



Allgemeines Lineares Modell

BSc Psychologie, SoSe 2026

Joram Soch

(15) Kovarianzanalyse

Vorbemerkungen

Additive Kovarianzanalyse

Kovarianzanalyse mit Interaktion

Anwendung/Praxis

Selbstkontrollfragen

Vorbemerkungen

Additive Kovarianzanalyse

Kovarianzanalyse mit Interaktion

Anwendung/Praxis

Selbstkontrollfragen

Variabilität der Werte einer abhängigen Variable

- Die Werte einer AV unterscheiden sich im Allgemeinen zwischen experimentellen Einheiten.
- Die Schwankungen der Werte einer AV bezeichnet man als *Datenvariabilität*.
- Ziel jeder Datenanalyse ist die Erklärung von Datenvariabilität durch Zerlegung.

Datenvariabilität und Allgemeines Lineares Modell

- Eine spezielle Art der Quantifizierung von Datenvariabilität ist die Stichprobenvarianz.
- Einen additiven Zugang zur Dekomposition von Datenvariabilität bietet das ALM.
- Den folgenden Überlegungen liegt datenanalytisch die Kovarianzanalyse zugrunde.

Gesamtvariabilität = Primärvariabilität + Residualvariabilität

Primärvariabilität

Systematische Veränderung der AV, die allein auf Variation der UV zurückzuführen ist.

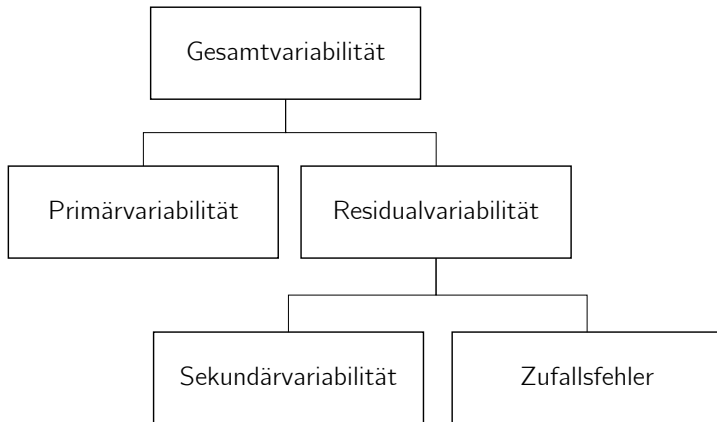
Residualvariabilität = Sekundärvariabilität + Zufallsfehler

Sekundärvariabilität

Systematische Veränderung der AV, die auf die Wirkung von unkontrollierten Störvariablen, nicht aber auf die Variation der UV zurückzuführen ist.

Zufallsfehler

Unsystematische Veränderung der AV, die weder auf die Variation der UV, noch auf den Einfluss von Störvariablen zurückzuführen ist.



Vorbemerkungen

Faktorielle und parametrische ALM-Designs

Faktorielle ALM-Designs:

- Designmatrizen mit 1en und 0en, manchmal -1 en.
- Betaparameter repräsentieren Erwartungswerte für Gruppen.
- Betaparameterschätzer repräsentieren Stichprobenmittel für Gruppen.
- $\Rightarrow$ T-Tests, einfaktorielle Varianzanalyse, zweifaktorielle Varianzanalyse

Parametrische ALM-Designs:

- Designmatrizen besitzen Spalten mit kontinuierlichen reellen Werten.
- Die Designmatrixspalten werden *Regressoren*, *Prädiktoren*, oder *Kovariaten* genannt.
- Betaparameter repräsentieren Steigungsparameter.
- Betaparameterschätzer ergeben sich als normalisierte Kovarianzen von Regressoren und Daten.
- Es besteht ein enger Bezug zur Theorie der Korrelation.
- $\Rightarrow$ einfache lineare Regression, multiple lineare Regression

Faktoriell-parametrische ALM-Designs:

- Designmatrizen mit mehreren faktoriellen und parametrischen Werten.
- Die parametrischen Regressoren werden oft als kontrollierte Kovariaten betrachtet.
- $\Rightarrow$ Kovarianzanalyse $\Leftrightarrow$ Allgemeines Lineares Modell

Wir betrachten hier exemplarisch den Fall 1 Faktor + 1 parametrischer Regressor.

Vorbemerkungen

Additive Kovarianzanalyse

Kovarianzanalyse mit Interaktion

Anwendung/Praxis

Selbstkontrollfragen

Additive Kovarianzanalyse

Modell der additiven einfaktoriellen Kovarianzanalyse

Im Modell der additiven einfaktoriellen Kovarianzanalyse möchte man die Erwartungswertparameter μ_{ij} der insgesamt $n = \sum_{i=1}^p n_i$ Datenvariablen

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad (1)$$

basierend auf den Leveln $i = 1, \dots, p$ des experimentellen Faktors und den entsprechenden Werten x_{ij} für $j = 1, \dots, n_i$ der Kovariate modellieren.

Dazu wählt man für jeden Erwartungswertparameter μ_{ij} zunächst einen Parameter μ_0 , der einen Datenvariablen-unspezifischen Offset modelliert. Weiterhin wählt man einen Parameter α_i , der den Beitrag des i ten Levels des experimentellen Faktors modelliert und schließlich einen Gewichtungssparameter β_0 , der den Beitrag der Kovariatenwerten x_{ij} zu μ_{ij} quantifiziert.

Das Modell für den Erwartungswertparameter von y_{ij} für $i = 1, \dots, p$, $j = 1, \dots, n_i$ nimmt im Fall der additiven einfaktoriellen Kovarianzanalyse also folgende Form an:

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \quad \text{mit} \quad \alpha_1 := 0, \quad (2)$$

wobei $\alpha_1 := 0$ wie im Falle der einfaktoriellen Varianzanalyse eine Überparameterisierung des Modells verhindert.

Man beachte, dass die Form des i ten Erwartungswertparameters damit eine einfache lineare Regression für jedes Level $i = 1, \dots, p$ des experimentellen Faktors definiert, wobei $\mu_0 + \alpha_i$ als Faktorlevel-spezifische Offsetparameter interpretiert werden können und der Steigungsparameter β_0 für alle Faktorlevel identisch ist.

Definition (Modell der additiven einfaktoriellen Kovarianzanalyse)

y_{ij} mit $i = 1, \dots, p$, $j = 1, \dots, n_i$ sei die Zufallsvariable, die den zum i ten Level des Faktors gehörenden j ten Datenpunkt modelliert und x_{ij} sei der entsprechende Wert der Kovariate. Dann hat das *Modell der additiven einfaktoriellen Kovarianzanalyse* die strukturelle Form

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} \quad \text{mit} \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, \dots, p, j = 1, \dots, n_i, \sigma^2 > 0 \quad (3)$$

und die Datenverteilungsform

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{u.v. für} \quad i = 1, \dots, p, j = 1, \dots, n_i, \quad (4)$$

wobei

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \quad \text{mit} \quad \alpha_1 := 0. \quad (5)$$

Bemerkungen

- Die Form des ij ten Erwartungswertparameters

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \quad \text{mit} \quad \alpha_1 := 0 \quad (6)$$

definiert eine einfache lineare Regression für jedes $i = 1, \dots, p$, wobei

- $\mu_0 + \alpha_i$ ein Faktorlevel-spezifischer Offsetparameter ist und
- der Steigungsparameter β_0 für alle Faktorlevel identisch ist.

Theorem (Designmatrixform der additiven einfaktoriellen Kovarianzanalyse)

Gegeben sei die strukturelle Form eines additiven einfaktoriellen Kovarianzanalysemodells mit Anzahl der Faktorlevel p und es sei $n := \sum_{i=1}^p n_i$ die Gesamtanzahl an Datenvariablen. Dann hat dieses Modell die Designmatrixform

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (7)$$

wobei

$$y := \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & & 0 & x_{11} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 0 & & 0 & x_{1n_1} \\ 1 & 1 & & 0 & x_{21} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 1 & & 0 & x_{2n_2} \\ \vdots & \vdots & & \dots & \vdots \\ 1 & 0 & & 1 & x_{p1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & 0 & & 1 & x_{pn_p} \end{pmatrix} \in \mathbb{R}^{n \times (p+1)}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \\ \beta_0 \end{pmatrix} \in \mathbb{R}^{p+1} \quad \text{und} \quad \sigma^2 > 0. \quad (8)$$

Bemerkungen

- Das Theorem ergibt sich direkt mit den Regeln der Matrixmultiplikation.

Additive Kovarianzanalyse

Designmatrix der additiven einfaktoriellen Kovarianzanalyse ($n = 18$, $p = 3$, $p + 1 = 4$)

Anwendungsbeispiel

experimenteller Faktor/Studiengruppe ($p := 2, i = 1, 2$)

- Face-to-face- vs. Online-Psychotherapie

parametrischer Regressor/Kovariate ($x_{ij}, j = 1, \dots, n_i$)

- Dauer der Depressionssymptomatik zu Beginn der Intervention

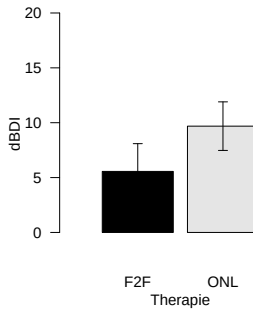
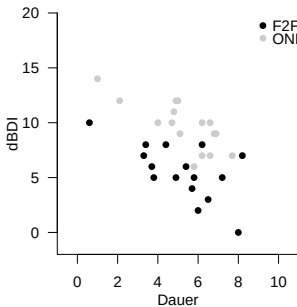
abhängige Variable/primäres Ergebnismaß ($y_{ij}, j = 1, \dots, n_i$)

- Prä-/Post-Differenzwerte von BDI-Scores

Beispieldatensatz ($n = 32$; hier: $i = 1, 2, j = 1, \dots, 10$)

	Therapie	Dauer	dBDI
1	F2F	3.7	6
2	F2F	5.4	6
3	F2F	3.3	7
4	F2F	8.2	7
5	F2F	5.7	4
6	F2F	3.4	8
7	F2F	6.0	2
8	F2F	6.5	3
9	F2F	6.2	8
10	F2F	4.4	8
17	ONL	5.0	12
18	ONL	6.9	9
19	ONL	6.6	7
20	ONL	6.2	7
21	ONL	6.8	9
22	ONL	6.6	10
23	ONL	5.1	9
24	ONL	1.0	14
25	ONL	6.2	10
26	ONL	4.9	12

Visualisierung



Modellformulierung

Modell 1 | keine Berücksichtigung der Kovariate

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{mit} \quad \mu_{ij} := \mu_0 + \alpha_i \quad \text{und} \quad \alpha_1 := 0 \quad \text{und} \quad \sigma^2 > 0 \quad (9)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad X := \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \end{pmatrix} \quad \text{und} \quad \sigma^2 > 0 \quad (10)$$

Modell 2 | additive Berücksichtigung der Kovariate

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{mit} \quad \mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \quad \text{und} \quad \alpha_1 := 0 \quad \text{und} \quad \sigma^2 > 0 \quad (11)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad X := \begin{pmatrix} 1 & 0 & x_{11} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{1n_1} \\ 1 & 1 & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n_2} \end{pmatrix}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_0 \end{pmatrix} \quad \text{und} \quad \sigma^2 > 0 \quad (12)$$

Additive Kovarianzanalyse

Implementation

```
fname      = "Daten/Kovarianzanalyse_A.csv"
D          = read.table(fname, sep = ",", header = TRUE)
n_i        = c(sum(D$Therapie == "F2F"), sum(D$Therapie == "ONL"))
y          = D$dBDI
n          = length(y)
XS         = list()
CS         = list()
XS[[1]]    = matrix(c(rep(1,n_i[1]),rep(1,n_i[2]),
                      rep(0,n_i[1]),rep(1,n_i[2])), nrow = n)
XS[[2]]    = matrix(c(rep(1,n_i[1]),rep(1,n_i[2]),
                      rep(0,n_i[1]),rep(1,n_i[2]),
                      D$Dauer), nrow = n)
CS[[1]]    = matrix(c(0,1), nrow = 2)
CS[[2]]    = matrix(c(0,1,0), nrow = 3)
B          = list()
S          = rep(NA,n)
T          = rep(NA,n)
for(i in 1:2){
  X         = XS[[i]]
  p         = ncol(X)
  beta_hat  = solve(t(X) %*% X) %*% t(X) %*% y
  eps_hat   = y - X %*% beta_hat
  sigsq_hat = (t(eps_hat) %*% eps_hat)/(n-p)
  c         = CS[[i]]
  t_num     = t(c) %*% beta_hat
  t_den     = sqrt(sigsqr_hat %*% t(c) %*% solve(t(X) %*% X) %*% c)
  t         = t_num/t_den
  B[[i]]    = beta_hat
  S[i]      = sigsq_hat
  T[i]      = t
}
```

Dateiname
Datensatz
Anzahl Datenpunkte pro Gruppe
Daten
Gesamtanzahldatenpunkte
Modell 1 und 2 Liste
Kontrastgewichtsvektorenliste
\mu_0 Regressor
\alpha_2 Regressor
\mu_0 Regressor
\alpha_2 Regressor
x Regressor
Kontrastgewichtsvektor \alpha_2
Kontrastgewichtsvektor \alpha_2
Betaparameterschätzerliste
Varianzparameterschätzervektor
T-Statistik-Vektor
Iteration über Modelle
Designmatrix
Anzahl Betaparameter
Betaparameterschätzer
Residuenvektor
Varianzparameterschätzer
Kontrastgewichtsvektor
Zähler der Zweistichproben-T-Teststatistik
Nenner der Zweistichproben-T-Teststatistik
Wert der Zweistichproben-T-Teststatistik
Betaparameterschätzer
Varianzparameterschätzer
T-Statistik

Ergebnisse

Modell 1 | Betaparameterschätzer : 5.56 4.12
Modell 2 | Betaparameterschätzer : 10.05 4.2 -0.86
Modell 1 | Varianzparameterschätzer : 5.65
Modell 2 | Varianzparameterschätzer : 3.14
Modell 1 | T-Statistik ONL-Therapie : 4.91
Modell 2 | T-Statistik ONL-Therapie : 6.7

In Modell 2 ist der Varianzparameterschätzer kleiner als in Modell 1.

⇒ Der parametrische Regressor erklärt Datenvarianz.

Die Betaparameterschätzer für den ONL-Therapieeffekt sind in Modell 1 und Modell 2 ähnlich.

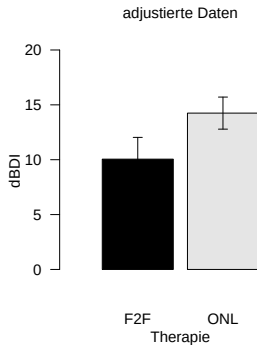
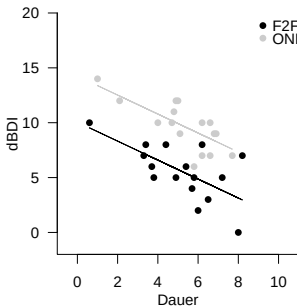
⇒ Die T-Statistik (Signal-zu-Rauschen-Verhältnis) für den ONL-Therapieeffekt ist Modell 2 höher.

Prinzipiell kann man die beobachteten Daten für den geschätzten Dauer-Effekt durch

$$\hat{y}_{ij}^{\text{adj}} = y_{ij} - \hat{\beta}_0 x_{ij} \quad (13)$$

“korrigieren” und erhält so einen “adjustierten Datensatz”.

Visualisierung



Vorbemerkungen

Additive Kovarianzanalyse

Kovarianzanalyse mit Interaktion

Anwendung/Praxis

Selbstkontrollfragen

Kovarianzanalyse mit Interaktion

Modell der einfaktoriellen Kovarianzanalyse mit Interaktion

Wie im Modell der additiven einfaktoriellen Kovarianzanalyse möchte man im Modell der einfaktoriellen Kovarianzanalyse mit Interaktion die Erwartungswertparameter μ_{ij} der insgesamt $n = \sum_{i=1}^p n_i$ Datenvariablen

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad (14)$$

basierend auf den Leveln $i = 1, \dots, p$ des experimentellen Faktors und den entsprechenden Werten x_{ij} für $j = 1, \dots, n_i$ der Kovariate modellieren. Über das Szenario der additiven einfaktoriellen Kovarianzanalyse hinaus möchte man dabei explizit einen Faktorlevel-spezifischen Einfluss der Kovariate auf die Erwartungswertparameter zulassen.

Dazu wählt man für jeden Erwartungswertparameter μ_{ij} zunächst wieder einen Parameter μ_0 , der einen Datenvariablen-unspezifischen Offset modelliert. Weiterhin wählt man einen Parameter α_i , der den Beitrag des i ten Levels des experimentellen Faktors modelliert. Den Wichtungsparmeter des Beitrags der Kovariaten x_{ij} zu μ_{ij} modelliert man nun hier mithilfe eines Faktorlevel-unspezifischen Parameter β_0 und eines Faktorlevel-spezifischen Parameters γ_i für $i = 1, \dots, p$. Das Modell für den Erwartungswertparameter von y_{ij} für $i = 1, \dots, p$, $j = 1, \dots, n_i$ nimmt im Fall der einfaktoriellen Kovarianzanalyse mit Interaktion damit folgende Form an:

$$\mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \quad \text{mit} \quad \alpha_1 := \gamma_1 := 0, \quad (15)$$

wobei $\alpha_1 := \gamma_1 := 0$ wie im Fall der zweifaktoriellen Varianzanalyse mit Interaktion eine Überparameterisierung des Modells verhindert.

Man beachte, dass die Form des i ten Erwartungswertparameters damit eine einfache lineare Regression für jedes Level $i = 1, \dots, p$ des experimentellen Faktors definiert, bei $\mu_0 + \alpha_i$ als Faktorlevel-spezifische Offsetparameter interpretiert werden können und sich die Steigungsparameter $\beta_0 + \gamma_i$ über die Faktorlevel unterscheiden können. Es werden also Unterschiede (zwischen den Faktorleveln) von Unterschieden (zwischen den Erwartungswertparametern eines Faktorlevels in Abhängigkeit vom Wert der Kovariate), d.h. Interaktionen, mit modelliert.

Definition (Modell der einfaktoriellen Kovarianzanalyse mit Interaktion)

y_{ij} mit $i = 1, \dots, p$, $j = 1, \dots, n_i$ sei die Zufallsvariable, die den zum i ten Level des Faktors gehörenden j ten Datenpunkt modelliert und x_{ij} sei der entsprechende Wert der Kovariate. Dann hat das *Modell einfaktoriellen Kovarianzanalyse mit Interaktion* die strukturelle Form

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} \quad \text{mit} \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad \text{u.i.v. für} \quad i = 1, \dots, p, j = 1, \dots, n_i, \sigma^2 > 0 \quad (16)$$

und die Datenverteilungsform

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{u.v. für} \quad i = 1, \dots, p, j = 1, \dots, n_i, \quad (17)$$

wobei

$$\mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \quad \text{mit} \quad \alpha_1 := \gamma_1 := 0. \quad (18)$$

Bemerkungen

- Die Form des ij ten Erwartungswertparameters

$$\mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \quad \text{mit} \quad \alpha_1 := \gamma_1 := 0 \quad (19)$$

definiert eine einfache lineare Regression für jedes $i = 1, \dots, p$, wobei

- $\mu_0 + \alpha_i$ ein Faktorlevel-spezifischer Offsetparameter ist und
- $\beta_0 + \gamma_i$ ein Faktorlevel-spezifischer Steigungsparameter ist.

Theorem (Designmatrixform der einfaktoriellen Kovarianzanalyse mit Interaktion)

Gegeben sei die strukturelle Form eines einfaktoriellen Kovarianzanalysemodells mit Interaktion mit Anzahl der Faktorebene p und es sei $n := \sum_{i=1}^p n_i$ die Gesamtanzahl an Datenvariablen. Dann hat dieses Modell die Designmatrixform

$$y = X\beta + \varepsilon \quad \text{mit} \quad \varepsilon \sim N(0_n, \sigma^2 I_n), \quad (20)$$

wobei

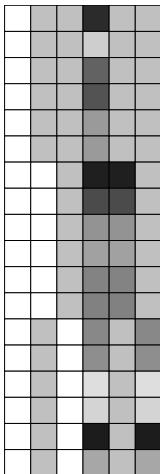
$$y := \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{p1} \\ \vdots \\ y_{pn_p} \end{pmatrix}, \quad X := \begin{pmatrix} 1 & 0 & & 0 & x_{11} & 0 & & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 0 & x_{1n_1} & 0 & & 0 \\ 1 & 1 & & 0 & x_{21} & x_{21} & & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 1 & & 0 & x_{2n_2} & x_{2n_2} & & 0 \\ \vdots & \vdots & & \dots & \vdots & \vdots & & \dots \\ 1 & 0 & & 1 & x_{p1} & 0 & & x_{p1} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & & 1 & x_{pn_p} & 0 & & x_{pn_p} \end{pmatrix} \in \mathbb{R}^{n \times (2p)}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \vdots \\ \alpha_p \\ \beta_0 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix} \in \mathbb{R}^{2p} \quad \text{und} \quad \sigma^2 > 0. \quad (21)$$

Bemerkungen

- Das Theorem ergibt sich direkt mit den Regeln der Matrixmultiplikation.

Kovarianzanalyse mit Interaktion

Designmatrix der einfaktoriellen Kovarianzanalyse m. Interaktion ($n = 18$, $p = 3$, $2p = 6$)



Anwendungsbeispiel

experimenteller Faktor/Studiengruppe ($p := 2, i = 1, 2$)

- Face-to-face- vs. Online-Psychotherapie

parametrischer Regressor/Kovariate ($x_{ij}, j = 1, \dots, n_i$)

- Digitalaffinität der Studienteilnehmer

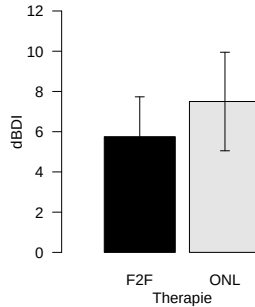
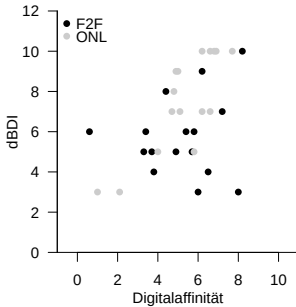
abhängige Variable/primäres Ergebnismaß ($y_{ij}, j = 1, \dots, n_i$)

- Prä-/Post-Differenzwerte von BDI-Scores

Beispieldatensatz ($n = 32$; hier: $i = 1, 2$, $j = 1, \dots, 10$)

	Therapie	Dig.Affin	dBDI
1	F2F	3.7	5
2	F2F	5.4	6
3	F2F	3.3	5
4	F2F	8.2	10
5	F2F	5.7	5
6	F2F	3.4	6
7	F2F	6.0	3
8	F2F	6.5	4
9	F2F	6.2	9
10	F2F	4.4	8
17	ONL	5.0	9
18	ONL	6.9	10
19	ONL	6.6	7
20	ONL	6.2	7
21	ONL	6.8	10
22	ONL	6.6	10
23	ONL	5.1	7
24	ONL	1.0	3
25	ONL	6.2	10
26	ONL	4.9	9

Visualisierung



Modellformulierung

Modell 1 | additive Berücksichtigung der Kovariate

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{mit} \quad \mu_{ij} := \mu_0 + \alpha_i + \beta_0 x_{ij} \quad \text{und} \quad \alpha_1 := 0 \quad \text{und} \quad \sigma^2 > 0 \quad (22)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad X := \begin{pmatrix} 1 & 0 & x_{11} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{1n_1} \\ 1 & 1 & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n_2} \end{pmatrix}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_0 \end{pmatrix} \quad \text{und} \quad \sigma^2 > 0 \quad (23)$$

Modell 2 | additive und interaktive Berücksichtigung der Kovariate

$$y_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{mit} \quad \mu_{ij} := \mu_0 + \alpha_i + (\beta_0 + \gamma_i)x_{ij} \quad \text{und} \quad \alpha_1 := \gamma_1 := 0 \quad \text{und} \quad \sigma^2 > 0 \quad (24)$$

$$\Leftrightarrow y \sim N(X\beta, \sigma^2 I_n) \quad \text{mit} \quad X := \begin{pmatrix} 1 & 0 & x_{11} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_{1n_1} & 0 \\ 1 & 1 & x_{21} & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n_2} & x_{2n_2} \end{pmatrix}, \quad \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_0 \\ \gamma_2 \end{pmatrix} \quad \text{und} \quad \sigma^2 > 0 \quad (25)$$

Kovarianzanalyse mit Interaktion

Implementation

```
1290 fname = "Daten/Kovarianzanalyse_B.csv" # Dateiname
1291 D = read.table(fname, sep = ",", header = TRUE) # Datensatz
1292 n_i = c(sum(D$Therapie == "F2F"), sum(D$Therapie == "ONL")) # Anzahl Datenpunkte pro Gruppe
1293 y = D$dBDI # Daten
1294 n = length(y) # Gesamtanzahl Datenpunkte
1295 XS = list() # Model 1 und 2 Liste
1296 CS = list() # Kontrastgewichtsvektorenliste
1297 XS[[1]] = matrix(c(rep(1,n_i[1]), rep(1,n_i[2]), # \mu_0 Regressor
1298 rep(0,n_i[1]), rep(1,n_i[2]), # \alpha_2 Regressor
1299 D$Dig.Affin, nrow = n) # \beta_2 Regressor
1300 XS[[2]] = matrix(c(rep(1,n_i[1]), rep(1,n_i[2]), # \mu_0 Regressor
1301 rep(0,n_i[1]), rep(1,n_i[2]), # \alpha_2 Regressor
1302 D$Dig.Affin, # \beta_2 Regressor
1303 D$Dig.Affin*(c(rep(0,n_i[1]),rep(1,n_i[2]))), # \gamma_2 Regressor
1304 nrow = n) # Anzahl Zeilen der Designmatrix
1305 CS[[1]] = matrix(c(0,1,0), nrow = 3) # Kontrastgewichtsvektor \alpha_2
1306 CS[[2]] = matrix(c(0,1,0), nrow = 4) # Kontrastgewichtsvektor \alpha_2
1307 B = list() # Betaparameterschätzerliste
1308 S = rep(NaN,2) # Varianzparameterschätzervektor
1309 T = rep(NaN,2) # T-Statistik-Vektor
1310 for(i in 1:2){ # Iteration über Modelle
1311 X = XS[[i]] # Designmatrix
1312 p = ncol(X) # Anzahl Betaparameter
1313 beta_hat = solve(t(X) %*% X) %*% t(X) %*% y # Betaparameterschätzer
1314 eps_hat = y - X %*% beta_hat # Residuenvektor
1315 sigsq_hat = (t(eps_hat) %*% eps_hat)/(n-p) # Varianzparameterschätzer
1316 c = CS[[i]] # Kontrastgewichtsvektor
1317 t_num = t(c) %*% beta_hat # Zähler der Zweistichproben-T-Teststatistik
1318 t_den = sqrt(sigsqr_hat %*% t(c) %*% solve(t(X) %*% X) %*% c) # Nenner der Zweistichproben-T-Teststatistik
1319 t = t_num/t_den # Wert der Zweistichproben-T-Teststatistik
1320 B[[i]] = beta_hat # Betaparameterschätzer
1321 S[i] = sigsq_hat # Varianzparameterschätzer
1322 T[i] = t # T-Statistik
1323 }
1324 ...
```

Ergebnisse

```
Modell 1 | Betaparameterschätzer      :  2.73  1.7  0.58
Modell 2 | Betaparameterschätzer      :  5.03 -3.47  0.14  0.99
Modell 1 | Varianzparameterschätzer      :  3.92
Modell 2 | Varianzparameterschätzer      :  3.15
Modell 1 | T-Statistik ONL-Therapie          :  2.43
Modell 2 | T-Statistik ONL-Therapie          : -1.79
```

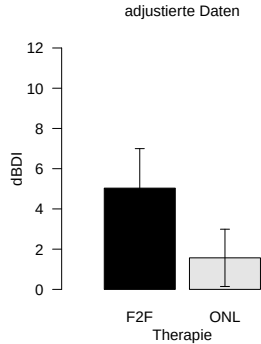
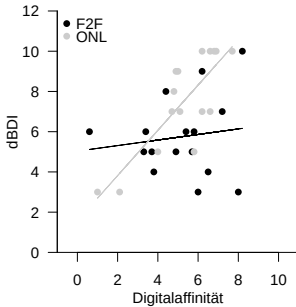
In Modell 1 wird der ONL-Therapieeffekt positiv, in Modell 2 negativ geschätzt.

Entsprechend ist die T-Statistik für ONL-Therapie in Modell 1 positiv, in Modell 2 negativ.

Die wahren, aber unbekannten Betaparameterwerte sind $\beta := (5, -3, 0, 1)^T$.

- ⇒ Die Abhängigkeit von Digitalaffinität kommt bei ONL zum Tragen, bei F2F nicht.
- ⇒ Der Haupteffekt der Therapieart ist schwierig zu beurteilen.
- ⇒ ONL-Therapie ist bei niedriger Digitalaffinität nicht so wirksam wie F2F-Therapie.
- ⇒ ONL-Therapie ist bei hoher Digitalaffinität wirksamer als F2F-Therapie.

Visualisierung



Vorbemerkungen

Additive Kovarianzanalyse

Kovarianzanalyse mit Interaktion

Anwendung/Praxis

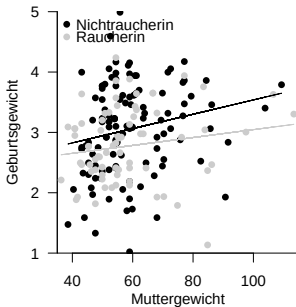
Selbstkontrollfragen

Datensatz: Geburtsgewicht und Muttergewicht für Mutter-Kind-Paare ($n = 189$; hier: $i = 1, \dots, 25$)

```
filename = "Daten/birth_weights.csv" # Dateiname  
D        = read.csv(filename)        # Datenframe
```

Birth_weight	Weight	Age	Smoker	Race	Hypertension	Visits
2.523	82.55374	19	no	black	no	0
2.551	70.30676	33	no	other	no	3
2.557	47.62716	20	yes	white	no	1
2.594	48.98794	21	yes	white	no	2
2.600	48.53434	18	yes	white	no	0
2.622	56.24541	21	no	other	no	0
2.637	53.52386	22	no	white	no	1
2.637	46.71998	17	no	other	no	1
2.663	55.79182	29	yes	white	no	1
2.665	51.25590	26	yes	white	no	0
2.722	43.09124	19	no	other	no	0
2.733	68.03880	19	no	other	no	1
2.751	43.09124	22	no	other	yes	0
2.750	48.53434	30	no	other	no	2
2.769	45.35920	18	yes	white	no	0
2.769	45.35920	18	yes	white	no	0
2.778	44.45202	15	no	black	no	0
2.782	53.52386	25	yes	white	no	3
2.807	54.43104	20	no	other	no	0
2.821	54.43104	28	yes	white	no	1
2.835	54.88463	32	no	other	no	2
2.835	45.35920	31	no	white	no	3
2.836	91.62558	36	no	white	no	1
2.863	54.43104	28	no	other	no	0
2.877	54.43104	25	no	other	no	2

Datensatz: Geburtsgewicht und Muttergewicht, getrennt nach Raucherstatus ($n = 189$)



Zweistichproben-T-Test: Geburtsgewicht, Nichtraucherinnen vs. Raucherinnen ($n_1 = 115$, $n_2 = 74$)

```
# Zweistichproben-T-Test
y1      = D$Birth_weight[D$Smoker=="no"]      # Gruppe 1
y2      = D$Birth_weight[D$Smoker=="yes"]     # Gruppe 2
n1      = length(y1)                          # Anzahl Gruppe 1
n2      = length(y2)                          # Anzahl Gruppe 2
n       = n1 + n2                             # Anzahl Datenpunkte
p       = 2                                   # Anzahl Regressoren
y       = matrix(c(y1,y2), nrow = n)          # Datenvektor
X       = matrix(c(rep(1,n1), rep(0,n2),      # Designmatrix
                  rep(0,n1), rep(1,n2))), ncol = p)
```

Datenpunkte pro Gruppe	:	115 74
Betaparameterschätzer	:	3.056 2.772
Varianzparameterschätzer	:	0.515
Zweistichproben-T-Teststatistik	:	2.653
Cohen's d	:	0.193
p-Wert	:	0.009

Kovarianzanalyse: T-Test auf Haupteffekt von Raucherstatus

```
# Kovarianzanalyse
D$Weight = D$Weight - mean(D$Weight) # Kovariate mittelwertzentrieren
y_1      = D$Birth_weight[D$Smoker=="no"] # Datenpunkte Gr. 1
y_2      = D$Birth_weight[D$Smoker=="yes"] # Datenpunkte Gr. 2
x_1      = D$Weight[D$Smoker=="no"] # Kovariatenwerte Gr. 1
x_2      = D$Weight[D$Smoker=="yes"] # Kovariatenwerte Gr. 2
n_1      = length(y_1) # Anzahl Datenpunkte Gr. 1
n_2      = length(y_2) # Anzahl Datenpunkte Gr. 2
n        = n_1 + n_2 # Anzahl Datenpunkte
p        = 4 # Anzahl Regressoren
y        = matrix(c(y_1, y_2), nrow = n) # Datenvektor
x        = matrix(c(x_1, x_2), nrow = n) # Kovariatenvektor
X        = matrix(c(rep(1,n_1), rep(1,n_2), # \mu_0 Regressor
                    rep(0,n_1), rep(1,n_2), # \alpha_2 Regressor
                    x, # \beta_0 Regressor
                    x*(c(rep(0,n_1), rep(1,n_2)))), # \gamma_2 Regressor
                  nrow = n) # Anzahl Zeilen der Designmatrix
c        = matrix(c(0,1,0,0), nrow = p) # Kontrastgewichtsvektor
```

	Estimate	Std. Error	t-value	Pr(> t)
Smoking	-0.2729737	0.1057379	-2.581606	0.01060802

Kovarianzanalyse: F-Test auf Effekt der Kovariate Muttergewicht

```
# Kovarianzanalyse
D$Weight      = D$Weight - mean(D$Weight)           # Kovariate mittelwertzentrieren
y_1           = D$Birth_weight[D$Smoker=="no"]      # Datenpunkte Gr. 1
y_2           = D$Birth_weight[D$Smoker=="yes"]     # Datenpunkte Gr. 2
x_1           = D$Weight[D$Smoker=="no"]            # Kovariatenwerte Gr. 1
x_2           = D$Weight[D$Smoker=="yes"]           # Kovariatenwerte Gr. 2
n_1           = length(y_1)                         # Anzahl Datenpunkte Gr. 1
n_2           = length(y_2)                         # Anzahl Datenpunkte Gr. 2
n             = n_1 + n_2                           # Anzahl Datenpunkte
p             = 4                                    # Anzahl Regressoren
y             = matrix(c(y_1, y_2), nrow = n)        # Datenvektor
x             = matrix(c(x_1, x_2), nrow = n)        # Kovariatenvektor
X             = matrix(c(rep(1,n_1), rep(1,n_2),     # \mu_0 Regressor
                        rep(0,n_1), rep(1,n_2),      # \alpha_2 Regressor
                        x,                          # \beta_0 Regressor
                        x*(c(rep(0,n_1), rep(1,n_2)))), # \gamma_2 Regressor
                        nrow = n)                   # Anzahl Zeilen der Designmatrix
p_0           = 2                                    # Anzahl Parameter reduziertes Modell
```

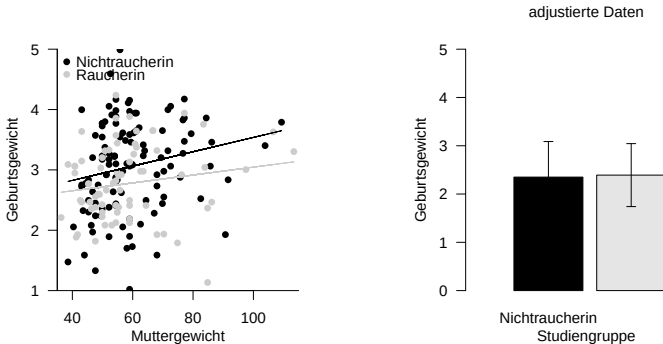
	numerator DFs	denominator DFs	F-value	Pr(>F)
Weight	2	185	3.389947	0.03581795

Kovarianzanalyse: T-Test auf Interaktion von Raucherstatus und Muttergewicht

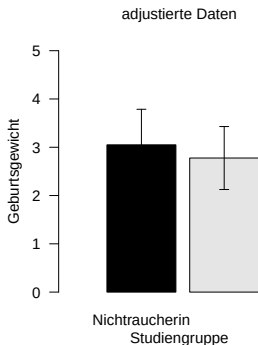
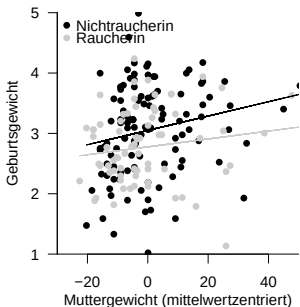
```
# Kovarianzanalyse
D$Weight = D$Weight - mean(D$Weight) # Kovariate mittelwertzentrieren
y_1      = D$Birth_weight[D$Smoker=="no"] # Datenpunkte Gr. 1
y_2      = D$Birth_weight[D$Smoker=="yes"] # Datenpunkte Gr. 2
x_1      = D$Weight[D$Smoker=="no"] # Kovariatenwerte Gr. 1
x_2      = D$Weight[D$Smoker=="yes"] # Kovariatenwerte Gr. 2
n_1      = length(y_1) # Anzahl Datenpunkte Gr. 1
n_2      = length(y_2) # Anzahl Datenpunkte Gr. 2
n        = n_1 + n_2 # Anzahl Datenpunkte
p        = 4 # Anzahl Regressoren
y        = matrix(c(y_1, y_2), nrow = n) # Datenvektor
x        = matrix(c(x_1, x_2), nrow = n) # Kovariatenvektor
X        = matrix(c(rep(1,n_1), rep(1,n_2), # \mu_0 Regressor
                    rep(0,n_1), rep(1,n_2), # \alpha_2 Regressor
                    x, # \beta_0 Regressor
                    x*(c(rep(0,n_1), rep(1,n_2)))), # \gamma_2 Regressor
                  nrow = n) # Anzahl Zeilen der Designmatrix
c        = matrix(c(0,0,0,1), nrow = p) # Kontrastgewichtsvektor
```

	Estimate	Std. Error	t-value	Pr(> t)
Smoking x Weight	-0.005338694	0.007470284	-0.7146574	0.4757216

Datensatz: Geburtsgewicht und Muttergewicht, getrennt nach Raucherstatus



Datensatz: Geburtsgewicht und Muttergewicht (mittelwertzentriert), getrennt nach Raucherstatus



Vorbemerkungen

Additive Kovarianzanalyse

Kovarianzanalyse mit Interaktion

Anwendung/Praxis

Selbstkontrollfragen

1. Erläutern Sie die Begriffe der Gesamt-/Primär-/Sekundär- und Residualvaribilität.
2. Geben Sie jeweils zwei Beispiele für parametrische bzw. faktorielle ALM-Designs an.
3. Geben Sie die Definition des Modells der additiven einfaktoriellen Kovarianzanalyse (EKA) wieder.
4. Erläutern Sie die Bedeutung der Parameter μ_0 , α_i und β_0 im Modell der additiven EKA.
5. Geben Sie die Designmatrixform des Modells der additiven EKA für $p := 2$ und $n_1 = n_2 = 3$ an.
6. Erläutern Sie den Unterschied zwischen dem Modell der einfaktoriellen Varianzanalyse und dem Modell der additiven einfaktoriellen Kovarianzanalyse.
7. Geben Sie die Definition des Modells der einfaktoriellen Kovarianzanalyse (EKA) mit Interaktion wieder.
8. Erläutern Sie die Bedeutung der Parameter μ_0 , α_i , β_0 und γ_i im Modell der EKA mit Interaktion.
9. Geben Sie die Designmatrixform des Modells der EKA mit Interaktion für $p := 2$ und $n_1 = n_2 = 3$ an.
10. Erläutern Sie den Unterschied zwischen dem Modell der additiven einfaktoriellen Kovarianzanalyse und dem Modell der einfaktoriellen Kovarianzanalyse mit Interaktion.