Vorlesung 'Multivariate Verfahren'

2. Sitzung

Uwe Altmann   Rolf Steyer

Institut für Psychologie
Friedrich-Schiller-Universität Jena

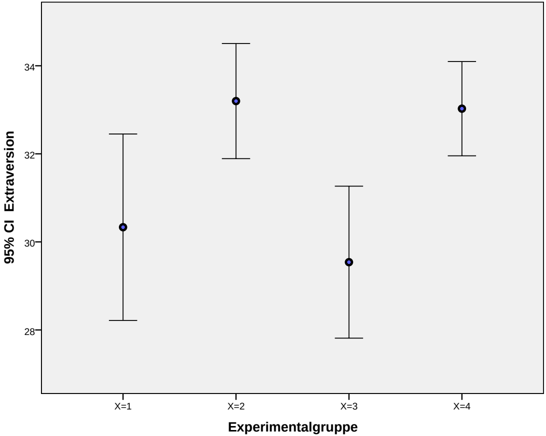
Jena, 26.10.2011

Überblick zur heutigen Vorlesung

1. Einfaktorielle Varianzanalyse (ANOVA)
   - Kontrasttests
2. Mehrfaktorielle Varianzanalyse
   - Datenbeispiel
   - Mehrfaktorielle Varianzanalyse mit Interaktion

Beispiel für SPSS Output der einfaktoriellen ANOVA

- Frage: Unterscheiden sich mindestens zwei der vier Gruppenmittelwerte?
- Stichprobenumfang $n = 176$, Anzahl der Faktorstufen $p = 4$



- Mit Kontrasttests kann innerhalb der ANOVA geprüft werden, ob sich ausgewählte Mittelwerte unterscheiden.
- Ein Kontrast ist ein Vektor, dessen Anzahl der Elemente gleich der Anzahl der Gruppen ist. Das erste Element des Kontrastes wird dem Mittelwert der ersten Gruppe zugeordnet, das zweite Element dem Mittelwert der zweiten Gruppe usw.

$$\mathbf{c} = (c_1, c_2, \ldots, c_p)$$

- Die allgemeine Nullhypothese lautet

$$\mathrm{H}_0: \; c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \ldots + c_p \cdot \mu_p = 0 \quad .$$

Die Prüfgröße ist der Schätzer des Kontrastwertes geteilt durch seinen Standardfehler. Sie ist t-verteilt mit $df = n - p$.

- Im vorangegangenen Beispiel wurden vier Experimentalgruppen betrachtet. Gruppe 1 und 3 bekamen beim Experiment die Instruktion „Antworten Sie so ehrlich wie möglich“. Dagegen lautete die Instruktion bei Gruppe 2 und 4: „Stellen Sie sich so gut wie möglich dar.“
- Wir wollen die Mittelwerte der Gruppe 1 (ehrlich) und Gruppe 2 (unehrlich) vergleichen. Es wird dafür folgender Kontrast formuliert:

$$\mathbf{c} = (1, -1, 0, 0) \quad .$$

- Die entsprechende Nullhypothese lautet

$$\begin{aligned} H_0: \quad 1 \cdot \mu_1 - 1 \cdot \mu_2 + 0 \cdot \mu_3 + 0 \cdot \mu_4 &= 0 \\ \mu_1 - \mu_2 &= 0 \end{aligned}$$

Einfaktorielle ANOVA und Kontrasttests – Beispiel 2

- Für den Vergleich von „ehrlichen" Antworten (Gruppe 1 und 3) und „unehrlichen" Antworten (Gruppe 2 und 4) wird folgender Kontrast formuliert:

$$\mathbf{c} = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right) \quad .$$

- Die entsprechende Nullhypothese lautet

$$\begin{aligned} H_0: \quad \tfrac{1}{2}\cdot\mu_1 - \tfrac{1}{2}\cdot\mu_2 + \tfrac{1}{2}\cdot\mu_3 - \tfrac{1}{2}\cdot\mu_4 &= 0 \\ \tfrac{1}{2}\cdot(\mu_1+\mu_3) - \tfrac{1}{2}\cdot(\mu_2+\mu_4) &= 0 \end{aligned}$$

SPSS Output der einfaktoriellen ANOVA, nun mit Kontrasttest

Kontrast-Koeffizienten

Kontrast	experimental group			
	Design: H-H-F	Design: H-F-H	Design: X-H-F	Design: X-F-H
1	,5	-,5	,5	-,5

Kontrast, wie er vor der Analyse formuliert wurde

Kontrast-Tests

		Kontrast	Kontrastwert	Standardfehler	T	df	Signifikanz (2-seitig)
NEO_E.2: Extraversion	Varianzen sind gleich	1	-3,18	,778	-4,082	172	,000
	Varianzen sind nicht gleich	1	-3,18	,795	-3,992	130,076	,000

1. Zeile: t-Test bei Varianzhomogenität

2. Zeile: Welch-Test bei Varianzheterogenität

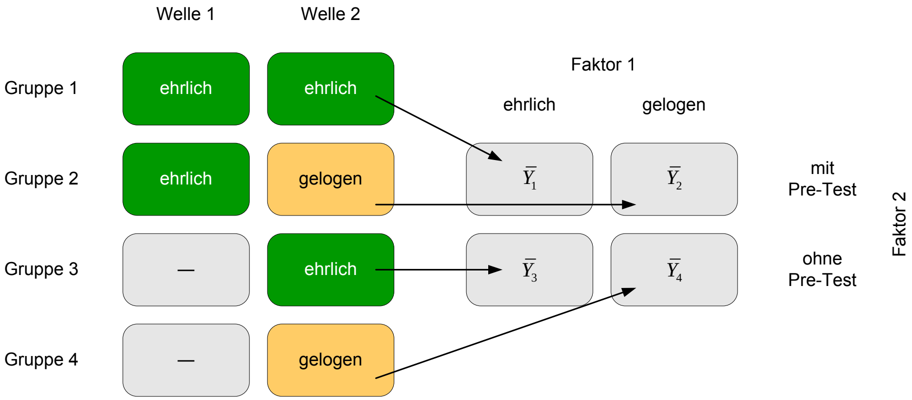
$$T = \frac{K}{SE} = \frac{-3.18}{0.778} = -4.082$$

$$df = n - p = 176 - 4 = 172$$

- Die Interpretation des Schätzers des Kontrastes ist von der Wahl der Kontrastwerte abhängig. In diesem Fall ist $\hat{K} = -3.18$ die Mittelwertsdifferenz von ehrlichen und unehrlichen Antworten.

Inhalt

Solomon-Desgin und zweifaktorielle ANOVA



- Viele Studien verwenden ein Between-Design und eine einmalige Erhebung der interessierenden Variablen.
- Kritik: Antworten könnten möglicherweise anders ausfallen, wenn Probanden schon einmal mit dem gleichen Fragebogen befragt wurden.
- Deshalb Experiment mit Solomon-Design → „pretest sensitization"

Modell der zweifaktoriellen Varianzanalyse mit Interaktion:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

bzw.

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

- Faktor A hat p_1 Faktorstufen und Faktor B hat p_2 Faktorstufen, d.h. $i \in \{1, \ldots, p_1\}$ und $j \in \{1, \ldots, p_2\}$. k ist hier der Index der Person innerhalb der jeweiligen Gruppe.
- Der Mittelwert von Y ist μ. Der Mittelwert der Gruppe i ist μ_i. Usw.
- $\alpha_i = \mu_i - \mu, \quad \beta_j = \mu_j - \mu, \quad (\alpha\beta)_{ij} = \mu_{ij} - \alpha_i - \beta_j - \mu$

Zerlegung der Quadratsummen

$$SS_{\text{total}} = SS_{\text{A}} + SS_{\text{B}} + SS_{\text{AB}} + SS_{\text{Residuum}}$$

Mehrfaktorielle Varianzanalyse mit Interaktion

- unkorrelierte Faktoren $\rightarrow$ orthogonale Varianzanalyse
- korrelierte Faktoren $\rightarrow$ nicht-orthogonale Varianzanalyse

- gleiche Umfänge der Sub-Stichproben $\rightarrow$ balanciertes Design
- ungleiche Umfänge der Sub-Stichproben $\rightarrow$ unbalanciertes Design

- Es sollten unkorrelierten Faktoren und gleiche Umfänge der Sub-Stichproben angestrebt werden. Solch „ideale“ Daten lassen sich in Experimenten gewinnen.
- Annahmen bei der ANOVA: Unabhängigkeit der Messungen, Homoskedastizität und Normalverteilung der Residuuen

Mehrfaktorielle Varianzanalyse mit Interaktion – Beispiel

Tests der Zwischensubjekteffekte

Abhängige Variable:NEO_E.2: Extraversion

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	
Korrigiertes Modell	466,023[a]	3	155,341	5,861	,001	
Konstanter Term	174075,677	1	174075,677	6568,067	,000	
with_pretest	10,218	1	10,218	,386	,535	$H_0: \alpha_i = 0 \quad \forall i$
faking.2	441,561	1	441,561	16,661	,000	$H_0: \beta_j = 0 \quad \forall j$
with_pretest * faking.2	4,229	1	4,229	,160	,690	$H_0: (\alpha\beta)_{ij} = 0 \quad \forall i, j$
Fehler	4558,573	172	26,503			
Gesamt	180348,545	176				
Korrigierte Gesamtvariation	5024,597	175				

a. R-Quadrat = .093 (korrigiertes R-Quadrat = .077)

$$MS_B = \frac{1}{df_B} SS_B = \frac{1}{1} 441.561 = 441.561$$

$$F = \frac{MS_B}{MS_{\text{Residuum}}} = \frac{441.561}{26.503} = 16.661$$

Gegeben sei eine metrische Outcomevariable, die unabhängige Variable $X \in \{1, 9\}$ und die unabhängige Variable $Z \in \{1, 9\}$.

Eine zur zweifaktoriellen Varianzanalyse mit Interaktion äquivalente Regression lautet:

$$E\,(\,Y\,|\,X, Z\,) = \beta_0 + \beta_1 I_{X=1} + \beta_2 I_{Z=1} + \beta_3 I_{X=1} I_{Z=1} \quad .$$

Es kann gezeigt werden, dass

- $\beta_0 = E\,(\,Y\,|\,X = 9, Z = 9\,)$,
- $\beta_1 = E\,(\,Y\,|\,X = 1, Z = 9\,) - \beta_0$,
- $\beta_2 = E\,(\,Y\,|\,X = 9, Z = 1\,) - \beta_0$ und
- $\beta_3 = E\,(\,Y\,|\,X = 1, Z = 1\,) - \beta_0 - \beta_1 - \beta_2$.

Parameterschätzer

Abhängige Variable :NEO_E.2: Extraversion

Parameter	Regressionskoeffizient B	Standardfehler	T	Sig.	95%-Konfidenzintervall Untergrenze	95%-Konfidenzintervall Obergrenze	
Konstanter Term	29,540	,759	38,916	,000	28,041	31,038	$H_0: \beta_0 = 0$
[Z2=1]	,794	1,121	,708	,480	-1,418	3,006	$H_0: \beta_2 = 0$
[Z2=9]	0[a]	.	.	.	.	.	
[X2=1]	3,486	1,073	3,248	,001	1,367	5,605	$H_0: \beta_1 = 0$
[X2=9]	0[a]	.	.	.	.	.	
[X2=1] * [Z2=1]	-,622	1,556	-,399	,690	-3,693	2,450	$H_0: \beta_3 = 0$
[X2=1] * [Z2=9]	0[a]	.	.				
[X2=9] * [Z2=1]	0[a]	.	.	.			
[X2=9] * [Z2=9]	0[a]	.	.	.			

$$T = \frac{\hat{\beta}_3}{SE} = \frac{-0.622}{1.556} = -0.399$$

a. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.