Vorlesung Statistik II

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sommersemester 2026

Einfaktorielle Varianzanalyse mit Messwiederholung

Für eine Anwendung der Varianzanalyse auf Messwiederholungs-Designs (Within-Subjects-Designs) werden die Versuchspersonen als Faktorstufen eines zusätzlichen Faktors (P) angesehen

- Da die Versuchspersonen in der Regel zufällig ausgewählt werden, handelt es sich hierbei um einen Faktor mit zufälligen Effekten

Der eigentlich interessierende Faktor, der die unterschiedlichen Versuchsbedingungen repräsentiert (Treatment, Wiederholungsfaktor), wird im Folgenden als Faktor mit festen Effekten konzeptualisert

- Die Kombination zufälliger und fester Effekte führt daher zu einer varianzanalytischen Auswertung der Daten aus Within-Subjects-Designs im Rahmen eines gemischten Modells

Datenschema

Faktor P	Faktor B				
	1	$\dots$	j	$\dots$	J
Vp 1	y_{11}	$\dots$	y_{1j}	$\dots$	y_{1J}
$\vdots$					
Vp i	y_{i1}	$\dots$	y_{ij}	$\dots$	y_{iJ}
$\vdots$					
Vp I	y_{I1}	$\dots$	y_{Ij}	$\dots$	y_{IJ}

Modell

Der Faktor P repräsentiert zufällige Effekte und der Faktor B feste Effekte

In Effektdarstellung lautet das Modell

$$Y_{ij} = \mu + \pi_i + \beta_j + (\pi\beta)_{ij} \quad \text{mit} \quad i = 1, \ldots, I, \quad j = 1, \ldots, J$$

Voraussetzungen

- Die Zufallsvariablen π_i und $(\pi\beta)_{ij}$ sind verteilt gemäß $\pi_i \sim N(0, \sigma_P^2)$ bzw. $(\pi\beta)_{ij} \sim N(0, \sigma_{PB}^2)$
- Die π_i, $(\pi\beta)_{ij}$ sind paarweise stochastisch unabhängig
- $\sum_j \beta_j = 0$

Streuungszerlegung

Als Zerlegung der Gesamtsumme der Abweichungsquadrate SS_T ergibt sich

$$SS_T = \underbrace{SS_P}_{SS_{\text{between subjects}}} + \underbrace{SS_B + SS_{P\times B}}_{SS_{\text{within subjects}}}$$

Man hat also folgende Interpretation der entsprechenden Summen der Abweichungsquadrate

SS_P	Haupteffekte Versuchspersonen (zufälliger Faktor)
SS_B	Haupteffekte Wiederholungsfaktor (fester Faktor, Treatment)
$SS_{P\times B}$	Interaktion Versuchspersonen $\times$ Treatment

Mittlere Quadrate

Damit erhält man die mittleren quadratischen Abweichungen und die Erwartungswerte

$$MS_P = \frac{SS_P}{I-1} \qquad E(MS_P) = \sigma^2_{PB} + J\,\sigma^2_P$$

$$MS_B = \frac{SS_B}{J-1} \qquad E(MS_B) = \sigma^2_{PB} + \frac{I\,\sum_{j=1}^{J} \beta_j^2}{J-1}$$

$$MS_{P\times B} = \frac{SS_{P\times B}}{(I-1)(J-1)} \qquad E(MS_{P\times B}) = \sigma^2_{PB}$$

Unter Nullhypothese $H_0\colon\ \beta_j = 0$ für alle $j = 1, \ldots, J$ gilt dann

$$E(MS_B) = \sigma^2_{PB}$$

Teststatistik

Unter H_0 sind damit MS_B und $MS_{P\times B}$ voneinander stochastisch unabhängige und erwartungstreue Schätzfunktionen für σ^2_{PB}

Für $MS_{P\times B}$ gilt dies auch unter der Alternativhypothese H_1: $\beta_j \neq 0$ für mindestens ein $j \in \{1, \ldots, J\}$

Somit eignet sich auch hier der Quotient

$$\frac{MS_B}{MS_{P\times B}}$$

als Teststatistik

Einfaktorielle Varianzanalyse mit Messwiederholung

Hypothesen

$$H_0\colon\ \beta_j = 0 \text{ für alle } j = 1, \ldots, J, \quad H_1\colon\ \beta_j \neq 0 \text{ für mindestens ein } j$$

Teststatistik

$$F = \frac{MS_B}{MS_{P\times B}}$$

Verteilung unter H_0

$$F \sim F(J-1,\ (I-1)(J-1))$$

Ablehnungsbereich

$$F > F_{1-\alpha}(J-1,\ (I-1)(J-1))$$

Zufälliger Personenfaktor

Ein Test der Effekte der Personen ist meist nicht interessant

- Diese Effekte reflektieren Leistungsunterschiede zwischen den Versuchspersonen, die unabhängig von den zu untersuchenden Effekten der Versuchsbedingungen bestehen

Man ist ausschließlich an den Effekten der Versuchsbedingungen (Treatment) innerhalb der einzelnen Versuchspersonen interessiert

Der Faktor "Versuchsperson" wird lediglich als Hilfskonstruktion eingeführt und repräsentiert aus Sicht des Designs eine Störgröße (im Englischen "nuisance factor")

Beispiel

Jede von I zufällig ausgewählten Versuchspersonen soll vier verschiedene Versuchsbedingungen durchlaufen

$$B_1, B_2, B_3, B_4$$

Diese sollen jeder einzelnen Versuchsperson in einer per Zufall bestimmten Reihenfolge vorgegeben werden (Randomisierung als Methode des Ausbalanzierens)

In der folgenden Poweranalyse sind die Effekte relativ zur Referenzkategorie zu interpretieren (*Dummykodierung*), dabei bezieht sich `mu` auf den Erwartungswert der Baseline-Bedingung B_1

Bestimmung des Stichprobenumfangs $n = I$, um mit ca. 80% Power ($\alpha = 0.05$) bei $\sigma_P = 4$ und $\sigma_{PB} = 6$ die Effekte $\beta_2 = 2$, $\beta_3 = 4$ und $\beta_4 = 6$ zeigen zu können, H_0: $\beta_j = 0$ für $j = 1, \ldots, 4$

Teststärke und Versuchsplanung

```
n <- 22                          # Anzahl der Personen (= I)
dat <- expand.grid(treat = factor(1:4),
                   subj = factor(1:n))
beta <- c(mu = 15, b2 = 2, b3 = 4, b4 = 6)                # b1 := 0
means <- model.matrix(~ treat, dat) %*% beta

pval <- replicate(2000, {
  y <- means +                   # y = mu + p + b + e
       model.matrix(~ 0 + subj, dat) %*% rnorm(n, sd = 4) +
       rnorm(4*n, sd = 6)
  m <- aov(y ~ treat + Error(subj), dat)
  summary(m)$"Error: Within"[[1]]$"Pr(>F)"[1]
})
mean(pval < 0.05)
```

Beipieldaten

Für diese Untersuchung resultiert ein Datenschema mit dem Faktor "Versuchspersonen" (10 Faktorstufen) und dem Treatment-Faktor B (4 Faktorstufen)

	B_1	B_2	B_3	B_4
Vp 1	32	23	16	30
Vp 2	19	20	27	32
Vp 3	16	14	23	16
Vp 4	32	26	16	32
Vp 5	21	12	14	33
Vp 6	19	19	16	29
Vp 7	20	32	12	29
Vp 8	19	19	27	27
Vp 9	23	16	31	34
Vp 10	27	26	25	35

Beispiel

Grand Mean

$$\hat{\mu} = \bar{y}_{..} = 23.475$$

Haupteffekte des Faktors B

$$\begin{aligned}
\hat{\beta}_1 &= \bar{y}_{.1} - \bar{y}_{..} = 22.8 - 23.475 = -0.675 \\
\hat{\beta}_2 &= \bar{y}_{.2} - \bar{y}_{..} = 20.7 - 23.475 = -2.775 \\
\hat{\beta}_3 &= \bar{y}_{.3} - \bar{y}_{..} = 20.7 - 23.475 = -2.775 \\
\hat{\beta}_4 &= \bar{y}_{.4} - \bar{y}_{..} = 29.7 - 23.475 = 6.225
\end{aligned}$$

Beispiel

Haupteffekte des Faktors P

$$\begin{aligned} \hat{\pi}_1 &= \hat{y}_{1.} - \bar{y}_{..} = 25.25 - 23.475 = 1.775 \\ \hat{\pi}_2 &= \hat{y}_{2.} - \bar{y}_{..} = 24.50 - 23.475 = 1.025 \end{aligned}$$

und entsprechend

$$\begin{aligned} \hat{\pi}_3 &= -6.225, & \hat{\pi}_4 &= 3.025 \\ \hat{\pi}_5 &= -3.475, & \hat{\pi}_6 &= -2.725 \\ \hat{\pi}_7 &= -0.225, & \hat{\pi}_8 &= -0.475 \\ \hat{\pi}_9 &= 2.525, & \hat{\pi}_{10} &= 4.775 \end{aligned}$$

Beispiel

Interaktionseffekte $(\widehat{\pi\beta})_{ij} = y_{ij} - (\hat{\mu} + \hat{\pi}_i + \hat{\beta}_j)$

$(\widehat{\pi\beta})_{ij}$	B_1	B_2	B_3	B_4
Vp 1	7.425	0.525	-6.475	-1.475
Vp 2	-4.825	-1.725	5.275	1.275
Vp 3	-0.575	-0.475	8.525	-7.475
Vp 4	6.175	2.275	-7.725	-0.725
Vp 5	1.675	-5.225	-3.225	6.775
Vp 6	-1.075	1.025	-1.975	2.025
Vp 7	-2.575	11.525	-8.475	-0.475
Vp 8	-3.325	-1.225	6.775	-2.225
Vp 9	-2.325	-7.225	7.775	1.775
Vp 10	-0.575	0.525	-0.475	0.525

Beispiel

Damit berechnet man

$$
\begin{aligned}
SS_B &= 10 \sum_{j=1}^{4} \hat{\beta}_j^2 \\
&= 10 \cdot [(-0.675)^2 + (-2.775)^2 + (-2.775)^2 + (6.225)^2] \\
&= 546.075
\end{aligned}
$$

$$
SS_{P\times B} = \sum_{i=1}^{10} \sum_{j=1}^{4} (\widehat{\pi\beta})_{ij}^2 = 881.675
$$

$$
MS_B = \frac{SS_B}{J-1} = \frac{546.075}{3} = 182.025
$$

$$
MS_{P\times B} = \frac{SS_{P\times B}}{(I-1)(J-1)} = \frac{881.675}{27} = 32.655
$$

Beispiel

Als Teststatistik ergibt sich

$$F = \frac{MS_B}{MS_{P\times B}} = \frac{182.025}{32.655} = 5.57$$

Für eine Irrtumswahrscheinlichkeit von $\alpha = 0.05$ gilt für das kritische Quantil

$$F_{0.95}(3, 27) = 2.96$$

Damit ist die Nullhypothese zu verwerfen, die Versuchsbedingungen unterscheiden sich also hinsichtlich ihrer Effekte

Allerdings

- wurde mit dem tatsächlichen Stichprobenumfang die angestrebte Teststärke nicht erreicht
- zeigen nicht alle Effekte in die interessierende Richtung

Varianzanalyse mit Messwiederholung in R: Dateneingabe

```
## Breites Format
dat_wide <- read.table(header = TRUE, text = "
  subj B1 B2 B3 B4
   Vp1 32 23 16 30
   Vp2 19 20 27 32
   Vp3 16 14 23 16
   Vp4 32 26 16 32
   Vp5 21 12 14 33
   Vp6 19 19 16 29
   Vp7 20 32 12 29
   Vp8 19 19 27 27
   Vp9 23 16 31 34
  Vp10 27 26 25 35
")
## Langes Format
dat <- reshape(dat_wide, varying = 2:5, v.names = "score", timevar = "treat",
               times = paste0("B", 1:4), idvar = "subj", direction = "long")
```

Varianzanalyse mit Messwiederholung in R

Mit dem Aufruf

```
m <- aov(score ~ treat + Error(subj + subj:treat), dat)
summary(m)
```

resultiert folgende ANOVA-Tabelle

```
Error: subj
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  404.2   44.91

Error: subj:treat
          Df Sum Sq Mean Sq F value  Pr(>F)
treat      3  546.1  182.03   5.574 0.00414
Residuals 27  881.7   32.65
```

Geplante Vergleiche und Kontraste

Gegeben die Erwartungswerte $\mu_1, \ldots, \mu_m$ von m verschiedenen Populationen, betrachtet man als Kontraste Linearkombinationen der Form

$$c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \ldots + c_m \cdot \mu_m$$

unter der Nebenbedingung

$$c_1 + c_2 + \ldots + c_m = 0$$

Der empirische Kontrast der entsprechenden Stichprobenmittelwerte $\bar{X}_1, \ldots, \bar{X}_m$ ist ein erwartungstreuer Schätzer für den Kontrast der Erwartungswerte der Populationen, d. h. es gilt

$$E\left(\sum_{i=1}^{m} c_i \cdot \bar{X}_i\right) = \sum_{i=1}^{m} c_i \cdot E(\bar{X}_i) = \sum_{i=1}^{m} c_i \cdot \mu_i$$

Kontraste: Beispiel

Werden in einer einfaktoriellen Varianzanalyse drei Gruppen verglichen, von denen die erste eine Kontrollgruppe ist und die übrigen zwei Treatmentgruppen, so ist man daran interessiert, ob die beiden Treatments generell einen Effekt haben

Man betrachtet hierzu den Kontrast

$$\frac{\mu_2 + \mu_3}{2} - \mu_1$$

für den gilt

$$c_1 + c_2 + c_3 = (-1) + \frac{1}{2} + \frac{1}{2} = 0$$

Testen von Kontrasten

Die Hypothesen

$$H_0\colon\ c_1 \cdot \mu_1 + \ldots + c_m \cdot \mu_m = 0, \quad H_1\colon\ c_1 \cdot \mu_1 + \ldots + c_m \cdot \mu_m \neq 0$$

lassen sich folgendermaßen statistisch testen

Man berechnet die Summe der Abweichungsquadrate für den betrachteten Kontrast, z. B. bei einer einfaktoriellen Varianzanalyse mit festen Effekten bei $i = 1, \ldots, m$ Faktorstufen und Zellenbesetzungen n_i durch

$$SS_{Kontrast} = \frac{(\sum_{i=1}^{m} c_i \cdot \bar{X}_i)^2}{\sum_{i=1}^{m} \frac{c_i^2}{n_i}}$$

$SS_{Kontrast}$ hat einen Freiheitsgrad und wird in Beziehung gesetzt zur mittleren Abweichung des Fehlers, im vorliegenden Beispiel also MS_W

Testen von Kontrasten

Es resultiert die $F(1, N - m)$-verteilte Prüfgröße

$$F = \frac{SS_{Kontrast}}{MS_W}$$

mit $N = \sum_{i=1}^{m} n_i$

Ist man an mehr als einem Kontrast interessiert, so ist für die Schlussfolgerungen sicher zu stellen, dass die verschiedenen Kontraste unabhängig sind

Vorausgesetzt die betrachteten unabhängigen Stichproben stammen aus normalverteilten Grundgesamtheiten mit gleichen Varianzen und haben alle denselben Umfang (d. h. gleiche Zellenbesetzungen), so lässt sich die Unabhängigkeit von Kontrasten anhand ihrer Koeffizienten wie folgt charakterisieren

Orthogonale Kontraste

Sind $c_1, \ldots, c_m$ und $c'_1, \ldots, c'_m$ die Koeffizienten von zwei Kontrasten, dann gilt Unabhängigkeit falls

$$c_1 \cdot c'_1 + \ldots + c_m \cdot c'_m = 0,$$

man spricht in diesem Fall von *orthogonalen* Kontrasten

Gegeben m unabhängige Stichprobenmittelwerte, so gibt es nicht mehr als $m - 1$ Kontraste, die vom Grand Mean und untereinander unabhängig sind

Man kann zeigen, dass die Summe der Abweichungsquadrate für $m - 1$ unabhängige Kontraste gerade die Summe der Abweichungsquadrate zwischen diesen m Gruppen (z. B. SS_B in einfaktorieller Varianzanalyse) ergibt, letztere wird also durch $m - 1$ unabhängige Kontraste vollständig zerlegt

Für die hierzu durchzuführenden multiplen Tests ist die Fehlerrate durch geeignete Korrekturen (z. B. Bonferroni-Korrektur) zu kontrollieren

Post-Hoc-Tests

Für nachträgliche Vergleiche zwischen Gruppen werden vor allem folgende zwei Verfahren verwendet

- Verfahren nach Scheffé für beliebige Vergleiche
- Tukeys "Honestly Significant Differences" (HSD) für paarweise Vergleiche

Beide Verfahren korrigieren ihre Aussagen für die Durchführung multipler Tests, wobei der Scheffé-Test bei Betrachtung paarweiser Vergleiche als zu konservativ gilt (tendiert zur Beibehaltung der Nullhypothese)

Tukeys HSD bestimmt eine Verteilung für die maximale auftretende Differenz zwischen zwei Gruppenmittelwerten, unter Berücksichtigung der mittleren Abweichungsquadrate der Fehler, und leitet daraus einen kritischen Wert für die paarweisen Vergleiche der Gruppenmittelwerte ab

Tukeys HSD in R

Die Funktion `TukeyHSD()` kann auf den Output der Funktion `aov()` angewendet werden

Ausgegeben werden dann Konfidenzintervalle und p-Werte (adjustiert für die Durchführung multipler Tests) für sämtliche Differenzen zwischen Gruppenmittelwerten

`TukeyHSD()` verfügt auch über eine Plotmethode, mit der die berechneten Konfidenzintervalle veranschaulicht werden

Wir betrachten die Anwendung auf das oben besprochene Beispiel einer zweifaktoriellen Varianzanalyse mit festen Effekten (Studienzufriedenheit)

```
m <- aov(score ~ motiv*fam, dat)
TukeyHSD(m)
```

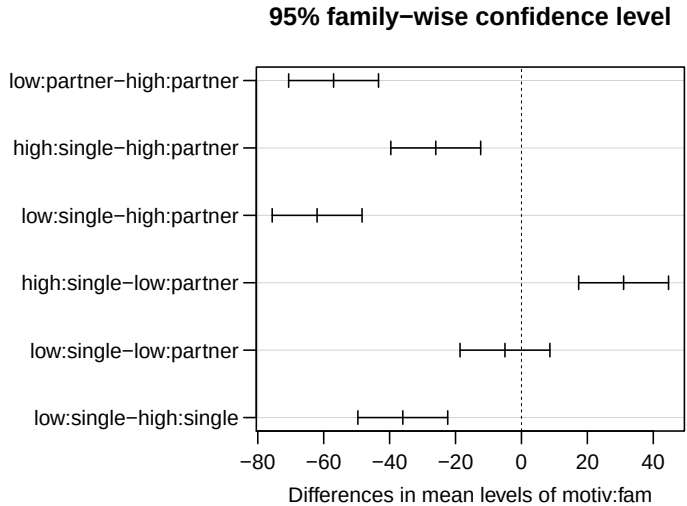
Tukeys HSD in R

```
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = score ~ motiv * fam, data = dat)

$motiv
          diff   lwr   upr p adj
low-high -46.5 -53.6 -39.4     0
$fam
                diff   lwr   upr p adj
single-partner -15.5 -22.6 -8.36     0
$`motiv:fam`
                          diff   lwr    upr p adj
low:partner-high:partner   -57 -70.6 -43.37 0.000
high:single-high:partner   -26 -39.6 -12.37 0.000
low:single-high:partner    -62 -75.6 -48.37 0.000
high:single-low:partner     31  17.4  44.63 0.000
low:single-low:partner      -5 -18.6   8.63 0.724
low:single-high:single     -36 -49.6 -22.37 0.000
```

Tukeys HSD in R

```
TukeyHSD(
  m, conf.level = 0.95
) |>
  plot()
```



Fazit: Einzelvergleiche von Erwartungswerten

Einzelvergleiche von Erwartungswerten können als a priori geplante Vergleiche bzw. Kontraste durchgeführt werden oder als Post-Hoc-Tests

- Geplante Vergleiche weisen dabei eine höhere Power auf
- Die Unabhängigkeit von Kontrasten setzt diesen Vergleichen aber auch engere Grenzen gegenüber den Post-Hoc-Tests

Ist der globale F-Test signifikant für eine Irrtumswahrscheinlichkeit α, dann muss bei Post-Hoc-Tests wenigstens einer der möglichen Vergleiche auf dem gegebenen Signifikanzniveau signifikant sein

- Dieser signifikante Vergleich muss nicht notwendig auch gefunden werden (z. B. bei Durchführung von Tukeys HSD oder dem Scheffé-Test)

Fazit: Einzelvergleiche von Erwartungswerten

Liegt für den globalen F-Wert keine Signifikanz vor, dann können prinzipiell auch Einzelvergleiche durchgeführt werden, aber

- in diesem Fall müssen Aussagen über die den Testentscheidungen unterliegenden Wahrscheinlichkeiten nicht notwendig korrekt sein
- Ergebnisse solcher Tests sind nur interpretierbar, wenn sie vorher spezifiziert wurden

Geplante Vergleiche dienen daher der Überprüfung von Fragestellungen, die sich aus der Anlage der Untersuchung ergeben und bieten die höchstmögliche Teststärke zu deren Beantwortung

Post-Hoc-Tests zielen zwar auf die Identifizierung der Ursachen für einen signifikanten globalen F-Wert ab; ihre Interpretation ist aber meist nicht eindeutig

Interpretation von Interaktionseffekten

Der Interaktionseffekt in einer zweifaktoriellen Varianzanalyse bedeutet, dass die Wirkung eines Faktors von der Faktorstufe des anderen Faktors abhängt

In drei- und mehrfaktoriellen Varianzanalysen können neben paarweisen Interaktionen auch Interaktionen höherer Ordnung auftreten

In einer dreifaktoriellen Varianzanalyse mit den Faktoren A, B und C hat man

- die möglichen paarweisen Interaktionen

$$A \times B$$
$$A \times C$$
$$B \times C$$

- die dreifache Interaktion

$$A \times B \times C$$

Interpretation von Interaktionseffekten

Grundsätzlich lässt sich eine dreifache Interaktion folgendermaßen erklären

- Die paarweise Interaktion zwischen zwei Faktoren verändert sich, abhängig von der Faktorstufe des dritten Faktors
- Beispielsweise hängt die Stärke bzw. die Art der Interaktion $A \times B$ davon ab, welche Stufe hinsichtlich des Faktors C vorliegt
- In gleicher Weise mögliche Interpretationen beziehen sich aber auch auf die Interaktion $A \times C$ abhängig von der Stufe des Faktors B, oder die Interaktion $B \times C$ abhängig von Stufe des Faktors A