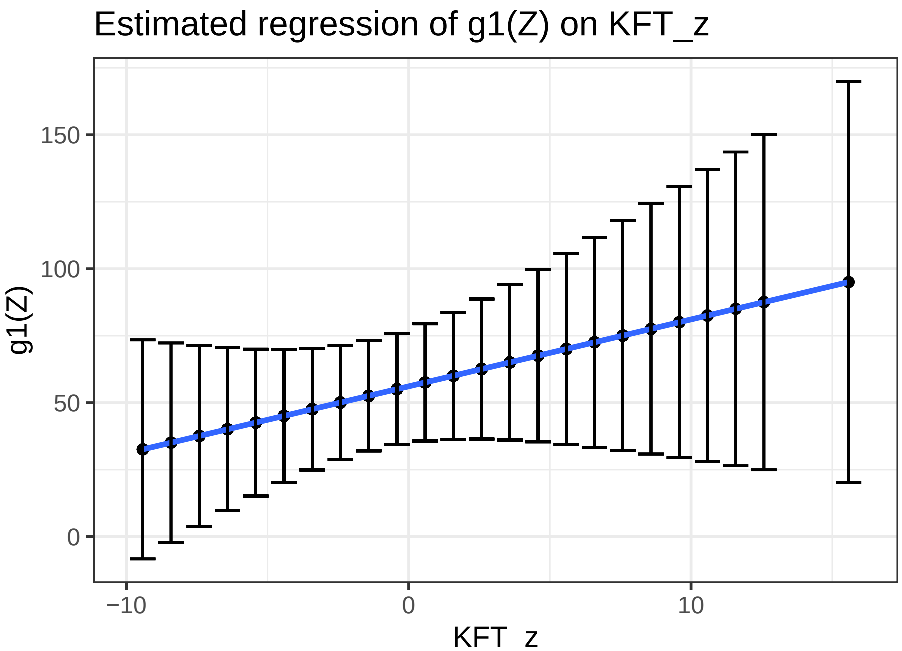
- $\widehat{ATE} = 56,1; d = 0,56$ -> signifikanter mittelstarker durchschnittlicher kausaler Effekt

- $\widehat{ATT}_{X=0} = 52,1; d = 0,52$ und
- $\widehat{ATT}_{X=1} = 57,6; d = 0,58$
- In dem hier dargestellten Modell (eine kontinuierliche, am Gesamtmittelwert *zentrierte* Kovariate) entspricht das Estimate für den Average Effect hier dem Wert von `g100` aus dem Abschnitt `Regression Model`. In anderen Fällen muss dies nicht so sein (z.B. bei unzentrierter Kovariate).

2.6.4 Grafische Darstellung der kovariaten-bedingten Effekte

```r
m1 <- effectLite(y = "lesen", x = "schule",
           control = "0", data = spf1,
           z = "KFT_z")

conditionalEffectsPlot(m1, zsel = "KFT_z", gxsel = "g1",
                       colour = "", show.ci = TRUE)
```


Estimated regression of g1(Z) on KFT_z
g1(Z)
150
100
50
0
−10
0
10
KFT_z