Tutorium Allgemeines Lineares Modell

BSc Psychologie SoSe 2022

13. Termin: (11) Zweifaktorielle Varianzanalyse

Belinda Fleischmann

Selbstkontrollfragen

1. Erläutern Sie das Anwendungsszenario der zweifaktoriellen Varianzanalyse (ZVA).
2. Aus wie vielen Datenpunkten besteht ein Datensatz eines 3 x 4 ZVA Designs mit 10 Datenpunkten pro Zelle?
3. Erläutern Sie die intuitive Bedeutung eines Haupteffektes in einem ZVA Design.
4. Erläutern Sie die intuitive Bedeutung einer Interkation ein einem ZVA Design.
5. Geben Sie die Definition des additiven Modells der ZVA mit Referenzgruppe (RG) wieder.
6. Erläutern Sie die Bedeutung der Parameter μ_0, α_2 und β_2 im additiven Modell der ZVA mit RG.
7. Bestimmen Sie μ_{ij} für $\mu_0 := 2, \alpha_2 = -1$ und $\beta_2 := 3$ im additiven Modell der ZVA mit RG.
8. Geben Sie die Designmatrixform des Modells einer additiven 2 x 2 ZVA mit RG für $n_{ij} := 1$ an.
9. Geben Sie die Designmatrixform des Modells einer additiven 2 x 2 ZVA mit RG für $n_{ij} := 3$ an.
10. Geben Sie die Definition des Modells der ZVA mit Interaktion und RG wieder.
11. Erläutern Sie die Bedeutung der Parameter μ_0, α_2, β_2 und γ_{22} im Modell der ZVA mit Interaktion und RG.
12. Geben Sie die Designmatrixform des Modells einer 2 x 2 ZVA mit Interaktion und RG für $n_{ij} := 1$ an.
13. Geben Sie die Designmatrixform des Modells einer 2 x 2 ZVA mit Interaktion und RG für $n_{ij} := 3$ an.
14. Geben Sie das Theorem zur Betaparameterschätzung im additiven 2 x 2 ZVA Modell mit RG wieder.
15. Geben Sie das Theorem zur Betaparameterschätzung im 2 x 2 ZVA Modell mit Interaktion und RG wieder.

Selbstkontrollfragen

16. Erzeugen Sie einen Beispieldatensatz des Anwendungsbeispiels einer 2 x 2 ZVA mit den Faktoren Therapie (F2F, ONL) und Alter (YA,OA) basierend auf dem Modell der ZVA mit Interaktion und Referenzgruppe für von Ihnen selbst gewählte Parameterwerte $\mu_0, \alpha_2, \beta_2, \gamma_{22}$ und σ^2. Visualisieren Sie den erzeugten Datensatz mithilfe eines Barplots, eines Barplots und eines Lineplots.
17. Bestimmen Sie Betaparameterschätzer für den von Ihnen erzeugten Datensatz im Modell der additiven 2 x 2 ZVA mit Referenzgruppe und im Modell der 2 x 2 mit Interaktion und Referenzgruppe. Erläutern Sie die Gemeinsamkeiten und Unterschiede der Betaparameterschätzer vor dem Hintergrund der von Ihnen gewählten wahren, aber unbekannten, Parameterwerte des Datensatzes.
18. Führen Sie basierend auf dem von Ihnen erzeugten Datensatz einen F-Test für den Haupteffekt des Faktors Therapie, einen F-Test für den Haupteffekt des Faktors Alter und einen F-Test für die Interaktion der Faktoren durch. Dokumentieren und erläutern Sie Ihre Ergebnisse.

1. Erläutern Sie das Anwendungsszenario der zweifaktoriellen Varianzanalyse (ZVA).

- Eine univariate abhängige Variable bestimmt an randomisierten experimentellen Einheiten.
- Zwei diskrete unabhängige Variablen, die mindestens zweistufig sind.
- Die unabhängigen Variablen werden **Faktoren** genannt.
- Die Stufen der Faktoren werden auch **Faktorlevel** genannt.
- Jedes Level eines Faktors wird mit allen Level des anderen Faktors kombiniert.
- Die Kombination zweier spezifischer Faktorlevel wird **Zelle** des Designs genannt.

Zweifaktorielle Studiendesigns werden üblicherweise anhand ihrer Faktorlevel bezeichnet

2 x 2 ANOVA : Faktor A mit Level 1,2	Faktor B mit Level 1,2
2 x 3 ANOVA : Faktor A mit Level 1,2	Faktor B mit Level 1,2,3
4 x 2 ANOVA : Faktor A mit Level 1,2,3,4	Faktor B mit Level 1,2
3 x 1 ANOVA : Faktor A mit Level 1,2,3	Faktor B mit Level 1

Die Zellen eines 2 x 2 Designs werden auch als Gruppen bezeichnet.

2. Aus wie vielen Datenpunkten besteht ein Datensatz eines 3 x 4 ZVA Designs mit 10 Datenpunkten pro Zelle?

$3 \times 4 = 12$ Gruppen 10 Datenpunkte pro Zelle $\times$ 12 Gruppen $= 120$ Datenpunkte

3. Erläutern Sie die intuitive Bedeutung eines Haupteffektes in einem ZVA Design.

- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor A*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor A, jeweils gemittelt über die zwei Level von Faktor B, unterscheiden.
- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor B*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor B, jeweils gemittelt über die zwei Level von Faktor A, unterscheiden.
- Intuitiv beziehen sich Haupteffekte also auf (marginale) Unterschiede (Differenzen), während sich Interaktionen auf Unterschiede von Unterschieden (Differenzen von Differenzen) beziehen.

4. Erläutern Sie die intuitive Bedeutung einer Interkation ein einem ZVA Design.

- Intuitiv spricht man vom Vorliegen einer *Interaktion der Faktoren A und B*, wenn der Unterschied der Gruppenmittelwerte von Faktor A zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor B ausgeprägt ist bzw. wenn der Unterschied der Gruppenmittelwerte von Faktor B zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor A ausgeprägt ist.
- Das Vorhandensein einer Interaktion besagt lediglich, dass sich die Unterschiede der Gruppenmittelwerte zwischen den Leveln eines experimentellen Faktors in Abhängigkeit von den Leveln des anderen experimentellen Faktors ändern, es macht aber keine Aussage darüber, warum dies so ist.

5. Geben Sie die Definition des additiven Modells der ZVA mit Referenzgruppe (RG) wieder.

Definition (Modell der additiven ZVA mit Referenzgruppe)

y_{ijk} mit $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij}$ sei die Zufallsvariable, die den $k$ten Datenpunkt zum $i$ten Level von Faktor A und dem $j$ten Level von Faktor B in einem ZVA Anwendungsszenario modelliert. Dann hat das *Modell der additiven ZVA mit Referenzgruppe* die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij} \tag{1}$$

und die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. für } i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij} \tag{2}$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j \text{ für } i = 1, \ldots, I, j = 1, \ldots, J \text{ mit } \alpha_1 := \beta_1 := 0. \tag{3}$$

und $\sigma^2 > 0$.

Bemerkungen

- Das Modell der additiven ZVA modelliert ausschließlich Haupteffekte, keine Interaktionen.

6. Erläutern Sie die Bedeutung der Parameter μ_0, α_2 und β_2 im additiven Modell der ZVA mit RG.

μ_0 entspricht dem Erwartungswert der Faktorlevelkombination A1B1, α_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A und β_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B

7. Bestimmen Sie μ_{ij} für $\mu_0 := 2, \alpha_2 = -1$ und $\beta_2 := 3$ im additiven Modell der ZVA mit RG.

$$\mu_{11} = \mu_0 + \alpha_1 + \beta_1 = 2 + 0 + 0 = 2 \qquad \mu_{12} = \mu_0 + \alpha_1 + \beta_2 = 2 + 0 + 3 = 5$$
$$\mu_{21} = \mu_0 + \alpha_2 + \beta_1 = 2 + (-1) + 0 = 1 \qquad \mu_{22} = \mu_0 + \alpha_2 + \beta_2 = 2 + (-1) + 3 = 4$$

$\Rightarrow$ Haupteffekt von Faktor A, Haupteffekt von Faktor B

8. Geben Sie die Designmatrixform des Modells einer additiven 2 x 2 ZVA mit RG für $n_{ij} := 1$ an.

$$y \sim N(X\beta, \sigma^2 I_4),$$

wobei

$$y := \begin{pmatrix} y_{111} \\ y_{121} \\ y_{211} \\ y_{221} \end{pmatrix}, \; X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{4\times 3}, \; \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0.$$

9. Geben Sie die Designmatrixform des Modells einer additiven 2 x 2 ZVA mit RG für $n_{ij} := 3$ an.

$$y \sim N(X\beta, \sigma^2 I_{12}),$$

wobei

$$y := \begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{221} \\ y_{222} \\ y_{223} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{12 \times 3}, \ \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \end{pmatrix} \in \mathbb{R}^3 \text{ und } \sigma^2 > 0.$$

10. Geben Sie die Definition des Modells der ZVA mit Interaktion und RG wieder.

Definition (Modell der ZVA mit Interaktion und Referenzgruppe)

y_{ijk} mit $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij}$ sei die Zufallsvariable, die den $k$ten Datenpunkt zum $i$ten Level von Faktor A und dem $j$ten Level von Faktor B in einem ZVA Anwendungsszenario modelliert. Dann hat das *Modell der ZVA mit Interaktion und Referenzgruppe* die strukturelle Form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij} \tag{4}$$

und die Datenverteilungsform

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2) \text{ u.i.v. für } i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij} \tag{5}$$

mit

$$\mu_{ij} := \mu_0 + \alpha_i + \beta_j + \gamma_{ij} \tag{6}$$

sowie

$$\alpha_1 := \beta_1 := \gamma_{i1} := \gamma_{1j} := 0 \text{ für } i = 1, \ldots, I, j = 1, \ldots, J \tag{7}$$

und $\sigma^2 > 0$.

11. Erläutern Sie die Bedeutung der Parameter μ_0, α_2, β_2 und γ_{22} im Modell der ZVA mit Interaktion und RG.

μ_0 entspricht dem Erwartungswert der Faktorlevelkombination A1B1, α_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor A, β_2 der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B und γ_{22} der Differenz beim Übergang von Level 1 zu Level 2 von Faktor B im Unterschiede zum Übergang von Level 1 zu Level 2 von Faktor A.

12. Geben Sie die Designmatrixform des Modells einer 2 x 2 ZVA mit Interaktion und RG für $n_{ij} := 1$ an.

$$y \sim N(X\beta, \sigma^2 I_4), \text{ mit}$$

$$y := \begin{pmatrix} y_{111} \\ y_{121} \\ y_{211} \\ y_{221} \end{pmatrix}, \; X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{4 \times 4}, \; \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0.$$

13. Geben Sie die Designmatrixform des Modells einer 2 x 2 ZVA mit Interaktion und RG für $n_{ij} := 3$ an.

$$y \sim N(X\beta, \sigma^2 I_1 2), \text{ mit}$$

$$y := \begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{221} \\ y_{222} \\ y_{223} \end{pmatrix}, \; X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{12\times 4}, \; \beta := \begin{pmatrix} \mu_0 \\ \alpha_2 \\ \beta_2 \\ \gamma_{22} \end{pmatrix} \in \mathbb{R}^4 \text{ und } \sigma^2 > 0.$$

14. Geben Sie das Theorem zur Betaparameterschätzung im additiven 2 x 2 ZVA Modell mit RG wieder.

Theorem (Betaparameterschätzung im additiven 2 x 2 ZVA Modell mit Referenzgruppe)

Gegeben sei die Designmatrixform eines balancierten additiven 2 x 2 ZVA Modells mit Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4}\bar{y}_{11} + \frac{1}{4}(\bar{y}_{12} + \bar{y}_{21}) - \frac{1}{4}\bar{y}_{22} \\ \frac{1}{2}(\bar{y}_{21} + \bar{y}_{22}) - \frac{1}{2}(\bar{y}_{11} + \bar{y}_{12}) \\ \frac{1}{2}(\bar{y}_{12} + \bar{y}_{22}) - \frac{1}{2}(\bar{y}_{11} + \bar{y}_{21}) \end{pmatrix}, \tag{8}$$

wobei

$$\bar{y}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \text{ für } 1 \le i, j \le 2 \tag{9}$$

das Stichprobenmittel der i, jten Gruppe des 2 x 2 ZVA Designs bezeichnet.

15. Geben Sie das Theorem zur Betaparameterschätzung im 2 x 2 ZVA Modell mit Interaktion und RG wieder.

Theorem (Betaparameterschätzung im 2 x 2 ZVA Modell mit Interaktion und Referenzgruppe)

Gegeben sei die Designmatrixform eines balancierten 2 x 2 ZVA Modells mit Interaktion und Referenzgruppe. Dann ergibt sich für den Betaparameterschätzer

$$\hat{\beta} := \begin{pmatrix} \hat{\mu}_0 \\ \hat{\alpha}_2 \\ \hat{\beta}_2 \\ \hat{\gamma}_{22} \end{pmatrix} = \begin{pmatrix} \bar{y}_{11} \\ \bar{y}_{21} - \bar{y}_{11} \\ \bar{y}_{12} - \bar{y}_{11} \\ \bar{y}_{11} + \bar{y}_{22} - \bar{y}_{12} - \bar{y}_{21} \end{pmatrix}, \tag{10}$$

wobei

$$\bar{y}_{ij} := \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk} \text{ für } 1 \leq i, j \leq 2 \tag{11}$$

das Stichprobenmittel der i, jten Gruppe des 2 x 2 ZVA Designs bezeichnet.

16. Erzeugen Sie einen Beispieldatensatz des Anwendungsbeispiels einer 2 x 2 ZVA mit den Faktoren Therapie (F2F, ONL) und Alter (YA,OA) basierend auf dem Modell der ZVA mit Interaktion und Referenzgruppe für von Ihnen selbst gewählte Parameterwerte $\mu_0, \alpha_2, \beta_2, \gamma_{22}$ und σ^2. Visualisieren Sie den erzeugten Datensatz mithilfe eines Barplots, eines Barplots und eines Lineplots.

Modellformulierung

```
# Modellformulierung
library(MASS)                                  # Multivariate Normalverteilung
I       = 2                                    # Anzahl Level Faktor A
J       = 2                                    # Anzahl Level Faktor B
n_ij    = 40                                   # Anzahl von Datenpunkten der i,jten Gruppe
n       = I*J*n_ij                             # Anzahl Datenpunkte
p       = 1 + (I-1)+(J-1)+(I*J-3)              # Anzahl Parameter
D       = matrix(c(1,0,0,0,                    # Prototypische Designmatrix für balancierte Designs
                   1,0,1,0,
                   1,1,0,0,
                   1,1,1,1),
                 nrow  = p,
                 byrow = TRUE)
C       = matrix(rep(1,n_ij),nrow = n_ij)      # Prototypischer Zellenvektor für balancierte Designs
X       = kronecker(D,C)                       # Kroneckerprodukt Designmatrix Erzeugung für balancierte Designs
I_n     = diag(n)                              # n x n Einheitsmatrix
beta    = matrix(c(1,0,0,-1), nrow = p)         # \beta = (\mu_0,\alpha_2,\alpha_3,\alpha_4)
sigsqr = 12                                    # \sigma^2

# Datenrealisierung
y       = mvrnorm(1, X %*% beta, sigsqr*I_n) # eine Realisierung eines n-dimensionalen ZVs
print(y)
```

```
>   [1]  3.78656  3.71185 -0.23454  1.13337  3.34757  4.64610 -0.93302
>   [8] -2.94219  1.42298  2.48661  6.30957 -4.22768  2.78263 -2.87879
>  [15]  5.57878  4.39467  3.65503  7.50041 -1.86589  0.43647  1.23315
>  [22]  4.10024  0.00617 -1.96251  2.57389 -2.27327 -3.52106  1.28309
>  [29]  3.93127 -2.97055  1.66563 -0.82996  0.09015  3.09097  1.07723
>  [36]  1.37268  2.47767  5.82139  2.55695  8.06458  4.71642  1.86395
>  [43] -4.83452  2.47175  8.36763  2.96969 -0.53044  1.89560 -3.84770
>  [50]  0.27641  4.38947  1.78679 -0.85906 -0.93485 -1.39209 -5.08392
>  [57] -1.75584 -2.66064 -1.72082 -0.34135 -0.02759  5.56162  4.27825
>  [64] -0.22754  6.79504  2.68407  0.86801  4.63285  4.68269  2.46943
>  [71] -0.32312  2.50304  0.46066 -8.04939  4.84701 -0.89922  2.92682
>  [78]  3.59389  5.49776  7.66612  0.68710  1.59645  0.61917  3.10184
>  [85]  2.81488  5.56670  1.81994  5.63261  1.98433 -4.09478 -4.88707
>  [92]  2.69266  1.45612  4.68947 -2.38045 -3.07920  0.14819  4.85884
>  [99] -1.30393  4.72164  1.73909 -0.25131 10.29265 -6.87776  2.90466
> [106]  1.10240  1.06643  0.28609  0.15740  3.98636 -5.62239  1.40944
> [113]  3.00124  1.16119 10.54505  2.54082  0.12884  0.82692 -1.21399
> [120]  1.24261  3.69726 -1.08581 -0.96731 -6.61674 -0.46418  0.38811
> [127]  5.62414  0.84415  0.61137  3.24019 -5.23911 -2.86408 -0.05525
> [134] -3.24187 -0.33756  6.54197  2.55573 -4.51907  1.25730  1.62173
```

Anwendungsbeispiel - Modellformulierung und Datensimulation - SKF 16

Beispieldatensatzerzeugung

```
# Datensimulation
library(MASS)                                         # Multivariate Normalverteilung
set.seed(1)                                           # reproduzierbare Daten
y             = mvrnorm(1, X %*% beta, sigsqr*I_n)    # eine Realisierung eines n-dimensionalen ZVs

# Dataframeformatierung
library(writexl)                                      # Excel Output
D             = data.frame("ID" = 1:n)                # Dataframe Initialisierung und ID Variable
D$Therapy     = c(rep("F2F", J*n_ij), rep("ONL", J*n_ij)) # Therapiebedingung
D$Age         = rep(c(rep("YA", n_ij), rep("OA",n_ij)),I) # Alter
D$BDI         = y                                     # PrePost-BDI Differenzwerte

# Datenspeicherung
write_xlsx(D, file.path(getwd()       , "11_Daten", "SKF_16_Daten.xlsx"))
write.csv( D, file = file.path(getwd(), "11_Daten", "SKF_16_Daten.csv"))
```

Anwendungsbeispiel - Modellformulierung und Datensimulation - SKF 16

Beispieldatensatzerzeugung

	Therapy	Age	BDI
1	F2F	YA	7.475
2	F2F	YA	-3.796
3	F2F	YA	-1.152
4	F2F	YA	4.464
5	F2F	YA	-2.725
6	F2F	YA	-4.153
41	F2F	OA	0.386
42	F2F	OA	2.712
43	F2F	OA	0.033
44	F2F	OA	-0.108
45	F2F	OA	-0.361
81	ONL	YA	-1.042
82	ONL	YA	1.258
83	ONL	YA	1.004
84	ONL	YA	-0.536
85	ONL	YA	2.010
121	ONL	OA	2.644
122	ONL	OA	3.811
123	ONL	OA	-0.205
124	ONL	OA	-1.366
125	ONL	OA	-1.438