

12.9 Das allgemeine lineare Modell

Das allgemeine lineare Modell (ALM) liefert einen formalen Rahmen, mit dessen Hilfe sich lineare Regression (Abschn. 6.3, 12.5), Varianzanalyse (Abschn. 7.3, 12.7) und Kovarianzanalyse (Abschn. 7.8) auf dieselbe Weise formulieren und ihre Hypothesen testen lassen. Es integriert dabei die uni- wie multivariaten Varianten dieser Verfahren. R verwendet das ALM implizit beim Anpassen dieser Modelle, was bisweilen für den Anwender sichtbar wird. Daher soll die Grundidee des ALM hier skizziert werden. Weiterführende Darstellungen finden sich bei Andres (1996) und Mardia et al. (1980). Für die verwendeten Konzepte der linearen Algebra s. Abschn. 12.1.

12.9.1 Modell der multiplen linearen Regression

Das Modell der univariaten multiplen Regression geht von Beobachtungsobjekten $i = 1, \dots, n$ aus, die Daten y_i eines Kriteriums Y und Werte x_{ij} von p Prädiktoren X_j liefern, die jeweils in Vektoren \mathbf{y} und \mathbf{x}_j zusammengefasst werden (Abschn. 6.3).

$$\begin{aligned} E(y_i) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip} \\ E(\mathbf{y}) &= \beta_0 + \beta_1 \mathbf{x}_1 + \dots + \beta_j \mathbf{x}_j + \dots + \beta_p \mathbf{x}_p \end{aligned}$$

Dabei ist $E(\mathbf{y})$ der n -Vektor $(E(y_1), \dots, E(y_n))^\top$ der Erwartungswerte von y_i . Von den skalaren Parametern β_0 (additive Konstante, absoluter Term) und β_j (theoretische Regressionsgewichte) wird angenommen, dass sie für alle Beobachtungsobjekte identisch sind. Ferner sei vorausgesetzt, dass mehr Beobachtungen als Parameter vorhanden sind, hier also $n > p + 1$ gilt. In Matrix-Schreibweise lässt sich das Modell so formulieren:

$$E(\mathbf{y}) = \mathbf{1}\beta_0 + \mathbf{X}_p\boldsymbol{\beta}_p = [\mathbf{1}|\mathbf{X}_p] \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_p \end{pmatrix} = \mathbf{X}\boldsymbol{\beta}$$

Hier ist $\mathbf{1}$ der n -Vektor $(1, \dots, 1)^\top$, $\boldsymbol{\beta}_p$ der p -Vektor $(\beta_1, \dots, \beta_p)^\top$, $\boldsymbol{\beta}$ der $(p+1)$ -Vektor $(\beta_0, \boldsymbol{\beta}_p^\top)^\top$ und \mathbf{X}_p die $(n \times p)$ -Matrix der spaltenweise zusammengestellten Vektoren \mathbf{x}_j . Die Prädiktoren sollen linear unabhängig sein, womit \mathbf{X}_p vollen Spaltenrang p besitzt. $\mathbf{X} = [\mathbf{1}|\mathbf{X}_p]$ ist die $(n \times (p+1))$ -Designmatrix, deren Spalten den Unterraum V mit Dimension $\text{Rang}(\mathbf{X}) = p + 1$ aufspannen.⁴⁵ Das Modell lässt sich als Behauptung verstehen, dass $E(\mathbf{y})$ in V liegt. Für einen solchen modellverträglichen Vektor von Erwartungswerten existiert ein Parametervektor $\boldsymbol{\beta}$, mit dem $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ gilt.⁴⁶

⁴⁵Die Designmatrix erhält man mit der Funktion `model.matrix()`, die als Argument ein mit `lm()` erstelltes Modell, oder auch nur die rechte Seite einer Modellformel akzeptiert (s. Abschn. 5.2 und Venables & Ripley, 2002, p. 144 ff.).

⁴⁶Die \mathbf{x}_j sind feste Realisierungen eines Zufallsvektors, also *stochastische Prädiktoren*. Sie enthalten damit nicht alle möglichen Prädiktorwerte, sondern nur jeweils n viele. Man könnte daher auch vom Vektor $E(\mathbf{y}|\mathbf{X})$ der auf eine konkrete Designmatrix \mathbf{X} bedingten Erwartungswerte von \mathbf{y} sprechen, worauf hier aber verzichtet wird. Die \mathbf{x}_j müssen fehlerfrei sein, bei den x_{ij} muss es sich also um die wahren Prädiktorwerte handeln. Ohne diese Annahme kommen lineare Strukturgleichungsmodelle zur Auswertung in Betracht (Abschn. 12.3, Fußnote 32).

Im ALM ergeben sich die personenweisen Erwartungswerte eines Kriteriums als lineare Funktion der Prädiktoren. Der Zusammenhang zwischen einer AV Y und einer UV X selbst muss im ALM dagegen nicht linear sein. Ein quadratischer Zusammenhang zwischen Y und X könnte etwa als Modell $E(\mathbf{y}) = \beta_0 + \beta_1 \mathbf{x} + \beta_2 \mathbf{x}^2$ formuliert werden – hier gäbe es bei einer UV X zwei Prädiktoren, nämlich X und X^2 .

Im Fall der multivariaten multiplen linearen Regression mit r Kriterien Y_l stellt man die Vektoren $E(\mathbf{y}_l)$ der Erwartungswerte der einzelnen Kriterien spaltenweise zu einer $(n \times r)$ -Matrix $E(\mathbf{Y})$ zusammen. Genauso werden die r vielen p -Vektoren $\beta_{p,l} = (\beta_{p,1l}, \dots, \beta_{p,jl}, \dots, \beta_{p,pl})^\top$ der theoretischen Regressionsgewichte jeweils aus der Regression mit den Prädiktoren X_j und dem Kriterium Y_l spaltenweise zu einer $(p \times r)$ -Matrix \mathbf{B}_p zusammengefasst. Die r absoluten Terme $\beta_{0,l}$ bilden einen r -Vektor β_0 . Das Modell lautet dann:

$$E(\mathbf{Y}) = \beta_0 + \mathbf{X}_p \mathbf{B}_p = [\mathbf{1} | \mathbf{X}_p] \begin{pmatrix} \beta_0^\top \\ \mathbf{B}_p \end{pmatrix} = \mathbf{X} \mathbf{B}$$

Die Designmatrix im multivariaten Fall stimmt mit jener im univariaten Fall überein. Die Modellparameter sind im univariaten Fall bei Kenntnis von \mathbf{X} und $E(\mathbf{y})$ als $\beta = \mathbf{X}^+ E(\mathbf{y})$ identifizierbar – der Lösung der *Normalgleichungen* (Abschn. 12.1.7). Dies gilt analog auch im multivariaten Fall mit $\mathbf{B} = \mathbf{X}^+ E(\mathbf{Y})$. Dabei ist $\mathbf{X}^+ = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ die Pseudoinverse von \mathbf{X} , also $\mathbf{X}^+ \mathbf{X} = \mathbf{I}$ mit der $(n \times n)$ -Einheitsmatrix \mathbf{I} . Die Parameter sind gleich den Koordinaten des Vektors der Erwartungswerte, der orthogonal auf V projiziert wurde, bzgl. der durch \mathbf{X} definierten Basis (Abschn. 12.1.7).

Für eine univariate multiple Regression mit zwei Prädiktoren X_j und drei Beobachtungsobjekten, die Prädiktorwerte x_{ij} besitzen, ergibt sich folgendes Modell:

$$\begin{aligned} E(\mathbf{y}) &= \begin{pmatrix} E(y_1) \\ E(y_2) \\ E(y_3) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}_{\mathbf{1}} \beta_0 + \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}_{\mathbf{X}_p} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}_{\beta_p} \\ &= \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \end{pmatrix}_{\mathbf{X}} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}_{\beta} \\ &= \begin{pmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} \\ \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} \end{pmatrix} \end{aligned}$$

Im Fall einer moderierten Regression wird das Modell um zusätzliche Prädiktoren erweitert, die als Interaktionsterm gleich dem Produkt von Einzelprädiktoren sind (Abschn. 6.3.4). Hierfür kann man aus Produkten der Spalten von \mathbf{X}_p eine weitere Matrix \mathbf{X}_{pI} mit den neuen Prädiktoren erstellen, so dass die neue Designmatrix $\mathbf{X} = [\mathbf{1} | \mathbf{X}_p | \mathbf{X}_{pI}]$ ist. Der Parametervektor β ist entsprechend um passend viele Interaktionsparameter zu ergänzen.

12.9.2 Modell der einfaktoriellen Varianzanalyse

Analog zur Regression wird in der univariaten einfaktoriellen Varianzanalyse mit p unabhängigen Bedingungen (CR- p Design, Abschn. 7.3) zunächst die Zugehörigkeit zu jeder der p Faktorstufen zu einem eigenen dichotomen Prädiktor X_j^* , der als *Indikatorvariable* bezeichnet wird. Beobachtungsobjekte erhalten für ein X_j^* den Wert 1, wenn sie sich in der zugehörigen Bedingung j befinden, sonst 0.⁴⁷ Jedes Beobachtungsobjekt erhält also auf einem X_j^* die 1 und auf den verbleibenden $p - 1$ Indikatorvariablen die 0. Die spaltenweise aus den X_j^* zusammengestellte $(n \times p)$ -Matrix \mathbf{X}_p^* heißt *Inzidenzmatrix* und hat vollen Spaltenrang p . Die Komponenten des Vektors \mathbf{y} und die Zeilen der Inzidenzmatrix \mathbf{X}_p^* seien gemeinsam entsprechend der Gruppenzugehörigkeit geordnet.⁴⁸

Für die einfaktorielle Varianzanalyse mit drei Gruppen, zwei Personen pro Gruppe sowie den Parametern β_0 und β_j^* ergibt sich folgendes Modell. Beobachtungen aus unterschiedlichen Gruppen sind durch waagerechte Linien getrennt.

$$\begin{aligned}
 E(\mathbf{y}) &= \begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{pmatrix} \beta_p^* \\
 &= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{pmatrix} \beta_{p+1}^* = \begin{pmatrix} \beta_0 + \beta_1^* \\ \beta_0 + \beta_1^* \\ \beta_0 + \beta_2^* \\ \beta_0 + \beta_2^* \\ \beta_0 + \beta_3^* \\ \beta_0 + \beta_3^* \end{pmatrix}
 \end{aligned}$$

Der Vektor $\mathbf{1}$ ist die Summe der Spalten von \mathbf{X}_p^* , liegt also in ihrem Erzeugnis, weshalb die $(n \times (p + 1))$ -Designmatrix $\mathbf{X}^* = [\mathbf{1} | \mathbf{X}_p^*]$ wie \mathbf{X}_p^* selbst nur Rang p besitzt. $(\mathbf{X}^*)^+$ ist nun nicht eindeutig bestimmt, da $(\mathbf{X}^*)^\top \mathbf{X}^*$ nicht invertierbar ist. Die Parameter sind hier deswegen anders als in der Regression zunächst nicht identifizierbar. Aus diesem Grund verändert man das Modell, indem man eine Nebenbedingung $\mathbf{v}^\top \beta_p^* = 0$ über die Parameter einführt und so deren frei variierbare Anzahl um 1 reduziert. Dabei ist \mathbf{v} ein p -Vektor, der eine Linearkombination der Parameter β_j^* festlegt, die 0 ist.⁴⁹

⁴⁷ Anders als in der Regression werden die Werte der Indikatorvariablen hier durch die Zuordnung von Beobachtungen zu Gruppen systematisch hergestellt, sind also keine stochastischen Prädiktoren.

⁴⁸ Eine der folgenden Zusammenfassungen ähnlicher Exposition enthält die Vignette des Pakets `codingMatrices` (Venables, 2016). Dabei entsprechen sich folgende Bezeichnungen: Die Inzidenzmatrix \mathbf{X}_p^* hier ist dort \mathbf{X} . Die reduzierte Designmatrix \mathbf{X} hier ist dort $\tilde{\mathbf{X}}$, die Codiermatrix \mathbf{C} hier ist dort \mathbf{B}_* , die Matrix $[\mathbf{1} | \mathbf{C}]$ hier ist dort $\mathbf{B} = [\mathbf{1} | \mathbf{B}_*] = \mathbf{C}^{-1}$, der Parametervektor β_{p-1}^* hier ist dort β_* .

⁴⁹ Alternativ kann auch der Parameter $\beta_0 = 0$ gesetzt werden, womit $\mathbf{X}^* = \mathbf{X}_p^*$ ist. Diese Möglichkeit zur Parametrisierung soll hier nicht weiter verfolgt werden, um das Modell wie jenes der Regression formulieren zu können (Fußnote 53).

Es sei nun $\beta_{p-1} = (\beta_1, \dots, \beta_{p-1})^\top$ der $(p-1)$ -Vektor der neuen, *reduzierten* Parameter, den *Kontrasten*. Weiter sei \mathbf{C} eine $(p \times (p-1))$ -Matrix, so dass die $(p \times p)$ -Matrix $[\mathbf{1}|\mathbf{C}]$ vollen Rang p besitzt und unter Erfüllung der Nebenbedingung $\mathbf{v}^\top \beta_p^* = 0$ gilt:⁵⁰

$$\begin{aligned}\beta_p^* &= \mathbf{C}\beta_{p-1} \\ \beta^* &= [\mathbf{1}|\mathbf{C}] \begin{pmatrix} \beta_0 \\ \beta_{p-1} \end{pmatrix} = [\mathbf{1}|\mathbf{C}] \beta\end{aligned}$$

Die *Codiermatrix* \mathbf{C} definiert für jede in den p Zeilen stehende Gruppenzugehörigkeit die zugehörigen Werte auf den neuen, nun $p-1$ Prädiktoren X_j , die mit den Spalten von \mathbf{C} korrespondieren. Die Nebenbedingung $\mathbf{v}^\top \beta_p^* = 0$ ist erfüllt, wenn \mathbf{C} so gewählt wird, dass $\mathbf{v}^\top \mathbf{C} = \mathbf{0}^\top$ ist.⁵¹ Dabei können für ein bestimmtes \mathbf{v} mehrere Wahlmöglichkeiten für \mathbf{C} bestehen. Da $[\mathbf{1}|\mathbf{C}]$ vollen Rang p besitzt, existiert $[\mathbf{1}|\mathbf{C}]^{-1}$, und es gilt:

$$\beta = [\mathbf{1}|\mathbf{C}]^{-1} \beta^*$$

Die *Kontrastmatrix* $[\mathbf{1}|\mathbf{C}]^{-1}$ bildet den Vektor β^* aller ursprünglichen Parameter auf den Vektor aller neuen Parameter β im Unterraum ab, in dem β^* mit der gewählten Nebenbedingung frei variieren kann. Anders gesagt ergeben sich die Komponenten von β als Linearkombinationen der ursprünglichen Parameter β^* , wobei $[\mathbf{1}|\mathbf{C}]^{-1}$ die zeilenweise aus den Koeffizientenvektoren zusammengestellte Matrix mit Rang p ist. $[\mathbf{1}|\mathbf{C}]^{-1}$ induziert so die inhaltliche Bedeutung der neuen Parameter. Das reduzierte Modell lautet insgesamt:

$$E(\mathbf{y}) = \mathbf{1}\beta_0 + \mathbf{X}_p^* \beta_p^* = \mathbf{1}\beta_0 + \mathbf{X}_p^* \mathbf{C} \beta_{p-1} = [\mathbf{1}|\mathbf{X}_{p-1}] \begin{pmatrix} \beta_0 \\ \beta_{p-1} \end{pmatrix} = \mathbf{X} \beta$$

Hier ist $\mathbf{X}_{p-1} = \mathbf{X}_p^* \mathbf{C}$ die reduzierte $(n \times (p-1))$ -Inzidenzmatrix und $\mathbf{X} = [\mathbf{1}|\mathbf{X}_{p-1}]$ die reduzierte $(n \times p)$ -Designmatrix mit vollem Spaltenrang p , wobei $n > p$ vorausgesetzt wird. Der p -Vektor β der neuen Parameter ist so wie im Fall der Regression über $\mathbf{X}^+ E(\mathbf{y})$ identifizierbar.

Die inhaltliche Bedeutung der Parameter in β hängt von der Wahl der Nebenbedingung \mathbf{v} und der Codiermatrix \mathbf{C} ab. R verwendet in der Voreinstellung die *Dummy-Codierung* (*Treatment-Kontraste*), bei denen die Indikatorvariablen X_j^* zunächst erhalten bleiben. In der Voreinstellung wird dann der ursprüngliche, zur ersten Gruppe gehörende Parameter $\beta_1^* = 0$ gesetzt.⁵² Hier ist die Nebenbedingung also $\beta_1^* = (1, 0, \dots, 0) \beta_p^* = 0$, d. h. $\mathbf{v} = (1, 0, \dots, 0)^\top$. `contr.treatment(<Anzahl>, base=<Referenzgruppe>)` gibt die Matrix \mathbf{C}_t für Treatment-Kontraste aus. Als Argument ist dabei die Anzahl der Gruppen p sowie optional die Nummer der Referenzstufe zu nennen – in der Voreinstellung die Stufe 1. Die Spalten von \mathbf{C}_t sind paarweise orthogonal und besitzen die Länge 1 ($\mathbf{C}_t^\top \mathbf{C}_t = \mathbf{I}$).

⁵⁰Für die Parameterschätzungen gilt dies analog (Abschn. 12.9.4, Fußnote 59).

⁵¹Denn dann gilt $\mathbf{v}^\top \beta_p^* = \mathbf{v}^\top \mathbf{C} \beta_{p-1} = \mathbf{0}^\top \beta_{p-1} = 0$. \mathbf{v} steht senkrecht auf den Spalten von \mathbf{C} , ist also eine Basis des orthogonalen Komplements des von den Spalten von \mathbf{C} aufgespannten Unterraums. Mit anderen Worten ist \mathbf{v} wegen $\mathbf{0} = (\mathbf{v}^\top \mathbf{C})^\top = \mathbf{C}^\top \mathbf{v}$ eine Basis des Kerns von \mathbf{C}^\top .

⁵²Die in einem linearen Modell mit kategorialen Variablen von R weggelassene Gruppe ist die erste Stufe von `levels(<Faktor>)`.

```
> contr.treatment(4)           # Treatment-Kontraste für 4 Gruppen
  2 3 4
1 0 0 0
2 1 0 0
3 0 1 0
4 0 0 1
```

Treatment-Kontraste bewirken, dass die reduzierte Inzidenzmatrix \mathbf{X}_{p-1} durch Streichen der ersten Spalte von \mathbf{X}_p^* entsteht. Die Bezeichnung dieses Codierschemas leitet sich aus der Situation ab, dass die erste Faktorstufe eine Kontrollgruppe darstellt, während die übrigen zu Treatment-Gruppen gehören. Die Parameter können dann als Wirkung der Stufe j i. S. der Differenz zur Kontrollgruppe verstanden werden: Für die Mitglieder der ersten Gruppe ist $E(y_i) = \mu_1 = \beta_0$, da für diese Gruppe alle verbleibenden $X_j = 0$ sind (mit $j = 2, \dots, p$). Für Mitglieder jeder übrigen Gruppe j erhält man $E(y_i) = \mu_j = \beta_0 + \beta_j = \mu_1 + \beta_j$, da dann $X_j = 1$ ist. Damit ist $\beta_j = \mu_j - \mu_1$ gleich der Differenz des Erwartungswertes der Gruppe j zum Erwartungswert der Referenzgruppe.⁵³

Für $p = 3$, zwei Personen pro Gruppe und Treatment-Kontraste ergibt sich folgendes Modell mit den reduzierten Parametern β_0 und β_j :

$$\begin{aligned}
 E(\mathbf{y}) &= \begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}_{\mathbf{1}} \beta_0 + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}_{\mathbf{X}_p^*} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}_{\mathbf{C}_t} \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix}_{\beta_{p-1}} \\
 &= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}_{\mathbf{X}} \begin{pmatrix} \beta_0 \\ \beta_2 \\ \beta_3 \end{pmatrix}_{\beta} = \begin{pmatrix} \beta_0 \\ \beta_0 \\ \beta_0 + \beta_2 \\ \beta_0 + \beta_2 \\ \beta_0 + \beta_3 \\ \beta_0 + \beta_3 \end{pmatrix}
 \end{aligned}$$

Die von \mathbf{C}_t induzierte Bedeutung der Parameter in β lässt sich direkt an $[\mathbf{1}|\mathbf{C}_t]^{-1}$ ablesen. Die erste Zeile definiert das Verhältnis von β_0 zu den Gruppenerwartungswerten μ_j – hier $\beta_0 = \mu_1$. Die weiteren Zeilen definieren, wie sich die Kontraste in β_{p-1} als Differenz jedes Gruppenerwartungswerts zum Erwartungswert der Referenzstufe ergeben – etwa $\beta_1 = -\mu_1 + \mu_2$.

```
> solve(cbind(1, contr.treatment(4)))
  1 2 3 4
  1 0 0 0
```

⁵³Die in Fußnote 49 erwähnte Möglichkeit der Parametrisierung führt zum *cell means* Modell, bei dem $\mathbf{X} = \mathbf{X}^*$ gilt und die Parameter β_j^* direkt die Bedeutung der Gruppenerwartungswerte μ_j erhalten.

```
2 -1 1 0 0
3 -1 0 1 0
4 -1 0 0 1
```

In der mit `getOption("contrasts")` einsehbaren Voreinstellung verwendet R Treatment-Kontraste, wenn ungeordnete kategoriale Prädiktoren in einem linearen Modell vorhanden sind, und polynomiale Kontraste für ordinale Prädiktoren (für andere Möglichkeiten vgl. `?contrasts`). Deren Stufen werden dabei als gleichabständig vorausgesetzt.

```
> getOption("contrasts")           # eingestelltes Codierschema
      unordered      ordered
"contr.treatment" "contr.poly"
```

In der einfaktoriellen Varianzanalyse ist die Konzeption der Parameter jedoch oft eine andere. Hier sind dies die Effektgrößen $\alpha_j = \mu_j - \mu$, also die jeweilige Differenz der Gruppenerwartungswerte zum (ungewichteten, also gleichgewichteten) mittleren Erwartungswert $\mu = \frac{1}{p} \sum_j \mu_j$. Die Summe der so definierten Effektgrößen α_j in der Rolle der Parameter β_j^* ist 0. Die Nebenbedingung lautet hier also $\sum_j \beta_j^* = \mathbf{1}^\top \beta_p^* = 0$, d. h. $\mathbf{v} = \mathbf{1}$.

Die genannte Parametrisierung lässt sich über die ungewichtete *Effektcodierung* ausdrücken:⁵⁴ Hierfür wird zunächst die Zugehörigkeit zu jeder Faktorstufe zu einem separaten Prädiktor X_j , der die Werte $-1, 0$ und 1 annehmen kann. Beobachtungsobjekte aus der Gruppe j (mit $j = 1, \dots, p-1$) erhalten für X_j den Wert 1 , sonst 0 . Beobachtungsobjekte aus der Gruppe p erhalten auf allen X_j den Wert -1 . Zur Beseitigung der Redundanz wird dann der ursprüngliche, zur letzten Stufe gehörende Parameter β_p^* aus dem Modell gestrichen.⁵⁵ `contr.sum(⟨Anzahl ↘ Gruppen⟩)` gibt die Matrix C_e für die Effektcodierung aus, wobei als Argument die Anzahl der Gruppen p zu nennen ist.

```
> contr.sum(4)           # Effektcodierung für 4-stufigen Faktor
  [,1] [,2] [,3]
1     1     0     0
2     0     1     0
3     0     0     1
4    -1    -1    -1
```

Mit der ungewichteten Effektcodierung erhält der Parameter β_0 die Bedeutung des ungewichteten mittleren Erwartungswertes $\mu = \frac{1}{p} \sum_j \mu_j$ und die β_j die Bedeutung der ersten $p-1$ Effektgrößen α_j . Für Mitglieder der ersten $p-1$ Gruppen ist nämlich $E(y_i) = \mu_j = \beta_0 + \beta_j$, für Mitglieder der Gruppe p gilt $E(y_i) = \mu_p = \beta_0 - (\beta_1 + \dots + \beta_{p-1})$. Dabei stellt $\beta_1 + \dots + \beta_{p-1}$ die Abweichung $\alpha_p = \mu_p - \mu$ dar, weil sich die Abweichungen der Gruppenerwartungswerte vom mittleren Erwartungswert über alle Gruppen zu 0 summieren müssen. In jeder Komponente ist somit $E(y_i) = \mu + \alpha_j$.

⁵⁴Ein andere Wahl für C unter der Nebenbedingung $\mathbf{v} = \mathbf{1}$ ist die Helmert-Codierung mit paarweise orthogonalen Spalten von $[\mathbf{1}|C]$ (vgl. `?contr.helmert`). Die Parameter haben dann jedoch eine andere Bedeutung.

⁵⁵Alternativ ließe sich $\mu = \sum_j \frac{n_j}{n} \mu_j$ als mit den anteiligen Zellbesetzungen $\frac{n_j}{n}$ gewichtetes Mittel der μ_j definieren. Die zugehörige Nebenbedingung für die β_j^* i. S. der α_j lautet dann $\sum_j \frac{n_j}{n} \beta_j^* = 0$, d. h. $\mathbf{v} = (\frac{n_1}{n}, \dots, \frac{n_p}{n})^\top$. Diese Parametrisierung lässt sich mit der gewichteten Effektcodierung umsetzen, die zunächst der ungewichteten gleicht. In der letzten Zeile der Matrix C erhalten hier jedoch Mitglieder der Gruppe p für die X_j nicht den Wert -1 , sondern $-\frac{n_j}{n_p}$.

Für $p = 3$, zwei Personen pro Gruppe und Effektcodierung ergibt sich folgendes Modell mit den reduzierten Parametern β_0 und β_j :

$$\begin{aligned}
 E(\mathbf{y}) &= \begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \\
 &= \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 \\ \beta_0 + \beta_1 \\ \beta_0 + \beta_2 \\ \beta_0 + \beta_2 \\ \beta_0 - (\beta_1 + \beta_2) \\ \beta_0 - (\beta_1 + \beta_2) \end{pmatrix}
 \end{aligned}$$

Die von \mathbf{C}_e induzierte Bedeutung der Parameter in $\boldsymbol{\beta}$ lässt sich direkt an $[\mathbf{1}|\mathbf{C}_e]^{-1}$ ablesen. Die erste Zeile definiert das Verhältnis von β_0 zu den Gruppenerwartungswerten μ_j – hier ist dies das arithmetische Mittel $\beta_0 = \frac{1}{4}\mu_1 + \frac{1}{4}\mu_2 + \frac{1}{4}\mu_3 + \frac{1}{4}\mu_4$. Die weiteren Zeilen definieren, wie sich die Kontraste in $\boldsymbol{\beta}_{p-1}$ als Vergleich zwischen Gruppenerwartungswert und mittlerem Erwartungswert ergeben, etwa:

$$\begin{aligned}
 \beta_1 &= \frac{3}{4}\mu_1 - \frac{1}{4}\mu_2 - \frac{1}{4}\mu_3 - \frac{1}{4}\mu_4 \\
 &= \frac{3}{4}\mu_1 + \frac{1}{4}\mu_1 - \frac{1}{4}\mu_1 - \frac{1}{4}\mu_2 - \frac{1}{4}\mu_3 - \frac{1}{4}\mu_4 \\
 &= \mu_1 - \frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4)
 \end{aligned}$$

```

> solve(cbind(1, contr.sum(4)))
      1      2      3      4
[1,] 0.25 0.25 0.25 0.25
[2,] 0.75 -0.25 -0.25 -0.25
[3,] -0.25 0.75 -0.25 -0.25
[4,] -0.25 -0.25 0.75 -0.25

```

Soll R bei ungeordneten kategorialen Prädiktoren in diesem Sinne jede Gruppe nicht wie bei Treatment-Kontrasten mit der Referenzstufe, sondern durch die Effektcodierung generell mit dem Gesamtmittel verglichen, ist dies mit `options()` einzustellen.

```

# gehe zur Effektcodierung für ungeordnete Faktoren über
> options(contrasts=c("contr.sum", "contr.poly"))

# wechsele zurück zu Treatment-Kontrasten
> options(contrasts=c("contr.treatment", "contr.poly"))

```

Neben dieser Grundeinstellung existiert auch die Möglichkeit, das Codierschema für einen Faktor mit `C()` direkt festzulegen.

```
> C(object=<Faktor>, contr=<Codiermatrix>)
```

Ihr erstes Argument ist ein Faktor. Für das Argument `contr` ist entweder eine selbst erstellte Codiermatrix C oder eine Funktion wie etwa `contr.sum` zu übergeben, die eine passende Codiermatrix erzeugt. C wird als Attribut des von `C()` zurückgegebenen Faktors gespeichert und automatisch von Funktionen wie `lm()` oder `aov()` verwendet, wenn der neue Faktor Teil der Modellformel ist.

```
> IV <- gl(3, 5) # Faktor: 3 Gruppen à 5 Personen
> (IVe <- C(IV, contr.sum)) # Effektcodierung für Faktor IVE
[1] 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3
attr(,"contrasts")
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
Levels: 1 2 3
```

`contrasts(<Faktor>)` gibt die passende Codiermatrix C für die Stufen des übergebenen Faktors aus. Ist mit dem Faktor keine Codiermatrix in Form eines Attributs fest verbunden, wird hierfür die Grundeinstellung verwendet, die mit `options("contrasts")` einsehbar ist.

```
> contrasts(IV) # Codiermatrix C nach Grundeinstellung
  2 3
1 0 0
2 1 0
3 0 1
```

Schließlich lässt sich das Codierschema temporär auch direkt beim Aufruf der `lm()` Funktion angeben, indem ihr Argument `contrasts` verwendet wird. Das Argument erwartet eine Liste, deren Komponenten als Namen die Bezeichnungen der Faktoren besitzen, die in der Modellformel auftauchen. Die Komponente selbst ist dann eine Kontrastfunktion wie etwa `contr.sum`.

```
> DV <- rnorm(15)
> lm(DV ~ IV, contrasts=list(IV=contr.treatment)) # ...
```

12.9.3 Modell der zweifaktoriellen Varianzanalyse

In der univariaten zweifaktoriellen Varianzanalyse mit p Stufen der ersten und q Stufen der zweiten UV (CRF- pq Design, Abschn. 7.5) ist zunächst für jeden der beiden Haupteffekte eine Inzidenzmatrix analog zum einfaktoriellen Fall zu bilden. Dies sollen hier die $(n \times p)$ -Matrix \mathbf{X}_1^* für die erste UV und die $(n \times q)$ -Matrix \mathbf{X}_2^* für die zweite UV sein. Der zugehörige p -Vektor der Parameter für die erste UV sei $\beta_1^* = (\beta_{1,1}^*, \dots, \beta_{1,p}^*)^\top$, der q -Vektor der Parameter für die zweite UV sei $\beta_2^* = (\beta_{2,1}^*, \dots, \beta_{2,q}^*)^\top$.

Die $(n \times (p \cdot q))$ -Inzidenzmatrix $\mathbf{X}_{1 \times 2}^*$ für den Interaktionseffekt wird als spaltenweise Zusammenstellung aller $p \cdot q$ möglichen Produkte der Spalten von \mathbf{X}_1^* und \mathbf{X}_2^* gebildet (Abschn. 6.3.4). Der zugehörige $(p \cdot q)$ -Vektor der passend geordneten Parameter sei $\beta_{1 \times 2}^* = (\beta_{1 \times 2.11}^*, \dots, \beta_{1 \times 2.pq}^*)^\top$. Die $(n \times (p + q + p \cdot q))$ -Gesamt-Inzidenzmatrix ist dann $[\mathbf{X}_1^* | \mathbf{X}_2^* | \mathbf{X}_{1 \times 2}^*]$ und der zugehörige $(p + q + p \cdot q)$ -Vektor aller genannten Parameter $(\beta_1^{*\top}, \beta_2^{*\top}, \beta_{1 \times 2}^{*\top})^\top$. Entsprechend ist die ursprüngliche $(n \times (p + q + p \cdot q + 1))$ -Designmatrix $\mathbf{X}^* = [\mathbf{1} | \mathbf{X}_1^* | \mathbf{X}_2^* | \mathbf{X}_{1 \times 2}^*]$ und das Modell analog zur einfaktoriellen Situation:

$$E(\mathbf{y}) = [\mathbf{1} | \mathbf{X}_1^* | \mathbf{X}_2^* | \mathbf{X}_{1 \times 2}^*] \begin{pmatrix} \beta_0 \\ \beta_1^* \\ \beta_2^* \\ \beta_{1 \times 2}^* \end{pmatrix} = \mathbf{X}^* \beta^*$$

Für die zweifaktorielle Varianzanalyse mit $p = 3$, $q = 2$ und einer Person pro Gruppe ergibt sich folgendes Modell:

$$E(\mathbf{y}) = \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{31} \\ \mu_{12} \\ \mu_{22} \\ \mu_{32} \end{pmatrix} = \begin{pmatrix} 1 & | & 1 & 0 & 0 & | & 1 & 0 & | & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & | & 0 & 1 & 0 & | & 1 & 0 & | & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & | & 0 & 0 & 1 & | & 1 & 0 & | & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & | & 1 & 0 & 0 & | & 0 & 1 & | & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & | & 0 & 1 & 0 & | & 0 & 1 & | & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & | & 0 & 0 & 1 & | & 0 & 1 & | & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_{1.1}^* \\ \beta_{1.2}^* \\ \beta_{1.3}^* \\ \beta_{2.1}^* \\ \beta_{2.2}^* \\ \beta_{1 \times 2.11}^* \\ \beta_{1 \times 2.21}^* \\ \beta_{1 \times 2.31}^* \\ \beta_{1 \times 2.12}^* \\ \beta_{1 \times 2.22}^* \\ \beta_{1 \times 2.32}^* \end{pmatrix} \beta^*$$

$$= \begin{pmatrix} \beta_0 + \beta_{1.1}^* + \beta_{2.1}^* + \beta_{1 \times 2.11}^* \\ \beta_0 + \beta_{1.2}^* + \beta_{2.1}^* + \beta_{1 \times 2.21}^* \\ \beta_0 + \beta_{1.3}^* + \beta_{2.1}^* + \beta_{1 \times 2.31}^* \\ \beta_0 + \beta_{1.1}^* + \beta_{2.2}^* + \beta_{1 \times 2.12}^* \\ \beta_0 + \beta_{1.2}^* + \beta_{2.2}^* + \beta_{1 \times 2.22}^* \\ \beta_0 + \beta_{1.3}^* + \beta_{2.2}^* + \beta_{1 \times 2.32}^* \end{pmatrix}$$

Wie in der einfaktoriellen Varianzanalyse besitzt \mathbf{X}^* nicht vollen Spaltenrang (sondern Rang $(p - 1) + (q - 1) + (p - 1) \cdot (q - 1) + 1 = p \cdot q$), weshalb die Parameter nicht identifizierbar sind. Das Modell muss deshalb durch separate Nebenbedingungen an die Parameter der Haupteffekte und Interaktion wieder in eines mit weniger Parametern überführt werden. Zunächst seien dafür $\mathbf{v}_1^\top \beta_1^* = 0$ und $\mathbf{v}_2^\top \beta_2^* = 0$ die Nebenbedingungen für die Parameter der Haupteffekte, deren frei variierbare Anzahl sich dadurch auf $p - 1$ bzw. $q - 1$ reduziert. Weiter bezeichne $\mathbf{B}_{1 \times 2}^*$ die in einer $(p \times q)$ -Matrix angeordneten ursprünglichen Parameter $\beta_{1 \times 2}^*$ der Interaktion.

Die Nebenbedingung für diese Parameter lässt sich dann als $\mathbf{v}_1^\top \mathbf{B}_{1 \times 2}^* = \mathbf{0}^\top$ und $\mathbf{B}_{1 \times 2}^* \mathbf{v}_2 = \mathbf{0}$ schreiben, wodurch deren frei variierbare Anzahl nunmehr $(p-1) \cdot (q-1)$ beträgt.

Analog zum einfaktoriellen Fall ist nun zunächst die $(p \times (p-1))$ -Codiermatrix \mathbf{C}_1 für die $p-1$ Parameter der ersten UV im Vektor β_1 nach demselben Schema zu bilden wie die $(q \times (q-1))$ -Codiermatrix \mathbf{C}_2 für die $q-1$ Parameter der zweiten UV im Vektor β_2 . Der Rang von $[\mathbf{1}|\mathbf{C}_1]$ sei also p , der Rang von $[\mathbf{1}|\mathbf{C}_2]$ sei q , und es gelte $\mathbf{v}_1^\top \mathbf{C}_1 = \mathbf{0}$ ebenso wie $\mathbf{v}_2^\top \mathbf{C}_2 = \mathbf{0}$. Die $((p \cdot q) \times (p-1) \cdot (q-1))$ -Codiermatrix $\mathbf{C}_{1 \times 2}$ für die $(p-1) \cdot (q-1)$ Parameter der Interaktion im Vektor $\beta_{1 \times 2}$ erhält man als Kronecker-Produkt \otimes von \mathbf{C}_1 und \mathbf{C}_2 .⁵⁶ Der gemeinsame Vektor der reduzierten Parameter sei schließlich $\beta = (\beta_0, \beta_1^\top, \beta_2^\top, \beta_{1 \times 2}^\top)^\top$.

Der Zusammenhang zwischen den neuen Parametern und den ursprünglichen Parametern, die den genannten Nebenbedingungen genügen, ist nun für jeden der drei Effekte wie im einfaktoriellen Fall. Der entsprechende Zusammenhang für die Parameter der Interaktion lässt sich dabei auf zwei verschiedene Arten formulieren – einmal mit $\beta_{1 \times 2}^*$ und einmal mit $\mathbf{B}_{1 \times 2}^*$.

$$\begin{aligned}\beta_1^* &= \mathbf{C}_1 \beta_1 \\ \beta_2^* &= \mathbf{C}_2 \beta_2 \\ \beta_{1 \times 2}^* &= \mathbf{C}_{1 \times 2} \beta_{1 \times 2} = (\mathbf{C}_2 \otimes \mathbf{C}_1) \beta_{1 \times 2} \\ \mathbf{B}_{1 \times 2}^* &= \mathbf{C}_1 \beta_{1 \times 2} \mathbf{C}_2^\top\end{aligned}$$

Aus dem Produkt jeweils einer ursprünglichen Inzidenzmatrix mit der zugehörigen Codiermatrix berechnen sich die zugehörigen Inzidenzmatrizen für die neuen Parameter: $\mathbf{X}_1 = \mathbf{X}_1^* \mathbf{C}_1$ für β_1 , $\mathbf{X}_2 = \mathbf{X}_2^* \mathbf{C}_2$ für β_2 und $\mathbf{X}_{1 \times 2} = \mathbf{X}_{1 \times 2}^* \mathbf{C}_{1 \times 2}$ für $\beta_{1 \times 2}$. Dabei kann $\mathbf{X}_{1 \times 2}$ äquivalent auch nach demselben Prinzip wie $\mathbf{X}_{1 \times 2}^*$ gebildet werden: Die $(p-1) \cdot (q-1)$ paarweisen Produkte der Spalten von \mathbf{X}_1 und \mathbf{X}_2 werden dazu spaltenweise zu $\mathbf{X}_{1 \times 2}$ zusammengestellt. Die reduzierte $(n \times p \cdot q)$ -Designmatrix mit vollem Spaltenrang ist $\mathbf{X} = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2|\mathbf{X}_{1 \times 2}]$, wobei $n > p \cdot q$ vorausgesetzt wird. Das Modell mit identifizierbaren Parametern lautet damit wieder analog zur einfaktoriellen Situation:

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2|\mathbf{X}_{1 \times 2}] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{1 \times 2} \end{pmatrix} = \mathbf{X} \beta$$

Der $p \cdot q$ -Vektor β der Parameter ist so wie im einfaktoriellen Fall über $\mathbf{X}^+ E(\mathbf{y})$ identifizierbar.

Die inhaltliche Bedeutung der Parameter in β hängt wie im einfaktoriellen Fall von der Wahl der Nebenbedingungen und Codiermatrizen ab. Voreinstellung in R sind Treatment-Kontraste: Für sie ergibt sich hier $\mathbf{v}_1 = (1, 0, \dots, 0)^\top$ und $\mathbf{v}_2 = (1, 0, \dots, 0)^\top$, d. h. es gilt $\beta_{1,1}^* = 0$, $\beta_{2,1}^* = 0$ sowie $\beta_{1 \times 2, 1k}^* = 0$ für alle k wie auch $\beta_{1 \times 2, j1}^* = 0$ für alle j . In der ersten Zeile und Spalte von

⁵⁶Hierbei ist die Reihenfolge relevant, mit denen man die Spalten von $\mathbf{X}_{1 \times 2}^*$ und entsprechend die Parameter in $\beta_{1 \times 2}^*$ ordnet, die zu den Kombinationen jk der Faktorstufen der UV 1 und UV 2 gehören: Variiert (wie hier) der erste Index j schnell und der zweite Index k langsam ($\beta_{1 \times 2}^* = (\beta_{11}^*, \beta_{21}^*, \beta_{31}^*, \beta_{12}^*, \beta_{22}^*, \beta_{32}^*)^\top$), gilt $\mathbf{C}_{1 \times 2} = \mathbf{C}_2 \otimes \mathbf{C}_1$. Dies ist die Voreinstellung in R, die sich etwa an der Ausgabe von `interaction((UV1), (UV2))` zeigt (Abschn. 2.6.2). Variiert dagegen j langsam und k schnell ($\beta_{1 \times 2}^* = (\beta_{11}^*, \beta_{12}^*, \beta_{21}^*, \beta_{22}^*, \beta_{31}^*, \beta_{32}^*)^\top$), ist $\mathbf{C}_{1 \times 2} = \mathbf{C}_1 \otimes \mathbf{C}_2$ zu setzen.

$\mathbf{B}_{1 \times 2}^*$ sind also alle Einträge 0. Mit Treatment-Kontrasten erhält man für die zweifaktorielle Varianzanalyse mit $p = 3, q = 2$ und einer Person pro Gruppe folgende Codiermatrizen \mathbf{C}_t :

$$\mathbf{C}_{t1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{C}_{t2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \mathbf{C}_{t1 \times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Die ursprünglichen Parameter erhalten mit Treatment-Kontrasten folgende Bedeutung: $\beta_0 = \mu_{11}$, $\beta_{1,j}^* = \mu_{j1} - \mu_{11}$, $\beta_{2,k}^* = \mu_{1k} - \mu_{11}$, $\beta_{1 \times 2,jk}^* = (\mu_{jk} - \mu_{j1}) - (\mu_{1k} - \mu_{11})$. Das zugehörige Modell mit reduzierter Designmatrix \mathbf{X} und dem reduzierten Parametervektor $\boldsymbol{\beta}$ lautet:

$$E(\mathbf{y}) = \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{31} \\ \mu_{12} \\ \mu_{22} \\ \mu_{32} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}_{\mathbf{X}} \begin{pmatrix} \beta_0 \\ \beta_{1.2} \\ \beta_{1.3} \\ \beta_{2.2} \\ \beta_{1 \times 2.22} \\ \beta_{1 \times 2.32} \end{pmatrix}_{\boldsymbol{\beta}}$$

$$= \begin{pmatrix} \beta_0 \\ \beta_0 + \beta_{1.2} \\ \beta_0 + \beta_{1.3} \\ \beta_0 + \beta_{2.2} \\ \beta_0 + \beta_{1.2} + \beta_{2.2} + \beta_{1 \times 2.22} \\ \beta_0 + \beta_{1.3} + \beta_{2.2} + \beta_{1 \times 2.32} \end{pmatrix}$$

Mit der ungewichteten Effektcodierung ergibt sich hier für die Nebenbedingungen $\mathbf{v}_1 = \mathbf{1}$ und $\mathbf{v}_2 = \mathbf{1}$, d.h. es gilt $\sum_j \beta_{1,j}^* = 0$, $\sum_k \beta_{2,k}^* = 0$ sowie $\sum_j \beta_{1 \times 2,jk}^* = 0$ für alle k wie auch $\sum_k \beta_{1 \times 2,jk}^* = 0$ für alle j (alle Zeilen- und Spaltensummen von $\mathbf{B}_{1 \times 2}^*$ sind 0). Für die zweifaktorielle Varianzanalyse mit $p = 3, q = 2$ und einer Person pro Gruppe erhält man mit Effektcodierung folgende Codiermatrizen \mathbf{C}_e :

$$\mathbf{C}_{e1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \quad \mathbf{C}_{e2} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \mathbf{C}_{e1 \times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{pmatrix}$$

Die ursprünglichen Parameter $\beta_0, \beta_{1,jk}^*, \beta_{2,jk}^*$ und $\beta_{1 \times 2,jk}^*$ erhalten durch die Nebenbedingungen die Bedeutung der Parameter μ, α_j, β_k und $(\alpha\beta)_{jk}$ der häufig gewählten Formulierung des

Modells der zweifaktoriellen Varianzanalyse als $\mu_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$.⁵⁷ Mit Effektcodierung lautet das Modell mit reduzierter Designmatrix \mathbf{X} und dem reduzierten Parametervektor β :

$$E(\mathbf{y}) = \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{31} \\ \mu_{12} \\ \mu_{22} \\ \mu_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}_{\mathbf{X}} \begin{pmatrix} \beta_0 \\ \beta_{1.1} \\ \beta_{1.2} \\ \beta_{2.1} \\ \beta_{1 \times 2.11} \\ \beta_{1 \times 2.21} \end{pmatrix}_{\beta}$$

$$= \begin{pmatrix} \beta_0 + \beta_{1.1} + \beta_{2.1} + \beta_{1 \times 2.11} \\ \beta_0 + \beta_{1.2} + \beta_{2.1} + \beta_{1 \times 2.21} \\ \beta_0 - (\beta_{1.1} + \beta_{1.2}) + \beta_{2.1} - (\beta_{1 \times 2.11} + \beta_{1 \times 2.21}) \\ \beta_0 + \beta_{1.1} - \beta_{2.1} - \beta_{1 \times 2.11} \\ \beta_0 + \beta_{1.2} - \beta_{2.1} - \beta_{1 \times 2.21} \\ \beta_0 - (\beta_{1.1} + \beta_{1.2}) - \beta_{2.1} + \beta_{1 \times 2.11} + \beta_{1 \times 2.21} \end{pmatrix}$$

12.9.4 Parameterschätzungen, Vorhersage und Residuen

Der n -Vektor der Beobachtungen $\mathbf{y} = (y_1, \dots, y_n)^\top$ ergibt sich im Modell des ALM als Summe von $E(\mathbf{y})$ und einem n -Vektor zufälliger Fehler $\epsilon = (\epsilon_1, \dots, \epsilon_n)^\top$.

$$y_i = E(y_i) + \epsilon_i$$

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Die Fehler sollen dabei gemeinsam unabhängig und auch unabhängig von \mathbf{X} sein. Ferner soll für alle $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ gelten, womit ϵ multivariat normalverteilt ist mit $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. Die Voraussetzung gleicher Fehlervarianzen wird im Fall der Regression als Homoskedastizität bezeichnet, im Fall der Varianzanalyse als Varianzhomogenität (Abb. 12.5). Die Beobachtungen sind dann ihrerseits multivariat normalverteilt mit $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$.⁵⁸

Der Vektor $\hat{\beta}$ der i. S. der geringsten Quadratsumme der Residuen optimalen Parameterschätzungen ist gleich $\mathbf{X}^+ \mathbf{y} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$.⁵⁹ Man erhält $\hat{\beta}$ also als Koordinaten des Kriteriums,

⁵⁷Hierbei seien $\alpha_j = \mu_{.j} - \mu$ die Effektgrößen des Haupteffekts der ersten UV, $\beta_k = \mu_{.k} - \mu$ die des Haupteffekts der zweiten UV und $(\alpha\beta)_{jk} = \mu_{jk} - (\mu + \alpha_j + \beta_k) = \mu_{jk} - \mu_{.j} - \mu_{.k} + \mu$ die der Interaktion. Dabei seien $\mu_{.j} = \frac{1}{q} \sum_k \mu_{jk}$, $\mu_{.k} = \frac{1}{p} \sum_j \mu_{jk}$ und $\mu = \frac{1}{p \cdot q} \sum_j \sum_k \mu_{jk}$ ungewichtete mittlere Erwartungswerte. Liegen gleiche, oder zumindest proportional ungleiche Zellbesetzungen vor ($\frac{n_{jk}}{n_{jk'}} = \frac{n_{j'k}}{n_{j'k'}}$ sowie $\frac{n_{jk}}{n_{j'k}} = \frac{n_{jk'}}{n_{j'k'}}$ für alle j, j', k, k'), lässt sich die Parametrisierung mit gewichteten mittleren Erwartungswerten durch die gewichtete Effektcodierung umsetzen (Fußnote 55).

⁵⁸Zunächst ist $E(\mathbf{y}) = E(\mathbf{X}\beta + \epsilon) = \mathbf{X}\beta + E(\epsilon) = \mathbf{X}\beta$. Weiter gilt $V(\mathbf{y}) = V(\mathbf{X}\beta + \epsilon) = V(\epsilon) = \sigma^2 \mathbf{I}$.

⁵⁹In der Varianzanalyse werden die reduzierten Parameter geschätzt. Für die Beziehung zwischen geschätzten ursprünglichen Parametern und geschätzten reduzierten Parametern gilt $\hat{\beta}^* = \mathbf{C}\hat{\beta}$. Auf Basis eines mit `aov()` oder `lm()` angepassten linearen Modells erhält man $\hat{\beta}$ mit `coef()` und $\hat{\beta}^*$ mit `dummy.coef()`.

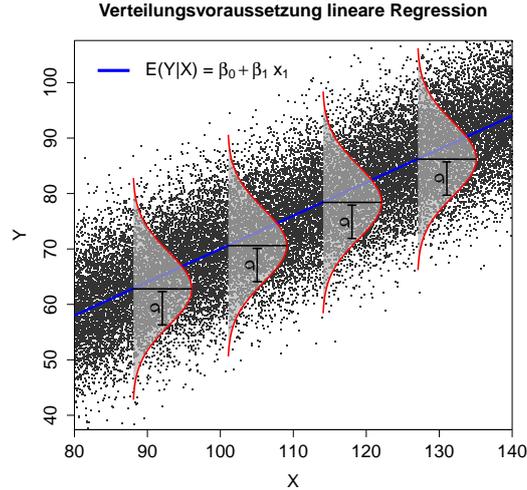


Abbildung 12.5: Verteilungsvoraussetzungen im allgemeinen linearen Modell am Beispiel der einfachen linearen Regression

das orthogonal auf V projiziert wurde, bzgl. der durch \mathbf{X} definierten Basis (Abschn. 12.1.7). Es gilt $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$, $\hat{\beta}$ ist damit erwartungstreuer Schätzer für β .⁶⁰

Der n -Vektor $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)^\top$ der vorhergesagten Werte berechnet sich durch $\mathbf{X}\mathbf{X}^+\mathbf{y} = \mathbf{P}\mathbf{y}$, wobei die orthogonale Projektion $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$ auch als *Hat-Matrix* bezeichnet wird. Man erhält $\hat{\mathbf{y}}$ also als Koordinaten des Kriteriums, das orthogonal auf V projiziert wurde, bzgl. der Standardbasis. Die Vorhersage $\hat{\mathbf{y}}$ liegt damit in V .

Für den n -Vektor der Residuen $\mathbf{e} = (e_1, \dots, e_n)^\top$ gilt $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{I}\mathbf{y} - \mathbf{P}\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y}$. Die Residuen ergeben sich also als Projektion des Kriteriums auf das orthogonale Komplement von V . Die Residuen \mathbf{e} liegen damit in V^\perp und sind senkrecht zur Vorhersage. V^\perp wird auch als *Fehlerraum* bezeichnet und besitzt die Dimension $\text{Rang}(\mathbf{I} - \mathbf{P}) = n - \text{Rang}(\mathbf{X})$. Weiter gilt $E(\mathbf{e}) = \mathbf{0}$ und $V(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{P})$.⁶¹

Erwartungstreuer Schätzer für σ^2 ist die Quadratsumme der Residuen $SS_e = \|\mathbf{e}\|^2 = \mathbf{y}^\top (\mathbf{I} - \mathbf{P})\mathbf{y}$ geteilt durch die Anzahl der Fehler-Freiheitsgrade, die gleich der Dimension von V^\perp ist.⁶² In der Regression mit p Prädiktoren und absolutem Term ist $\hat{\sigma}^2 = \frac{\|\mathbf{e}\|^2}{n - (p+1)}$ der quadrierte Standardschätzfehler, in der einfaktoriellen Varianzanalyse mit p Gruppen ist $\hat{\sigma}^2 = \frac{\|\mathbf{e}\|^2}{n - p}$ die mittlere Quadratsumme der Residuen.

Im multivariaten Fall werden mehrere Kriteriums- bzw. AV-Vektoren \mathbf{y}_i spaltenweise zu einer Matrix \mathbf{Y} zusammengestellt. Setzt man in den genannten Formeln \mathbf{Y} für \mathbf{y} ein, berechnen sich

⁶⁰Zunächst ist $E(\hat{\beta}) = E(\mathbf{X}^+\mathbf{y}) = \mathbf{X}^+E(\mathbf{y}) = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{X}\beta = \beta$. Weiter gilt $V(\hat{\beta}) = V(\mathbf{X}^+\mathbf{y}) = \mathbf{X}^+V(\mathbf{y})(\mathbf{X}^+)^\top = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \sigma^2 \mathbf{I} \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$.

⁶¹Zunächst ist $E(\mathbf{e}) = E((\mathbf{I} - \mathbf{P})\mathbf{y}) = E(\mathbf{y}) - E(\mathbf{P}\mathbf{y}) = \mathbf{X}\beta - \mathbf{P}\mathbf{X}\beta$. Da die Spalten von \mathbf{X} in V liegen, bleiben sie durch \mathbf{P} unverändert, es folgt also $E(\mathbf{e}) = \mathbf{X}\beta - \mathbf{P}\mathbf{X}\beta = \mathbf{0}$. Weiter gilt $V(\mathbf{e}) = V((\mathbf{I} - \mathbf{P})\mathbf{y}) = (\mathbf{I} - \mathbf{P})V(\mathbf{y})(\mathbf{I} - \mathbf{P})^\top = \sigma^2(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})^\top$. Als orthogonale Projektion ist $\mathbf{I} - \mathbf{P}$ symmetrisch und idempotent, es gilt also $V(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})^\top = \sigma^2(\mathbf{I} - \mathbf{P})$.

⁶² $SS_e = \sum_i e_i^2 = \|\mathbf{e}\|^2 = \mathbf{e}^\top \mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}})^\top (\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \hat{\mathbf{y}} - \hat{\mathbf{y}}^\top \mathbf{y} + \hat{\mathbf{y}}^\top \hat{\mathbf{y}} = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}\mathbf{y} - (\mathbf{P}\mathbf{y})^\top \mathbf{y} + (\mathbf{P}\mathbf{y})^\top (\mathbf{P}\mathbf{y}) = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}\mathbf{y} - \mathbf{y}^\top \mathbf{P}\mathbf{y} + \mathbf{y}^\top \mathbf{P}\mathbf{P}\mathbf{y} = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}\mathbf{y} = \mathbf{y}^\top (\mathbf{I} - \mathbf{P})\mathbf{y}$.

Parameterschätzungen, die spaltenweise aus den Vektoren der Vorhersagen $\hat{\mathbf{y}}_l$ zusammengestellte Matrix $\hat{\mathbf{Y}}$ und die spaltenweise aus den Vektoren der Residuen \mathbf{e}_l zusammengestellte Matrix \mathbf{E} wie in der univariaten Formulierung.

12.9.5 Hypothesen über parametrische Funktionen testen

Im ALM können verschiedenartige Hypothesen über die Modellparameter β_j getestet werden. Eine Gruppe solcher Hypothesen ist jene über den Wert einer *parametrischen Funktion*. Im univariaten Fall ist eine parametrische Funktion $\psi = \sum_j c_j \beta_j = \mathbf{c}^\top \boldsymbol{\beta}$ eine Linearkombination der Parameter, wobei die Koeffizienten c_j zu einem Vektor \mathbf{c} zusammengestellt werden. Beispiele für parametrische Funktionen sind etwa a-priori Kontraste aus der Varianzanalyse (Abschn. 7.3.6). Auch ein Parameter β_j selbst (z. B. ein Gewicht in der Regression) ist eine parametrische Funktion, bei der \mathbf{c} der j -te Einheitsvektor ist. Die H_0 lässt sich dann als $\psi = \psi_0$ mit einem festen Wert ψ_0 formulieren.

Eine parametrische Funktion wird mit Hilfe der Parameterschätzungen als $\hat{\psi} = \mathbf{c}^\top \hat{\boldsymbol{\beta}}$ erwartungstreu geschätzt, und es gilt $\hat{\psi} \sim \mathcal{N}(\psi, \sigma^2 \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c})$.⁶³ $\hat{\psi}$ lässt sich auch direkt als Linearkombination der Beobachtungen y_i im Vektor \mathbf{y} formulieren:

$$\hat{\psi} = \mathbf{c}^\top \hat{\boldsymbol{\beta}} = \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = (\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c})^\top \mathbf{y} = \mathbf{a}^\top \mathbf{y}$$

Dabei ist $\mathbf{a} = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c}$ der zu \mathbf{c} gehörende n -Vektor der *Schätzerkoeffizienten*, mit dem auch $\hat{\psi} = \mathbf{a}^\top \hat{\mathbf{y}}$ ⁶⁴ sowie $\sigma_{\hat{\psi}}^2 = \mathbf{a}^\top \mathbf{a} \sigma^2 = \|\mathbf{a}\|^2 \sigma^2$ gilt.⁶⁵ Damit lässt sich die t -Teststatistik in der üblichen Form $\frac{\hat{\psi} - \psi_0}{\hat{\sigma}_{\hat{\psi}}}$ definieren:

$$t = \frac{\hat{\psi} - \psi_0}{\hat{\sigma}_{\hat{\psi}}} = \frac{\hat{\psi} - \psi_0}{\|\mathbf{a}\| \sqrt{\|\mathbf{e}\|^2 / (n - \text{Rang}(\mathbf{X}))}}$$

Unter H_0 ist t zentral t -verteilt mit $n - \text{Rang}(\mathbf{X})$ Freiheitsgraden (im Fall der Regression $n - (p + 1)$, im Fall der einfaktoriellen Varianzanalyse $(n - p)$).

12.9.6 Lineare Hypothesen als Modellvergleiche formulieren

Analog zur einzelnen parametrischen Funktion können Hypothesen über einen Vektor $\boldsymbol{\psi}$ von m parametrischen Funktionen $\psi_j = \mathbf{c}_j^\top \boldsymbol{\beta}$ gleichzeitig formuliert werden, wobei die ψ_j linear unabhängig sein sollen. Diese *linearen* Hypothesen besitzen die Form $\boldsymbol{\psi} = \mathbf{L} \boldsymbol{\beta}$. Dabei ist \mathbf{L} eine Matrix aus den zeilenweise zusammengestellten Koeffizientenvektoren \mathbf{c}_j für die Linearkombinationen der allgemein u Parameter im Vektor $\boldsymbol{\beta}$ ($u > m$). \mathbf{L} ist dann eine $(m \times u)$ -Matrix mit Rang m . Der unter H_0 für $\boldsymbol{\psi}$ angenommene Vektor sei $\boldsymbol{\psi}_0$.

⁶³Zunächst ist $E(\hat{\psi}) = E(\mathbf{c}^\top \hat{\boldsymbol{\beta}}) = \mathbf{c}^\top E(\hat{\boldsymbol{\beta}}) = \mathbf{c}^\top \boldsymbol{\beta} = \psi$. Weiter gilt $V(\hat{\psi}) = V(\mathbf{c}^\top \hat{\boldsymbol{\beta}}) = \mathbf{c}^\top V(\hat{\boldsymbol{\beta}}) \mathbf{c} = \mathbf{c}^\top \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c}$. Bei $\hat{\psi}$ handelt es sich um einen *Gauß-Markoff-Schätzer*, also den linearen erwartungstreuen Schätzer mit der geringsten Varianz.

⁶⁴Zunächst ist $\mathbf{X}^\top \mathbf{a} = \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c} = \mathbf{c}$, also $\mathbf{c}^\top = \mathbf{a}^\top \mathbf{X}$. Damit gilt $\mathbf{a}^\top \mathbf{y} = \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{a}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \mathbf{a}^\top \mathbf{P} \mathbf{y} = \mathbf{a}^\top \hat{\mathbf{y}}$.

⁶⁵ $\|\mathbf{a}\|^2 = \mathbf{a}^\top \mathbf{a} = (\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c})^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c} = \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c} = \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c}$.

Im multivariaten Fall mit r Variablen Y_l und der $(u \times r)$ -Matrix der Parameter \mathbf{B} hat eine lineare Hypothese die Form $\boldsymbol{\Psi} = \mathbf{L}\mathbf{B}$, wobei die unter H_0 für $\boldsymbol{\Psi}$ angenommene Matrix $\boldsymbol{\Psi}_0$ sei.

Die hier vorgestellten linearen Hypothesen aus der Varianzanalyse und Regression haben die Form $\boldsymbol{\psi}_0 = \mathbf{0}$ (bzw. $\boldsymbol{\Psi}_0 = \mathbf{0}$) und beziehen sich auf den Vergleich zweier nested Modelle (Abschn. 6.3.3, 7.3.3): Das umfassendere (*unrestricted*) H_1 -Modell mit Designmatrix \mathbf{X}_u besitzt dabei im univariaten Fall allgemein u freie Parameter. Für das eingeschränkte (*restricted*) H_0 -Modell mit Designmatrix \mathbf{X}_r nimmt man an, dass m der Parameter des umfassenderen Modells 0 sind, wodurch es noch $r = u - m$ freie Parameter besitzt. Man erhält \mathbf{X}_r , indem man die zu den auf 0 gesetzten Parametern gehörenden m Spalten von \mathbf{X}_u streicht.

Die freien Parameter des eingeschränkten Modells bilden eine echte Teilmenge jener des umfassenderen Modells. Daher liegt der von den Spalten von \mathbf{X}_r aufgespannte Unterraum V_r der Vorhersage des eingeschränkten Modells vollständig in jenem des umfassenderen Modells V_u , dem Erzeugnis der Spalten von \mathbf{X}_u . Umgekehrt liegt der Fehlerraum des umfassenderen Modells V_u^\perp vollständig in jenem des eingeschränkten Modells V_r^\perp . Die H_0 lässt sich auch so formulieren, dass $E(\mathbf{y})$ in V_r liegt.

Im multivariaten Fall besitzt \mathbf{B} allgemein u Zeilen mit freien Parametern. Für das eingeschränkte H_0 -Modell nimmt man an, dass m Zeilen von \mathbf{B} gleich 0 sind. Dadurch besitzt \mathbf{B} unter H_0 noch $r = u - m$ Zeilen mit freien Parametern. \mathbf{X}_u und \mathbf{X}_r stehen wie im univariaten Fall zueinander, unter H_0 soll also jede Spalte von $E(\mathbf{Y})$ in V_r liegen. Für die inferenzstatistische Prüfung s. Abschn. 12.9.7.

Univariate einfaktorielle Varianzanalyse

In der univariaten einfaktoriellen Varianzanalyse sind unter H_0 alle Gruppenerwartungswerte gleich. Dies ist äquivalent zur Hypothese, dass $\boldsymbol{\beta}_{p-1} = \mathbf{0}$, also $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_{p-1}^\top)^\top = (\beta_0, \mathbf{0}^\top)^\top$ gilt. In $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ hat \mathbf{L} hier damit die Form $[\mathbf{0}|\mathbf{I}]$, wobei $\mathbf{0}$ der $(p-1)$ -Vektor $(0, \dots, 0)^\top$ und \mathbf{I} die $((p-1) \times (p-1))$ -Einheitsmatrix ist. Das H_0 -Modell mit einem Parameter β_0 lässt sich als $E(\mathbf{y}) = \mathbf{1}\beta_0$ formulieren. Im umfassenderen H_1 -Modell mit p Parametern gibt es keine Einschränkung für $\boldsymbol{\beta}_{p-1}$, es ist damit identisch zum vollständigen Modell der einfaktoriellen Varianzanalyse $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$: Hier ist also $u = p$, $\mathbf{X}_u = \mathbf{X}$, $r = 1$, $\mathbf{X}_r = \mathbf{1}$ und $m = p - 1$ (Abschn. 12.9.2).

Univariate zweifaktorielle Varianzanalyse: Quadratsummen vom Typ I

Wie in Abschn. 7.5.2 erläutert, existieren in der zweifaktoriellen Varianzanalyse (CRF- pq Design) Typen von Quadratsummen, die bei ungleichen Zellbesetzungen zu unterschiedlichen Tests führen können. Proportional ungleiche Zellbesetzungen sind dabei gegeben, wenn $\frac{n_{jk}}{n_{j'k'}} = \frac{n_{jk}}{n_{j'k'}}$ sowie $\frac{n_{jk}}{n_{j'k}} = \frac{n_{jk'}}{n_{j'k'}}$ für alle j, j', k, k' gilt. Ist dies nicht der Fall, spricht man von unbalancierten Zellbesetzungen.

Zunächst sind für den Test jedes Effekts (beide Haupteffekte und Interaktionseffekt) das eingeschränkte H_0 -Modell sowie das umfassendere H_1 -Modell zu definieren. Die H_0 des Haupteffekts der ersten UV lässt sich mit Hilfe des zugehörigen Parametervektors als $\boldsymbol{\beta}_1 = \mathbf{0}$ formulieren. Bei

Quadratsummen vom Typ I wählt man für den Test des Haupteffekts der UV 1 als H_0 -Modell das Gesamt-Nullmodell, bei dem alle Parametervektoren $\beta_1, \beta_2, \beta_{1 \times 2}$ gleich $\mathbf{0}$ sind und damit $E(\mathbf{y}) = \mathbf{1}\beta_0$ gilt. Hier gilt also für die Designmatrix des eingeschränkten Modells $\mathbf{X}_r = \mathbf{X}_0 = \mathbf{1}$. Das sequentiell folgende H_1 -Modell für den Test der ersten UV ist jenes, bei dem man die zugehörigen $p - 1$ Parameter in β_1 dem eingeschränkten Modell hinzufügt und die Designmatrix entsprechend erweitert, d. h. $\mathbf{X}_u = [\mathbf{1}|\mathbf{X}_1]$:

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1] \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Die H_0 des Haupteffekts der UV 2 lautet mit Hilfe des zugehörigen Parametervektors $\beta_2 = \mathbf{0}$. Bei Quadratsummen vom Typ I fällt die Wahl für das H_0 -Modell der zweiten UV auf das H_1 -Modell der ersten UV ($\mathbf{X}_r = [\mathbf{1}|\mathbf{X}_1]$). Das H_1 -Modell für den Test der zweiten UV ist jenes mit der sequentiell erweiterten Designmatrix $\mathbf{X}_u = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2]$:

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Durch den sequentiellen Aufbau der Modellvergleiche ist die Reihenfolge der UVn beim Test der Haupteffekte in Fällen mit unbalancierten Zellbesetzungen rechnerisch bedeutsam.

Die H_0 des Interaktionseffekts lautet mit Hilfe des zugehörigen Parametervektors $\beta_{1 \times 2} = \mathbf{0}$. Die Wahl für das H_0 -Modell der Interaktion ist das sequentielle H_1 -Modell der UV 2 ($\mathbf{X}_r = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2]$). Das H_1 -Modell der Interaktion ist das in Abschn. 12.9.3 vorgestellte vollständige Modell der zweifaktoriellen Varianzanalyse ($\mathbf{X}_u = \mathbf{X}$):

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2|\mathbf{X}_{1 \times 2}] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{1 \times 2} \end{pmatrix} = \mathbf{X}\beta$$

In $H_0 : \mathbf{L}\beta = \mathbf{0}$ hat \mathbf{L} in den genannten Tests der Haupteffekte die Form $[\mathbf{0}|\mathbf{I}|\mathbf{0}]$ und beim Test der Interaktion die Form $[\mathbf{0}|\mathbf{I}]$. Dabei ist \mathbf{I} der Reihe nach die $((p - 1) \times (p - 1))$ -, $((q - 1) \times (q - 1))$ - und $((p - 1) \cdot (q - 1) \times (p - 1) \cdot (q - 1))$ -Einheitsmatrix und $\mathbf{0}$ eine passende Matrix aus 0-Einträgen.

Zweifaktorielle Varianzanalyse: Quadratsummen vom Typ II

Auch bei Quadratsummen vom Typ II unterscheiden sich die eingeschränkten und umfassenderen Modelle um jeweils $p - 1$ (erster Haupteffekt), $q - 1$ (zweiter Haupteffekt) und $(p - 1) \cdot (q - 1)$ (Interaktion) freie Parameter. Das H_1 -Modell beim Test beider Haupteffekte ist hier:

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Das jeweilige H_0 -Modell für den Test des Haupteffekts einer UV wird bei Quadratsummen vom Typ II dadurch gebildet, dass ausgehend vom H_1 -Modell die Parameter des zu testenden Effekts auf 0 gesetzt und die zugehörigen Spalten der Designmatrix entfernt werden.

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_2] \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix} \quad (\text{H}_0 - \text{Modell für Test der UV 1})$$

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1] \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad (\text{H}_0 - \text{Modell für Test der UV 2})$$

Die Reihenfolge der UVn ist anders als bei Quadratsummen vom Typ I beim Test der Haupteffekte unwesentlich, selbst wenn unbalancierte Zellbesetzungen vorliegen. Liegen proportional ungleiche Zellbesetzungen vor, stimmen die Ergebnisse für den Test der Haupteffekte mit jenen aus Quadratsummen vom Typ I überein. Bei Quadratsummen vom Typ II ist die Wahl für H_0 - und H_1 -Modell beim Test der Interaktion gleich jener bei Quadratsummen vom Typ I und III, die Ergebnisse sind daher identisch.

Zweifaktorielle Varianzanalyse: Quadratsummen vom Typ III

Auch bei Quadratsummen vom Typ III unterscheiden sich die eingeschränkten und umfassenden Modelle um jeweils $p - 1$ (erster Haupteffekt), $q - 1$ (zweiter Haupteffekt) und $(p - 1) \cdot (q - 1)$ (Interaktion) freie Parameter. Hier ist beim Test aller Effekte das H_1 -Modell das in Abschn. 12.9.3 vorgestellte vollständige Modell der zweifaktoriellen Varianzanalyse:

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2|\mathbf{X}_{1 \times 2}] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{1 \times 2} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta}$$

Das jeweilige H_0 -Modell für den Test aller Effekte wird bei Quadratsummen vom Typ III dadurch gebildet, dass ausgehend vom vollständigen Modell die Parameter des zu testenden Effekts auf 0 gesetzt und die zugehörigen Spalten der Designmatrix gestrichen werden:

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_2|\mathbf{X}_{1 \times 2}] \begin{pmatrix} \beta_0 \\ \beta_2 \\ \beta_{1 \times 2} \end{pmatrix} \quad (\text{H}_0 - \text{Modell für Test UV 1})$$

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_{1 \times 2}] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_{1 \times 2} \end{pmatrix} \quad (\text{H}_0 - \text{Modell für Test UV 2})$$

$$E(\mathbf{y}) = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2] \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad (\text{H}_0 - \text{Modell für Test Interaktion})$$

Die Reihenfolge der UVn ist anders als bei Quadratsummen vom Typ I beim Test der Haupteffekte unwesentlich, selbst wenn unbalancierte Zellbesetzungen vorliegen. Bei gleichen Zellbesetzungen liefern alle drei Typen von Quadratsummen dieselben Ergebnisse. Bei Quadratsummen vom Typ III stimmt die Wahl für H_0 - und H_1 -Modell beim Test der Interaktion mit jener bei Quadratsummen vom Typ I und II überein, die Ergebnisse sind daher identisch. Ein Test des absoluten Terms β_0 ist mit Quadratsummen vom Typ III ebenso wenig möglich wie ein Test von Haupteffekten, die nicht Teil einer Interaktion sind.

Die H_0 -Modelle für Haupteffekte machen bei Quadratsummen vom Typ III keine Einschränkung für den Parametervektor $\beta_{1 \times 2}$ der Interaktion, obwohl die Parameter des zu testenden, und an der Interaktion beteiligten Haupteffekts auf 0 gesetzt werden. Die Modellvergleiche verletzen damit anders als bei Quadratsummen vom Typ I und II das Prinzip, dass Vorhersageterme höherer Ordnung nur in das Modell aufgenommen werden, wenn alle zugehörigen Terme niedriger Ordnung ebenfalls Berücksichtigung finden.

Modellvergleiche für Quadratsummen vom Typ III können deshalb bei ungleichen Zellbesetzungen in Abhängigkeit von der verwendeten Nebenbedingung für die Parameter samt Codierschema unterschiedlich ausfallen (Venables, 2016). Richtige Ergebnisse erhält man mit $\mathbf{v} = \mathbf{1}$, was alle Codierschemata gewährleisten, deren Koeffizienten sich über die Spalten der Codiermatrix \mathbf{C} zu 0 summieren (Abschn. 12.9.2). Dazu zählen Effekt- und Helmert-Codierung, nicht aber Treatment-Kontraste (Dummy-Codierung) – die Voreinstellung in R. Bei Quadratsummen vom Typ I und II spielt die Wahl von Nebenbedingung und Codierschema dagegen keine Rolle für den inferenzstatistischen Test.

Multiple Regression

Die Modellvergleiche beim Test von p Prädiktoren X_j in der multiplen Regression werden nach demselben Prinzip wie in der zweifaktoriellen Varianzanalyse gebildet: Das Gesamt-Nullmodell ist $E(\mathbf{Y}) = \mathbf{1}\beta_0^\top$, bei dem für alle Kriterien y_l der zugehörige Parametervektor der p Prädiktoren $\beta_{p,l} = \mathbf{0}$ und damit $\mathbf{B}_{p,0} = \mathbf{0}$ sowie $\mathbf{X}_0 = \mathbf{1}$ ist. Das vollständige Modell ist jenes mit allen Prädiktoren $E(\mathbf{Y}) = \mathbf{X}\mathbf{B}$.

Die Prädiktoren werden über Modellvergleiche getestet, die wie in der zweifaktoriellen Varianzanalyse vom verwendeten Typ der Quadratsummen abhängen. Die Reihenfolge der Prädiktoren ist im multivariaten Fall bei Quadratsummen vom Typ I bedeutsam, sofern nicht alle Prädiktoren paarweise unkorreliert sind: Für den Test des ersten Prädiktors ist das eingeschränkte H_0 -Modell das Gesamt-Nullmodell, das zugehörige umfassendere H_1 -Modell ist jenes mit nur dem ersten Prädiktor. Für den Test des zweiten Prädiktors ist das eingeschränkte H_0 -Modell das sequentielle H_1 -Modell des ersten Prädiktors, das H_1 -Modell ist jenes mit den ersten beiden Prädiktoren. Für die Tests aller folgenden Prädiktoren gilt dies analog.

Bei Quadratsummen vom Typ II ist das H_1 -Modell beim Test aller Prädiktoren gleich dem vollständigen Modell mit allen Prädiktoren, jedoch ohne deren Interaktionen. Das H_0 -Modell beim Test eines Prädiktors X_j entsteht aus dem H_1 -Modell, indem der zu X_j gehörende Parametervektor $\beta_j^\top = \mathbf{0}^\top$ gesetzt und die passende Spalte der Designmatrix gestrichen wird.

Bei Quadratsummen vom Typ III ist das H_1 -Modell beim Test aller Prädiktoren gleich dem vollständigen Modell mit allen Prädiktoren und deren Interaktionen, sofern letztere berücksichtigt werden sollen. Das H_0 -Modell beim Test eines Prädiktors X_j entsteht aus dem H_1 -Modell, indem der zu X_j gehörende Parametervektor $\beta_j^\top = \mathbf{0}^\top$ gesetzt und die passende Spalte der Designmatrix gestrichen wird. Quadratsummen vom Typ II und III unterscheiden sich also nur, wenn die Regression auch Interaktionsterme einbezieht.

Sollen allgemein die Parameter von m Prädiktoren gleichzeitig daraufhin getestet werden, ob sie 0 sind, ist das Vorgehen analog, wobei unter H_0 die zugehörigen, aus \mathbf{B} stammenden m Parametervektoren $\beta_j^\top = \mathbf{0}^\top$ gesetzt und entsprechend ausgehend vom H_1 -Modell die passenden Spalten der Designmatrix gestrichen werden.

12.9.7 Lineare Hypothesen testen

Liegt eine lineare Hypothese der Form $\boldsymbol{\psi} = \mathbf{L}\boldsymbol{\beta}$ vor (Fox, Friendly & Weisberg, 2013), soll \mathbf{A} die zur $((u-r) \times u)$ -Matrix \mathbf{L} gehörende $((u-r) \times n)$ -Matrix bezeichnen, die zeilenweise aus den n -Vektoren der Schätzerkoeffizienten $\mathbf{a}_j = \mathbf{X}_u(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{c}_j$ gebildet wird (Abschn. 12.9.5). Es gilt also $\mathbf{A}^\top = \mathbf{X}_u(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{L}^\top$ und $\text{Rang}(\mathbf{A}) = \text{Rang}(\mathbf{L}) = u-r$. Aus $\mathbf{c} = \mathbf{X}^\top \mathbf{a}$ (Abschn. 12.9.6, Fußnote 64) folgt damit $\mathbf{L}^\top = \mathbf{X}_u^\top \mathbf{A}^\top$, also $\mathbf{L} = \mathbf{A}\mathbf{X}_u$.⁶⁶

Der Vektor der Schätzungen $\hat{\boldsymbol{\psi}} = \mathbf{L}\hat{\boldsymbol{\beta}}$ kann mit \mathbf{A} analog zur Schätzung einer einfachen parametrischen Funktion $\hat{\psi}_j = \mathbf{c}_j^\top \hat{\boldsymbol{\beta}} = \mathbf{a}_j^\top \mathbf{y}$ auch als Abbildung des Vektors der Beobachtungen \mathbf{y} formuliert werden:

$$\hat{\boldsymbol{\psi}} = \mathbf{L}\hat{\boldsymbol{\beta}} = \mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{X}_u^\top \mathbf{y} = \mathbf{A}\mathbf{y}$$

Mit der Vorhersage $\hat{\mathbf{y}}$ gilt zudem $\hat{\boldsymbol{\psi}} = \mathbf{A}\hat{\mathbf{y}}$.⁶⁷ $\hat{\boldsymbol{\psi}}$ ist ein erwartungstreuer Gauß-Markoff-Schätzer mit Verteilung $\hat{\boldsymbol{\psi}} \sim \mathcal{N}(\boldsymbol{\psi}, \sigma^2 \mathbf{A}\mathbf{A}^\top)$.⁶⁸ Der Rang der Kovarianzmatrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}} = \sigma^2 \mathbf{A}\mathbf{A}^\top$ beträgt $u-r$. Er ist also gleich der Differenz der Anzahl zu schätzender Parameter beider Modelle sowie gleich der Differenz der Dimensionen von V_u und V_r einerseits und von V_r^\perp und V_u^\perp andererseits.

Univariate Teststatistik

Um die Teststatistik für den univariaten Fall zu motivieren, ist zunächst festzustellen, dass für die quadrierte Mahalanobisdistanz (Abschn. 12.1.4) $\|\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\|_{M, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}}}^2$ von $\hat{\boldsymbol{\psi}}$ zu $\boldsymbol{\psi}_0 = \mathbf{0}$ bzgl. $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}}$ gilt (Fußnoten 67 und 68):

$$\begin{aligned} \|\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\|_{M, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}}}^2 &= (\hat{\boldsymbol{\psi}} - \mathbf{0})^\top \boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}}^{-1} (\hat{\boldsymbol{\psi}} - \mathbf{0}) = \frac{\hat{\boldsymbol{\psi}}^\top (\mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{L}^\top)^{-1} \hat{\boldsymbol{\psi}}}{\sigma^2} \\ &= \frac{\hat{\boldsymbol{\psi}}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \hat{\boldsymbol{\psi}}}{\sigma^2} = \frac{\mathbf{y}^\top \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{A}\mathbf{y}}{\sigma^2} \end{aligned}$$

⁶⁶ $\mathbf{A}\mathbf{X}_u = (\mathbf{X}_u(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{L}^\top)^\top \mathbf{X}_u = \mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{X}_u^\top \mathbf{X}_u = \mathbf{L}$.

⁶⁷ $\hat{\boldsymbol{\psi}} = \mathbf{L}\hat{\boldsymbol{\beta}} = \mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{X}_u^\top \mathbf{y} = \mathbf{A}\mathbf{X}_u(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{X}_u^\top \mathbf{y} = \mathbf{A}\hat{\mathbf{y}}$.

⁶⁸ Zunächst gilt $E(\mathbf{L}\hat{\boldsymbol{\beta}}) = \mathbf{L}E(\hat{\boldsymbol{\beta}}) = \mathbf{L}\boldsymbol{\beta} = \boldsymbol{\psi}$. Weiter ist $V(\hat{\boldsymbol{\psi}}) = V(\mathbf{L}\hat{\boldsymbol{\beta}}) = \mathbf{L}V(\hat{\boldsymbol{\beta}})\mathbf{L}^\top = \sigma^2 \mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{L}^\top = \sigma^2 \mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{X}_u^\top \mathbf{X}_u (\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{L}^\top = \sigma^2 \mathbf{A}\mathbf{A}^\top$. Insbesondere ist also $\mathbf{A}\mathbf{A}^\top = \mathbf{L}(\mathbf{X}_u^\top \mathbf{X}_u)^{-1} \mathbf{L}^\top$.

Hier ist $\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$ die Projektion auf das orthogonale Komplement von V_r in V_u (also $V_r^\perp \cap V_u$), dessen spaltenweise Basis \mathbf{A}^\top ist.⁶⁹

Für die Schätzung der Fehlervarianz σ^2 benötigt man das vollständige (*full*) Modell mit Designmatrix \mathbf{X}_f und Projektionsmatrix \mathbf{P}_f (bisher einfach als \mathbf{X} und \mathbf{P} bezeichnet), das alle Parameter beinhaltet. Der zugehörige Vektor der Residuen sei $\mathbf{e}_f = \mathbf{y} - \hat{\mathbf{y}}_f$, die Residual-Quadratsumme also $SS_{ef} = \|\mathbf{e}_f\|^2$ mit Freiheitsgraden $df_{ef} = n - \text{Rang}(\mathbf{X}_f)$, der Dimension des Fehlertraumes V_f^\perp (Abschn. 12.9.4). Als erwartungstreuer Schätzer $\hat{\sigma}^2$ dient in allen Tests $SS_{ef}/df_{ef} = \mathbf{y}^\top(\mathbf{I} - \mathbf{P}_f)\mathbf{y}/(n - \text{Rang}(\mathbf{X}_f))$ (Abschn. 12.9.5, Fußnote 62) – also auch dann, wenn das umfassendere Modell nicht mit dem vollständigen Modell übereinstimmt. Damit lassen sich lineare Nullhypothesen $\mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ im univariaten Fall mit der folgenden Teststatistik prüfen:

$$\begin{aligned} F &= \frac{\hat{\boldsymbol{\psi}}^\top(\mathbf{L}(\mathbf{X}_u^\top\mathbf{X}_u)^{-1}\mathbf{L}^\top)^{-1}\hat{\boldsymbol{\psi}}/\text{Rang}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}})}{\hat{\sigma}^2} = \frac{\hat{\boldsymbol{\psi}}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\hat{\boldsymbol{\psi}}/(u - r)}{\hat{\sigma}^2} \\ &= \frac{\mathbf{y}^\top\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{y}/(\text{Rang}(\mathbf{X}_u) - \text{Rang}(\mathbf{X}_r))}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_f)\mathbf{y}/(n - \text{Rang}(\mathbf{X}_f))} \end{aligned}$$

Unter H_0 ist F zentral F -verteilt mit $\text{Rang}(\mathbf{X}_u) - \text{Rang}(\mathbf{X}_r)$ Zähler- und $n - \text{Rang}(\mathbf{X}_f)$ Nenner-Freiheitsgraden. Das Modell einer Regression mit p Prädiktoren und absolutem Term hat $n - (p + 1)$ Fehler-Freiheitsgrade, das Modell einer einfaktorischen Varianzanalyse mit p Gruppen $n - p$.

Der Zähler der Teststatistik lässt sich äquivalent umformulieren, wobei explizit Bezug zum Modellvergleich genommen wird. Dafür sei \mathbf{e}_r der Vektor der Residuen des eingeschränkten Modells mit zugehöriger Projektionsmatrix \mathbf{P}_r und analog \mathbf{e}_u der Vektor der Residuen des umfassenderen Modells mit Projektionsmatrix \mathbf{P}_u . In der einfaktorischen Varianzanalyse sowie im Gesamt-Test der Regressionsanalyse sind das umfassendere Modell und das vollständige Modell identisch ($\mathbf{P}_u = \mathbf{P}_f$), ebenso ist dort das eingeschränkte Modell gleich dem Gesamt-Nullmodell ($\mathbf{P}_r = \mathbf{P}_0$).

Dann ist die Residual-Quadratsumme des eingeschränkten Modells $SS_{er} = \mathbf{y}^\top(\mathbf{I} - \mathbf{P}_r)\mathbf{y}$ mit Freiheitsgraden $df_{er} = n - \text{Rang}(\mathbf{X}_r)$ und die des umfassenderen Modells $SS_{eu} = \mathbf{y}^\top(\mathbf{I} - \mathbf{P}_u)\mathbf{y}$ mit Freiheitsgraden $df_{eu} = n - \text{Rang}(\mathbf{X}_u)$. Für die Differenz dieser Quadratsummen gilt $SS_{er} - SS_{eu} = \mathbf{y}^\top(\mathbf{P}_u - \mathbf{P}_r)\mathbf{y}$, wobei ihre Dimension $df_{er} - df_{eu} = \text{Rang}(\mathbf{X}_u) - \text{Rang}(\mathbf{X}_r)$ und $\mathbf{P}_u - \mathbf{P}_r = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$ ist. Damit lautet die Teststatistik:

$$\begin{aligned} F &= \frac{\mathbf{y}^\top(\mathbf{P}_u - \mathbf{P}_r)\mathbf{y}/(\text{Rang}(\mathbf{X}_u) - \text{Rang}(\mathbf{X}_r))}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_f)\mathbf{y}/(n - \text{Rang}(\mathbf{X}_f))} \\ &= \frac{(SS_{er} - SS_{eu})/(df_{er} - df_{eu})}{SS_{ef}/df_{ef}} \end{aligned}$$

⁶⁹Es sei $\mathbf{y}_r = \mathbf{X}_u\mathbf{r}$ aus V_r und $\mathbf{y}_a = \mathbf{A}^\top\mathbf{a}$ aus dem Erzeugnis der Spalten von \mathbf{A}^\top mit \mathbf{r} und \mathbf{a} als zugehörigen Koordinatenvektoren bzgl. \mathbf{X}_u und \mathbf{A}^\top . Mit $\mathbf{A}^\top = \mathbf{X}_u(\mathbf{X}_u^\top\mathbf{X}_u)^{-1}\mathbf{L}^\top$ liegt \mathbf{y}_a als Linearkombination der Spalten von \mathbf{X}_u in V_u . Da \mathbf{y}_r in V_r liegt, gilt $\mathbf{L}\mathbf{y}_r = \mathbf{0}$. Damit folgt $\mathbf{y}_a^\top\mathbf{y}_r = (\mathbf{A}^\top\mathbf{a})^\top\mathbf{X}_u\mathbf{r} = \mathbf{a}^\top\mathbf{A}\mathbf{X}_u\mathbf{r} = \mathbf{a}^\top\mathbf{L}(\mathbf{X}_u^\top\mathbf{X}_u)^{-1}\mathbf{X}_u^\top\mathbf{X}_u\mathbf{r} = \mathbf{a}^\top\mathbf{L}\mathbf{r} = \mathbf{a}^\top\mathbf{0} = 0$, d. h. $\mathbf{y}_a \perp \mathbf{y}_r$. Also liegt \mathbf{y}_a auch in V_r^\perp .

Multivariate Teststatistiken

Die Quadratsummen des univariaten Falls verallgemeinern sich im multivariaten Fall zu folgenden Matrizen, auf denen letztlich die Teststatistiken basieren (Fox et al., 2013):

- $\mathbf{T} = \mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_0) \mathbf{Y}$: Dies ist gleich der SSP-Matrix der Residuen des Gesamt-Nullmodells und damit gleich der SSP-Matrix der Gesamt-Daten.⁷⁰
- $\mathbf{W} = \mathbf{Y}^\top (\mathbf{I} - \mathbf{P}_f) \mathbf{Y}$: Dies ist gleich der SSP-Matrix der Residuen des vollständigen Modells (Fußnote 70), verallgemeinert also SS_{ef} .
- $\mathbf{B} = \mathbf{Y}^\top (\mathbf{P}_u - \mathbf{P}_r) \mathbf{Y}$: Dies ist gleich der SSP-Matrix der Vorhersagedifferenzen beider Modelle,⁷¹ verallgemeinert also $SS_{er} - SS_{eu}$.

In der einfaktoriellen Varianzanalyse handelt es sich bei den Matrizen um multivariate Verallgemeinerungen der Quadratsummen zwischen (*between*, \mathbf{B}) und innerhalb (*within*, \mathbf{W}) der Gruppen, sowie der totalen Quadratsumme $\mathbf{T} = \mathbf{B} + \mathbf{W}$. In der Diagonale dieser Gleichung findet sich für jede AV die zugehörige Quadratsummenzerlegung aus der univariaten Varianzanalyse wieder. \mathbf{B} ist hier gleichzeitig die SSP-Matrix der durch die zugehörigen Gruppenzentroide ersetzten Daten.

In der zweifaktoriellen Varianzanalyse ist \mathbf{W} die multivariate Verallgemeinerung der Residual-Quadratsumme. Es gibt nun für den Test der ersten UV eine Matrix \mathbf{B}_1 , für den Test der zweiten UV eine Matrix \mathbf{B}_2 und für den Test der Interaktion eine Matrix $\mathbf{B}_{1 \times 2}$ jeweils als multivariate Verallgemeinerung der zugehörigen Effekt-Quadratsumme. Mit diesen Matrizen gilt bei Quadratsummen vom Typ I dann analog zur univariaten Situation $\mathbf{T} = \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_{1 \times 2} + \mathbf{W}$ (Abschn. 7.5.2). Bei Quadratsummen vom Typ II und III gilt dies nur in orthogonalen Designs.

In der multivariaten multiplen Regression ist analog für den Test jedes Prädiktors j eine Matrix \mathbf{B}_j zu bilden, die sich aus der Differenzprojektion des zugehörigen Paares von eingeschränktem und umfassenderem Modell ergibt.

Auf Basis der Matrizen \mathbf{B} und \mathbf{W} berechnen sich die üblichen multivariaten Teststatistiken für lineare Hypothesen im ALM, die im univariaten Fall alle äquivalent zum oben aufgeführten F -Bruch sind.

- Wilks' Λ : $\frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})}$
- Roys Maximalwurzel: entweder der größte Eigenwert θ_1 von $(\mathbf{B} + \mathbf{W})^{-1} \mathbf{B}$ oder der größte Eigenwert λ_1 von $\mathbf{W}^{-1} \mathbf{B}$ (so definiert in R). Für die Umrechnung von θ_1 und λ_1 gilt $\lambda_1 = \frac{\theta_1}{1 - \theta_1}$ sowie $\theta_1 = \frac{\lambda_1}{1 + \lambda_1}$.
- Pillai-Bartlett-Spur: $\text{tr}((\mathbf{B} + \mathbf{W})^{-1} \mathbf{B})$
- Hotelling-Lawley-Spur: $\text{tr}(\mathbf{W}^{-1} \mathbf{B})$

⁷⁰ $\mathbf{X}_0 = \mathbf{1}$, daher ist die Matrix der Residuen $\mathbf{E} = (\mathbf{I} - \mathbf{P}_0) \mathbf{Y} = \mathbf{QY}$ zentriert (Abschn. 12.1.7, Fußnote 22) und $(\mathbf{QY})^\top (\mathbf{QY})$ deren SSP-Matrix. Als orthogonale Projektion ist \mathbf{Q} symmetrisch und idempotent, weshalb $(\mathbf{QY})^\top (\mathbf{QY}) = \mathbf{Y}^\top \mathbf{Q}^\top \mathbf{QY} = \mathbf{Y}^\top \mathbf{QY}$ gilt.

⁷¹ Zunächst ist die Matrix der Vorhersagedifferenzen $\hat{\mathbf{Y}}_u - \hat{\mathbf{Y}}_r = \mathbf{P}_u \mathbf{Y} - \mathbf{P}_r \mathbf{Y} = \mathbf{Y} - \mathbf{P}_r \mathbf{Y} - \mathbf{Y} + \mathbf{P}_u \mathbf{Y} = (\mathbf{I} - \mathbf{P}_r) \mathbf{Y} - (\mathbf{I} - \mathbf{P}_u) \mathbf{Y} = \mathbf{E}_r - \mathbf{E}_u$ gleich der Matrix der Differenzen der Residuen. Als Differenz zweier zentrierter Matrizen ist sie damit ihrerseits zentriert. Für die weitere Argumentation s. Fußnote 70.

12.9.8 Beispiel: Multivariate multiple Regression

Das Ergebnis der multivariaten multiplen Regression in Abschn. 12.5 lässt sich nun mit dem in Abschn. 12.9.6 und 12.9.7 dargestellten Vorgehen manuell prüfen. Im Beispiel sollen anhand der Prädiktoren Alter, Körpergröße und wöchentliche Dauer sportlicher Aktivitäten die Kriterien Körpergewicht und Gesundheit (i. S. eines geeigneten quantitativen Maßes) vorhergesagt werden. Zunächst sind die Designmatrix \mathbf{X} und die Projektion \mathbf{P}_f für das vollständige Regressionsmodell zu erstellen.⁷²

```
> Y <- cbind(weight, health)           # Matrix der Kriterien
> X <- cbind(1, height, age, sport)     # Designmatrix
> XR <- model.matrix(~ height + age + sport) # Designmatrix aus R
> all.equal(X, XR, check.attributes=FALSE) # Vergleich
[1] TRUE

> Xplus <- solve(t(X) %*% X) %*% t(X)   # X+
> B <- Xplus %*% Y                     # Parameterschätzungen
> Pf <- X %*% Xplus                    # Projektion (Hat-Matrix)
> Yhat <- Pf %*% Y                     # Vorhersage

# Kontrolle: vergleiche manuelle Berechnungen mit R
> fit <- lm(Y ~ height + age + sport)   # mult. Regr.-Modell
> all.equal(B, coef(fit), check.attributes=FALSE) # Parameter
[1] TRUE

> all.equal(Yhat, fitted(fit), check.attributes=FALSE) # Vorhersage
[1] TRUE
```

Im Gesamt-Nullmodell $E(\mathbf{Y}) = \mathbf{1}\beta_0^\top$ der multiplen Regression sind alle Parameter bis auf β_0 gleich 0, Designmatrix \mathbf{X}_0 ist der Vektor $\mathbf{1}$. Es folgt die Berechnung der zugehörigen Projektion \mathbf{P}_0 . Die Residuen ergeben sich aus der Projektion $\mathbf{I} - \mathbf{P}_0$ auf das orthogonale Komplement des von \mathbf{X}_0 aufgespannten Raumes.

```
# Gesamt-Nullmodell
> X0 <- X[, 1, drop=FALSE]             # Designmatrix X0 = 1-Vektor
> P0 <- X0 %*% solve(t(X0) %*% X0) %*% t(X0) # Projektion
> Id <- diag(N)                       # n x n Einheitsmatrix I
> WW <- t(Y) %*% (Id - Pf) %*% Y      # W
```

Schließlich muss für den Test jedes der drei Prädiktoren das zugehörige sequentielle Paar aus eingeschränktem H_0 -Modell und umfassenderem H_1 -Modell mit zugehörigen Designmatrizen \mathbf{X}_r und \mathbf{X}_u sowie ihren orthogonalen Projektionen \mathbf{P}_r und \mathbf{P}_u berechnet werden. Aus $\mathbf{P}_u - \mathbf{P}_r$ ergibt sich für jeden Prädiktor die Matrix \mathbf{B}_j , die in die zugehörigen Teststatistiken eingeht.

```
# Test Prädiktor 1: eingeschränktes Modell (= Gesamt-Nullmodell)
```

⁷²Der gewählte Weg zur Berechnung der Projektionsmatrizen soll die mathematischen Formeln direkt umsetzen, ist aber numerisch nicht stabil und weicht von in R-Funktionen implementierten Rechnungen ab (Bates, 2004).

```

> Xr1 <- X0 # Designmatrix
> Pr1 <- P0 # Projektion

# Test Prädiktor 1: umfassenderes Modell
> Xu1 <- X[ , c(1, 2)] # Designmatrix
> Pu1 <- Xu1 %>% solve(t(Xu1) %>% Xu1) %>% t(Xu1) # Projektion
> B1 <- t(Y) %>% (Pu1 - Pr1) %>% Y # Matrix B1

# Test Prädiktor 2: eingeschränktes Modell (= umfass. Modell Präd. 1)
> Xr2 <- Xu1 # Designmatrix
> Pr2 <- Pu1 # Projektion

# Test Prädiktor 2: umfassenderes Modell
> Xu2 <- X[ , c(1, 2, 3)] # Designmatrix
> Pu2 <- Xu2 %>% solve(t(Xu2) %>% Xu2) %>% t(Xu2) # Projektion
> B2 <- t(Y) %>% (Pu2 - Pr2) %>% Y # Matrix B2

# Test Prädiktor 3: eingeschränktes Modell (= umfass. Modell Präd. 2)
> Xr3 <- Xu2 # Designmatrix
> Pr3 <- Pu2 # Projektion

# Test Prädiktor 3: umfassenderes Modell (hier = vollst. Modell)
> Xu3 <- X # Designmatrix
> Pu3 <- Pf # Projektion
> B3 <- t(Y) %>% (Pu3 - Pr3) %>% Y # Matrix B3

```

Mit Hilfe der Matrizen B_j und W können die Teststatistiken für den Test jedes Prädiktors berechnet werden, was hier nur für den ersten Prädiktor gezeigt werden soll. Die Ergebnisse stimmen mit der Ausgabe von `summary(manova(...))` in Abschn. 12.5 überein.

```

> (WL1 <- det(WW) / det(B1 + WW)) # Wilks' Lambda
[1] 0.1104243

# Roys Maximalwurzel
> (RLR11 <- max(eigen(solve(WW) %>% B1)$values)) # lambda
[1] 8.055978

> (RLRt1 <- max(eigen(solve(B1 + WW) %>% B1)$values)) # theta
[1] 0.8895757

# Pillai-Bartlett-Spur: hier gleich Roys Maximalwurzel theta
> (PBT1 <- sum(diag(solve(B1 + WW) %>% B1)))
[1] 0.8895757

# Hotelling-Lawley-Spur: hier gleich Roys Maximalwurzel lambda
> (HLT1 <- sum(diag(solve(WW) %>% B1)))
[1] 8.055978

```

12.9.9 Beispiel: Einfaktorielle MANOVA

Das Ergebnis der einfaktoriellen MANOVA in Abschn. 12.7.1 lässt sich nun mit dem in Abschn. 12.9.6 und 12.9.7 dargestellten Vorgehen manuell prüfen. Im Beispiel sollen Daten von zwei AVn (Datenmatrix Y_{m1}) in drei Bedingungen (Faktor IV_{man}) vorliegen. Zunächst sind die Designmatrizen und Projektionen für das eingeschränkte H_0 -Modell sowie für das umfassendere H_1 -Modell zu erstellen.

```
# vollständiges Modell: ursprüngliche Inzidenzmatrix X*p
> XstarP <- cbind(as.numeric(IVman == 1),      # 1. Indikatorvariable
+               as.numeric(IVman == 2),      # 2. Indikatorvariable
+               as.numeric(IVman == 3))      # 3. Indikatorvariable

# vollständiges Modell für Treatment-Kontraste
> Ct <- contr.treatment(ncol(XstarP))         # Codiermatrix C
> Xpm1 <- XstarP %*% Ct                       # Xp-1 = X*p * C
> X <- cbind(1, Xpm1)                         # reduzierte Designmatrix [1|Xp-1]
> XR <- model.matrix(~ IVman)                 # Designmatrix aus R
> all.equal(X, XR, check.attributes=FALSE)   # Vergleich
[1] TRUE

# orthogonale Projektion auf durch Designmatrix aufgespannten Raum
> Pf <- X %*% solve(t(X) %*% X) %*% t(X)
> Pu <- Pf                                     # hier: H1-Modell = vollst. Modell, Pu = Pf

# Gesamt-Nullmodell: Designmatrix = 1-Vektor
> X0 <- X[, 1, drop=FALSE]

# orthogonale Projektion auf durch Gesamt-Nullmodell aufgespannten Raum
> P0 <- X0 %*% solve(t(X0) %*% X0) %*% t(X0)
> Pr <- P0                                     # hier: H0-Modell = Gesamt-0-Modell, Pr = P0
```

Die Kontrastschätzungen $\hat{B} = X^+Y$ bestehen bei den hier verwendeten Treatment-Kontrasten für jede AV l aus dem Mittelwert der ersten Gruppe ($\hat{\beta}_{0,l} = M_{1,l}$) sowie aus den Abweichungen der verbleibenden Gruppenmittel zu $M_{1,l}$ ($\hat{\beta}_{j,l} = M_{j,l} - M_{1,l}$ mit $j = 2, \dots, p$).

```
# Parameter- / Kontrastschätzungen für Treatment-Kontraste
> (Bt <- coef(lm(Ym1 ~ IVman, contrasts=list(IVman=contr.treatment))))
              [,1]  [,2]
(Intercept) -3.266667  5.60
IVman2       5.626667 -1.56
IVman3       2.916667 -6.00

# Kontrolle: teile Datenmatrix nach Gruppen auf und berechne Zentroide
> splDat <- split(data.frame(Ym1), IVman)
> (Mj <- t(sapply(splDat, colMeans)))
      X1      X2
1 -3.266667  5.60
```

```

2  2.360000  4.04
3 -0.350000 -0.40

> Mj[2, ] - Mj[1, ]          # Abweichung M2 - M1 = beta2
      X1      X2
5.626667 -1.560000

> Mj[3, ] - Mj[1, ]          # Abweichung M3 - M1 = beta3
      X1      X2
2.916667 -6.000000

```

Bei Treatment-Kontrasten stimmen die Schätzungen der ursprünglichen Parameter in $\hat{\mathbf{B}}_p^* = \mathbf{L}\hat{\mathbf{B}}_{p-1}$ für jede Zeile $j = 2, \dots, p$ mit jenen in $\hat{\mathbf{B}}$ überein, während durch die Nebenbedingung $\mathbf{v} = (1, 0, \dots, 0)^\top$ für die Parameter in der ersten Zeile $\hat{\beta}_{1,l}^* = 0$ gilt.

```

> (BstarPt <- Ct %% Bt[-1, ])      # B*p
      [,1] [,2]
1 0.000000 0.00
2 5.626667 -1.56
3 2.916667 -6.00

```

Bei der Effektcodierung bestehen die Kontrastschätzungen in $\hat{\mathbf{B}}$ für jede AV l aus dem ungewichteten Gesamtmittelwert ($\hat{\beta}_{0,l} = M_l = \frac{1}{p} \sum_j M_{j,l}$) sowie aus den Abweichungen der ersten $p - 1$ Gruppenmittel zu M_l ($\hat{\beta}_{j,l} = M_{j,l} - M_l$ mit $j = 1, \dots, p - 1$).

```

# Parameter- / Kontrastschätzungen für Effektcodierung
> (Be <- coef(lm(Ym1 ~ IVman, contrasts=list(IVman=contr.sum))))
      [,1] [,2]
(Intercept) -0.4188889 3.08
IVman1      -2.8477778 2.52
IVman2       2.7788889 0.96

> colMeans(Mj)          # ungew. Gesamtmittel = beta0
      X1      X2
-0.4188889 3.0800000

> scale(Mj, colMeans(Mj), scale=FALSE) # Mj.l - Ml = beta1, beta2
      X1      X2
1 -2.84777778 2.52
2  2.77888889 0.96
3  0.06888889 -3.48

```

Die Schätzungen der ursprünglichen Parameter in $\hat{\mathbf{B}}_p^*$ stimmen bei der Effektcodierung für jede Zeile $j = 1, \dots, p - 1$ mit jenen in $\hat{\mathbf{B}}$ überein. Die Nebenbedingung $\mathbf{v} = \mathbf{1}$ legt für jede AV l die $\hat{\beta}_{p,l}^*$ dadurch fest, dass die Beziehung $\sum_{j=1}^p \hat{\beta}_{j,l}^* = 0$ gelten muss, was zu $\hat{\beta}_{p,l}^* = -\sum_{j=1}^{p-1} \hat{\beta}_{j,l}^*$ führt.

```

> Ce <- contr.sum(ncol(XstarP))          # Codiermatrix Effektcodierung
> (BstarPe <- Ce %*% betaE[-1 , ])      # B*p
      [,1] [,2]
1 -2.84777778  2.52
2  2.77888889  0.96
3  0.06888889 -3.48

> colSums(BstarPe)                      # Summe über beta*p = 0
[1] 0.000000e+00 1.110223e-16

```

Bei der Parametrisierung ohne β_0 in der Rolle des theoretischen Zentroids $\boldsymbol{\mu}$ (cell means Modell, Abschn. 12.9.2, Fußnote 49) erhalten die ursprünglichen Parameter in \mathbf{B}_p^* die Bedeutung der Gruppenzentroide $\boldsymbol{\mu}_j$ und werden entsprechend über die Gruppenmittelwerte geschätzt. Für jede AV l gilt also $\hat{\beta}_{j,l}^* = M_{j,l}$, zudem stimmt \mathbf{B}_p^* mit \mathbf{B}_p überein.

```

# Parameterschätzungen B = B* für cell means Modell
> (Bcm <- coef(lm(Ym1 ~ IVman - 1)))
      [,1] [,2]
IVman1 -3.266667  5.60
IVman2  2.360000  4.04
IVman3 -0.350000 -0.40

```

Die Vorhersage $\hat{\mathbf{Y}}_f = \mathbf{P}_f \mathbf{Y}$ des vollständigen Modells liefert für eine Person in der Gruppe j für jede AV l den zugehörigen Gruppenmittelwert $M_{j,l}$.

```

> Yhat <- Pf %*% Ym1                    # Vorhersage vollst. Modell
> unique(Yhat)                           # vorkommende Werte
      [,1] [,2]
[1,] -3.266667  5.60
[2,]  2.360000  4.04
[3,] -0.350000 -0.40

```

Die für den inferenzstatistischen Test notwendigen Matrizen \mathbf{T} , \mathbf{B} und \mathbf{W} ergeben sich aus den ermittelten Projektionen $\mathbf{P}_r = \mathbf{P}_0$ und $\mathbf{P}_u = \mathbf{P}_f$ jeweils auf den Unterraum, der durch die Designmatrix des eingeschränkten und des umfassenderen Modells aufgespannt wird.

```

> Id <- diag(sum(Nj))                   # n x n Einheitsmatrix I
> BB <- t(Ym1) %*% (Pu-Pr) %*% Ym1     # B
> WW <- t(Ym1) %*% (Id-Pf) %*% Ym1     # W
> TT <- t(Ym1) %*% (Id-P0) %*% Ym1     # T

```

```

# B: SSP-Matrix der durch Vorhersagedifferenzen ersetzten Daten
> all.equal(BB, (sum(Nj)-1) * cov((Pu-Pr) %*% Ym1))
[1] TRUE

```

```

# W: SSP-Matrix der durch Differenz zum jeweiligen
# Gruppenzentroid ersetzten Daten
> all.equal(WW, (sum(Nj)-1) * cov((Id-Pf) %*% Ym1))

```

```
[1] TRUE

# T: SSP-Matrix der Residuen des Nullmodells sowie der Daten
> all.equal(TT, (sum(Nj)-1) * cov((Id-P0) %*% Ym1))
[1] TRUE

> all.equal(TT, (sum(Nj)-1) * cov(Ym1))          # SSP-Matrix Daten
[1] TRUE

> all.equal(TT, BB + WW)                        # Kontrolle: T = B + W
[1] TRUE
```

Mit Hilfe der Matrizen B und W können die Teststatistiken berechnet werden, die mit der Ausgabe von `summary(manova(...))` in Abschn. 12.7.1 übereinstimmen.

```
> (WL <- det(WW) / det(BB + WW))                # Wilks' Lambda
[1] 0.4201068

# Roys Maximalwurzel
> (RLRl <- max(eigen(solve(WW) %*% BB)$values))  # lambda
[1] 0.6956387

> (RLRt <- max(eigen(solve(BB + WW) %*% BB)$values)) # theta
[1] 0.4102517

> (PBT <- sum(diag(solve(BB + WW) %*% BB)))      # Pillai-Bartlett-Spur
[1] 0.6979024

> (HLT <- sum(diag(solve(WW) %*% BB)))          # Hotelling-Lawley-Spur
[1] 1.099444
```

12.9.10 Beispiel: Zweifaktorielle MANOVA

Das Ergebnis der zweifaktoriellen MANOVA in Abschn. 12.7.2 lässt sich nun ebenfalls mit dem in Abschn. 12.9.6 und 12.9.7 dargestellten Vorgehen manuell prüfen. Es sollen Daten auf zwei AVn (Datenmatrix $Ym2$) in 3×2 Bedingungen (Faktoren IV1 und IV2) vorliegen. Zunächst sind die ursprünglichen Inzidenzmatrizen X_1^* , X_2^* , $X_{1 \times 2}^*$ zu erstellen. Ihr jeweiliges Produkt mit der zugehörigen Codiermatrix C geht in die Designmatrizen der eingeschränkten und umfassenderen Modelle ein, die den Tests der drei Hypothesen (zwei Haupteffekte und Interaktionseffekt) zugrunde liegen.

```
# ursprüngliche Inzidenzmatrizen
> Xstar1 <- cbind(as.numeric(IV1 == 1),          # UV 1: X*1
+               as.numeric(IV1 == 2),
+               as.numeric(IV1 == 3))

> Xstar2 <- cbind(as.numeric(IV2 == 1),          # UV 2: X*2
```

```

+           as.numeric(IV2 == 2))

# Interaktion: alle paarweisen Produkte der Spalten von X*1 und X*2: X*1x2
> Xstar12 <- cbind(as.numeric(IV1 == 1) * as.numeric(IV2 == 1),
+                 as.numeric(IV1 == 2) * as.numeric(IV2 == 1),
+                 as.numeric(IV1 == 3) * as.numeric(IV2 == 1),
+                 as.numeric(IV1 == 1) * as.numeric(IV2 == 2),
+                 as.numeric(IV1 == 2) * as.numeric(IV2 == 2),
+                 as.numeric(IV1 == 3) * as.numeric(IV2 == 2))

# Codiermatrizen: Treatment-Kontraste
> C1 <- contr.treatment(ncol(Xstar1))      # UV 1
> C2 <- contr.treatment(ncol(Xstar2))      # UV 2
> C12 <- kronecker(C2, C1)                # Interaktion

# reduzierte Inzidenzmatrizen
> X1 <- Xstar1  %% C1                      # UV 1: X1 = X*1 * C1
> X2 <- Xstar2  %% C2                      # UV 2: X2 = X*2 * C2
> X12 <- Xstar12 %% C12                    # Interaktion: X*1x2 * C1x2

# vollständiges Modell: Designmatrix für identifizierbare Parameter
> X <- cbind(1, X1, X2, X12)               # X = [1|X1|X2|X1x2]
> XR <- model.matrix(~ IV1*IV2)           # Designmatrix aus R
> all.equal(X, XR, check.attributes=FALSE) # Vergleich
[1] TRUE

Im Gesamt-Nullmodell der zweifaktoriellen Varianzanalyse sind alle Parameter bis auf  $\beta_0$  gleich 0, als Designmatrix  $\mathbf{X}_0$  bleibt also der Vektor  $\mathbf{1}$ . Es folgt die Berechnung der zugehörigen Projektion  $\mathbf{P}_0$  sowie der zum vollständigen Modell mit Designmatrix  $\mathbf{X}_f$  gehörenden Projektion  $\mathbf{P}_f$ . Aus der Projektion auf das orthogonale Komplement des jeweils von den Spalten von  $\mathbf{X}_0$  und  $\mathbf{X}_f$  aufgespannten Unterraumes ergeben sich die Matrizen  $\mathbf{T}$  und  $\mathbf{W}$ .

# Gesamt-Nullmodell
> X0 <- X[, 1, drop=FALSE]                # Designmatrix X0 = 1-Vektor

# Nullmodell: orthogonale Projektion auf durch X0 aufgespannten Raum
> P0 <- X0 %% solve(t(X0) %% X0) %% t(X0)

# vollst. Modell: orth. Proj. auf durch Designmatrix aufgespannten Raum
> Pf <- X %% solve(t(X) %% X) %% t(X)

# Matrizen T und W für Teststatistiken
> Id <- diag(nrow(Ym2))                   # n x n Einheitsmatrix I
> TT <- t(Ym2) %% (Id - P0) %% Ym2        # T
> WW <- t(Ym2) %% (Id - Pf) %% Ym2        # W

```

Die Vorhersage $\hat{\mathbf{Y}}_f = \mathbf{P}_f \mathbf{Y}$ des vollständigen Modells liefert für eine Person in der Bedingungs-

kombination jk beider UVn für jede AV l den Gruppenmittelwert $M_{jk.l}$.

```
> Yhat <- Pf %*% Ym2           # Vorhersage vollständiges Modell
> unique(Yhat)                # vorkommende Werte
      [,1]      [,2]
[1,] -3.2666667  5.600000
[2,]  2.3600000  4.040000
[3,] -0.3500000 -0.400000
[4,]  0.0666667  4.666667
[5,]  3.6800000  7.960000
[6,]  4.0500000 -0.400000

# Kontrolle: teile Datenmatrix nach Gruppen auf und berechne Zentroide
> splDat <- split(data.frame(Ym2), list(IV1, IV2))
> t(sapply(splDat, colMeans))
      X1      X2
1.1 -3.2666667  5.600000
2.1  2.3600000  4.040000
3.1 -0.3500000 -0.400000
1.2  0.0666667  4.666667
2.2  3.6800000  7.960000
3.2  4.0500000 -0.400000
```

Schließlich muss für jede der drei Hypothesen das zugehörige Paar aus eingeschränktem H_0 -Modell und umfassenderem H_1 -Modell mit zugehörigen Designmatrizen \mathbf{X}_r und \mathbf{X}_u sowie ihren orthogonalen Projektionen \mathbf{P}_r und \mathbf{P}_u berechnet werden. Dies soll hier für Quadratsummen vom Typ I geschehen. Aus der Differenz beider Projektionen ergibt sich für jeden Effekt die Matrix \mathbf{B}_j , die als Verallgemeinerung der univariaten Effekt-Quadratsumme in die zugehörigen Teststatistiken eingeht.

```
# Test UV 1: eingeschränktes Modell (= Gesamt-Nullmodell)
> Xr1 <- X0                    # Designmatrix
> Pr1 <- P0                    # Projektion

# Test UV 1: umfassenderes Modell
> Xu1 <- X[ , c(1, 2, 3)]      # Designmatrix
> Pu1 <- Xu1 %*% solve(t(Xu1) %*% Xu1) %*% t(Xu1) # Projektion
> B1 <- t(Ym2) %*% (Pu1-Pr1) %*% Ym2           # Matrix B1

# Test UV 2: eingeschränktes Modell (= umfassenderes Modell UV 1)
> Xr2 <- Xu1                    # Designmatrix
> Pr2 <- Pu1                    # Projektion

# Test UV 2: umfassenderes Modell
> Xu2 <- X[ , c(1, 2, 3, 4)]    # Designmatrix
> Pu2 <- Xu2 %*% solve(t(Xu2) %*% Xu2) %*% t(Xu2) # Projektion
> B2 <- t(Ym2) %*% (Pu2-Pr2) %*% Ym2           # Matrix B2
```

```

# Test Interaktion: eingeschränktes Modell (= umfass. Modell UV 2)
> Xr12 <- Xu2                                # Designmatrix
> Pr12 <- Pu2                                # Projektion

# Test Interaktion: umfassenderes Modell (hier = vollst. Modell)
> Xu12 <- X                                  # Designmatrix
> Pu12 <- Pf                                 # Projektion
> B12 <- t(Ym2) %*% (Pu12 - Pr12) %*% Ym2    # Matrix B12

# Kontrolle: Verallgemeinerung der Quadratsummenzerlegung
# gilt allgemein nur für QS vom Typ I, sonst nur bei gleichen Njk
> all.equal(TT, B1 + B2 + B12 + WW)
[1] TRUE

Mit Hilfe der Matrizen  $B_1$ ,  $B_2$ ,  $B_{1 \times 2}$  und  $W$  können die Teststatistiken für den Test jedes Effekts berechnet werden, was hier nur für den ersten Haupteffekt gezeigt werden soll. Die Ausgabe stimmt mit jener von summary(manova(...)) in Abschn. 12.7.2 überein.

> (WL1 <- det(WW) / det(B1 + WW))            # Wilks' Lambda
[1] 0.3846785

# Roys Maximalwurzel
> (RLR11 <- max(eigen(solve(WW) %*% B1)$values)) # lambda
[1] 1.042646

> (RLRt1 <- max(eigen(solve(B1 + WW) %*% B1)$values)) # theta
[1] 0.5104389

> (PBT1 <- sum(diag(solve(B1 + WW) %*% B1)))    # Pillai-Bartlett-Spur
[1] 0.7246769

> (HLT1 <- sum(diag(solve(WW) %*% B1)))        # Hotelling-Lawley-Spur
[1] 1.315296

```