5.6 Vergleich k verbundener Stichproben

2-faktorielle Varianzanalyse

Modell:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim (0, \sigma^2)$$

$i = 1, \ldots, a, j = 1, \ldots, b$. (eine Beobachtung je Zelle)

Das Modell ist überparametrisiert,
Bedingung: $\sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0$.
Folgene Hypothesen sind zu testen:

$H_{0a}: \quad \alpha_1 = \cdots = \alpha_a = 0$ gegen
$H_{1a}: \quad \exists (i_1, i_2) : \alpha_{i_1} \neq \alpha_{i_2}$
$H_{0b}: \quad \beta_1 = \cdots = \beta_b = 0$ gegen
$H_{1a}: \quad \exists (j_1, j_2) : \beta_{j_1} \neq \beta_{j_2}$

`GLM_Synchro` `GLM_Cache`

2-faktorielle Varianzanalyse

$$
\begin{aligned}
\overline{Y}_{..} &= \frac{1}{a \cdot b} \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij} \quad \text{arith. Mittel aller Beob.} \\
\overline{Y}_{i.} &= \frac{1}{b} \sum_{j=1}^{b} Y_{ij} \quad \text{Mittel der i-ten Stufe von A} \\
\overline{Y}_{.j} &= \frac{1}{a} \sum_{i=1}^{a} Y_{ij} \quad \text{Mittel der j-ten Stufe von B} \\
SSA &:= b \sum_{i=1}^{a} (\overline{Y}_{i.} - \overline{Y}_{..})^2 \qquad SSB := a \sum_{j=1}^{b} (\overline{Y}_{.j} - \overline{Y}_{..})^2 \\
SSE &:= \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2 \\
SST &:= \sum_{i=1}^{a} \sum_{j=1}^{b} (Y_{ij} - \overline{Y}_{..})^2
\end{aligned}
$$

2-faktorielle Varianzanalyse

Quadratsummenzerlegung

Dependent Variable: Y

Source	DF	Sum Squ.	Mean Squ.	F-value	$Pr > F$ p-value
A	a-1	SSA	MSA	$\frac{MSA}{MSE}$	H_{1a}
B	b-1	SSB	MSB	$\frac{MSB}{MSE}$	H_{1b}
Model	a+b-2	SSM	MSM	$\frac{MSM}{MSE}$	H_1
Error	(a-1)(b-1)	SSE	MSE		
Total	a b - 1	SST			

$$SSM = SSA + SSB \qquad SST = SSA + SSB + SSE$$

$$MSA = \frac{SSA}{(a-1)} \qquad MSB = \frac{SSB}{(b-1)}$$

$$MSM = \frac{SSA + SSB}{a+b-2} \qquad MSE = \frac{SSE}{(a-1)(b-1)}$$

2-faktorielle Varianzanalyse

Tests

H_{0a} gegen H_{1a}:

$$F_1 = \frac{MSA}{MSE} = \frac{\text{mittl. Var. zwischen Stufen von A}}{\text{mittl. Var. innerhalb d. Gruppen}}$$

$$F_1 \quad \sim \quad F_{a-1,(a-1)(b-1)}$$

H_{0b} gegen H_{1b}:

$$F_2 = \frac{MSB}{MSE} = \frac{\text{mittl. Var. zwischen Stufen von B}}{\text{mittl. Var. innerhalb d. Gruppen}}$$

$$F_2 \quad \sim \quad F_{b-1,(a-1)(b-1)}$$

große Werte von F führen zur Ablehnung!

$$F_1 \quad > \quad F_{1-\alpha,a-1,(a-1)(b-1)} \rightarrow \text{Ablehnung von } H_{0a}$$
$$F_2 \quad > \quad F_{1-\alpha,b-1,(a-1)(b-1)} \rightarrow \text{Ablehnung von } H_{0b}$$

2-faktorielle Varianzanalyse
Tests

H_0: $\alpha_1 = \cdots \alpha_a = 0$ und $\beta_1 = \cdots \beta_a = 0$ gegen
H_1: $\exists(i_1, i_2)$: $\alpha_{i_1} \neq \alpha_{i_2} \vee \exists(j_1, j_2)$: $\beta_{j_1} \neq \beta_{j_2}$.

$$F = \frac{MSModell}{MSE} = \frac{SSA + SSB}{SSE} \cdot \frac{(a-1)(b-1)}{a+b-2}$$

$$\begin{aligned} MSModell &= \frac{SSModell}{a+b-2} \\ SSModell &= SSA + SSB. \end{aligned}$$

H_0 ablehnen, falls

$$F > F_{1-\alpha, a+b-2, (a-1)(b-1)}.$$

Zweifaktorielle Varianzanalyse

```
PROC GLM;
  CLASS A B; /*die beiden Faktoren*/
  MODEL Y = A B;
RUN;
```

Output

- Balanzierter Fall: Variante I und III identisch
- Unbalanzierter Fall: Typ III-Summen sind vorzuziehen, da der entsprechende Test unabhängig von den Stichprobenumfängen ist.

5.7 Weitere Varianzanalyse-Modelle

5.7.1 Mehrere Beobachtungen pro Kombination der Faktoren A und B

SAS-Prozedur ändert sich nicht!

Output ändert sich gegebenenfalls
a) balanzierter Fall $\rightarrow$ eindeutig
b) unbalanzierter Fall $\rightarrow$
Es gibt verschiedene Möglichkeiten die Fehlerquadratsummen zu zerlegen.
$\rightarrow$ SAS bietet die Varianten an

3 Forscher graben eine Reihe von Schädeln in 3 verschiedenen Schichten aus. Gemessen wird die Nasenlänge.
? Forschereffekt, Schichteneffekt

Weitere Varianzanalyse-Modelle

Mehrere Beobachtungen pro Kombination der Faktoren A und B

Klinische Untersuchung in mehreren Zentren

Ein Medikament zur Gewichtsreduktion soll getestet werden.

- 1: Medikament
- 0: Placebo
- 1-6: Zentren

Modell:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

Es interessiert nur das Medikament, nicht das Zentrum:

$$H_0: \quad \alpha_0 = \alpha_1 \qquad H_1: \alpha_0 < \alpha_1$$

Weitere Varianzanalyse-Modelle

```
PROC GLM;
   CLASS Medik Zentrum; /*die beiden Faktoren*/
   MODEL Y = Medik Zentrum;
RUN; (dieselbe Prozedur wie oben)
```

```
GLM_Drugeffect
```

Zum Output: wie bisher.

Balanzierter Fall: Variante I und III identisch.

Unbalanzierter Fall: Typ III-Summen zu bevorzugen, da der entsprechende Test unabhängig von den Stichprobenumfängen ist.

Weitere Varianzanalyse-Modelle

5.7.2 Wechselwirkungen ins Modell mit aufnehmen

$$Y_{ijk} = \alpha + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

(+Reparametrisierungsbedingungen)
geht nur, wenn für jede Faktorstufenkombination mehrere Beobachtungen vorliegen.

```
PROC GLM;
   CLASS A B; /*die beiden Faktoren*/
   MODEL Y = A B A*B;
RUN;
```

`GLM_Insekten`

Weitere Varianzanalyse-Modelle
Modell mit Wechselwirkungen

Folgene Hypothesen sind zu testen:

$H_{0a}:\quad \alpha_1 = \cdots = \alpha_a = 0$ gegen
$H_{1a}:\quad \exists(i_1, i_2) : \alpha_{i_1} \neq \alpha_{i_2}$

$H_{0b}:\quad \beta_1 = \cdots = \beta_b = 0$ gegen
$H_{1a}:\quad \exists(j_1, j_2) : \beta_{j_1} \neq \beta_{j_2}$

$H_{0c}:\quad \gamma_{11} = \cdots = \gamma_{a*b} = 0$ gegen
$H_{1c}:\quad \exists(j_1, j_2) : \gamma_{j_1, j_2} \neq 0$

Weitere Varianzanalyse-Modelle

5.7.3 Faktoren (Effekte, Faktorstufen) sind zufällig

hier ist Schätzung der Varianzkomponenten interessant und evtl. ein Hypothesentest

Preisrichter seien zufällig ausgewählt.

Die Frage ist, ob die Variabilität in den Scores an den Preisrichtern liegt?

$$Y_{ij} = \mu + \underbrace{A_i}_{\text{zufällig}} + b_j + \epsilon_{ij}$$

$$A_i \sim (0, \sigma_P^2)$$

$$\epsilon_{ij} \sim (0, \sigma^2)$$

Varianzkomponentenschätzung

```
PROC VARCOMP METHOD=Type1;
  CLASS Preisrichter Wettkaempfer;
  MODEL Score = Preisricher;
RUN;
```

GLM_syncro_zufaelligeEffekte

METHOD=Type1: Auf den Quadratsummen beruhende Varianzschätzungen

Annahme: A_i, B_j und ϵ_{ij} unabhängig.

$$var(Y_{ij}) = var(A_i) + var(B_j) + var(\epsilon_{ij})$$

Output: Schätzungen für die Varianzkomponenten.

Weitere Varianzanalyse-Modelle

7.4 Mehr als 2 Faktoren

- höherfaktorielle VA

Frequenzspektren

Gemessen wird die Amplitude bei 35 verschiedenen Frequenzen, 4 Füllungen, 3 Richtungen,
jede Messung wird 5 mal wiederholt.
? Füllungs-, Richtungseffekt, Wiederholungseffekt? Frequenzeffekt?
$\longrightarrow$ 4 Faktoren.

```
PROC GLM;
    CLASS A B C D;
    MODEL Y = A B C D; RUN;
```

~/Beratung/Vogt/Glaeser1

Weitere Varianzanalyse-Modelle

7.5 Hierarchische Modelle

Die Faktoren liegen in hierarch. Ordnung vor.

					A						
		A1			A2			A3			A4
B	B11	B12	B13	B21	B22	B23	B31	B32	B33	B41	B4

(mit zufäll. Effekten)

Kalzium-Gehalt verschiedener Pflanzen und von verschiedenen Blättern
4 Pflanzen werden zufällig ausgewählt
3 Blätter davon
2 Stichproben zu 100mg von jedem Blatt
Frage: Gibt es zwischen Pflanzen oder zwischen Blättern unterschiedliche CA-Konzentrationen?

Weitere Varianzanalyse-Modelle

Hierarchische Modelle (2)

Modell

$$Y_{ijk} = \mu + A_i + B_{ij} + \epsilon_{ijk}$$

$$A_i \sim \mathcal{N}(0, \sigma_a^2) \quad B_{ij} \sim \mathcal{N}(0, \sigma_b^2) \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

hier: n=2 a=4 b=3

$$\begin{aligned} \text{var} Y_{ijk} &= \text{var} A_i + \text{var} B_{ij} + \text{var} \epsilon_{ijk} \\ &= \sigma_a^2 + \sigma_b^2 + \sigma^2 \end{aligned}$$

$H_{0a} : \sigma_a^2 = 0 \qquad H_{0b} : \sigma_b^2 = 0$

`GLM_hierarch`

Weitere Varianzanalyse-Modelle

Hierarchische Modelle (3)

```
PROC GLM;
   CLASS A B;
   MODEL Y = A B(A);  hierarch. Struktur*
   RANDOM A B(A);  Faktoren sind zufaellig*
RUN;

PROC VARCOMP;
    CLASS A B;
    MODEL Y=A B(A);
RUN;
```