

14 Unterschiede zwischen mehreren abhängigen Stichproben: Varianzanalyse mit Messwiederholung und verwandte Verfahren

Was Sie in diesem Kapitel lernen

- ▶ Wie testet man, ob Bedingungsmittelwerte bei Experimenten mit intraindividueller Bedingungsvariation signifikant voneinander abweichen?
- ▶ In welche Bestandteile wird bei der Varianzanalyse mit Messwiederholung die Gesamtvariation zerlegt?
- ▶ Wie können bei der Varianzanalyse mit Messwiederholung spezifische Hypothesen über Mittelwertsunterschiede getestet werden?
- ▶ Wie werden Interaktionseffekte zwischen zwei messwiederholten Faktoren getestet?
- ▶ Wie werden Interaktionseffekte zwischen einem messwiederholten und einem nicht-messwiederholten Faktor getestet?
- ▶ Wie testet man auf der Basis eines nonparametrischen Tests, ob die Mediane mehrerer messwiederholter Bedingungen signifikant voneinander abweichen?

In Kapitel 13 haben wir die Varianzanalyse als Möglichkeit kennengelernt, Mittelwertsunterschiede aus mehreren unabhängigen Stichproben auf ihre statistische Bedeutsamkeit hin zu testen. Die Variation in den Messwerten, die ein Faktor (bzw. mehrere Faktoren und ihre Wechselwirkung) verursacht, haben wir als Effekte (Haupteffekte, Interaktionseffekte) bezeichnet und gesehen, dass solche Effekte auf der Basis eines *F*-Tests inferenzstatistisch abgesichert werden können. Wir haben festgestellt, dass eine notwendige Voraussetzung für die Anwendung des in Kapitel 13 beschriebenen *F*-Tests darin besteht, dass es zwischen Messwerten, die aus unterschiedlichen Stichproben (d.h. aus unterschiedlichen Stufen des Faktors) stammen, keinerlei gegenseitigen Abhängigkeiten gibt. Es muss sich also bei den Stufen eines Faktors jeweils um *unabhängige* Stichproben handeln: In den unterschiedlichen Faktorstufen müssen sich unterschiedliche Personen befinden.

In diesem Kapitel werden wir nun sehen, wie die Varianzanalyse funktioniert, wenn es sich bei den Faktorstufen um *abhängige* Stichproben handelt. Was mit abhängigen Stichproben gemeint ist (und was nicht mit ihnen gemeint ist), haben wir bereits in Kapitel 12 ausführlich behandelt. Abhängige Stichproben liegen z.B. vor, wenn es sich um Experimente mit intraindividueller Bedingungsvariation handelt (sog. messwiederholte Faktoren) oder wenn unterschiedliche Versuchspersonen in den unterschiedlichen Faktorstufen einander zugeordnet werden können (aufgrund einer »natürlichen Beziehung« zwischen ihnen oder aufgrund gleicher Ausprägungen auf einer Kontrollvariablen; vgl. die Technik des Parallelisierens in Abschn. 4.3.3). Da die wiederholte Messung an den gleichen Personen den typischen Anwendungsfall abhängiger Stichproben darstellt, wird die entsprechende Auswertungsprozedur als »Varianzanalyse mit Messwiederholung« (engl. *repeated-measures analysis of variance*, kurz: RM-ANOVA oder auch *within-subjects ANOVA*) bezeichnet. Den Fall einer einfaktoriellen Varianzanalyse mit Messwiederholung werden wir in Abschnitt 14.1 detailliert behandeln.

Im Falle von mehrfaktoriellen Designs können Messwiederholungen bei keinem, einigen oder allen Faktoren vorliegen. Sind alle Faktoren messwiederholt, spricht man von einem komplett messwiederholten Design; sind nur einige Faktoren messwiederholt, andere hingegen nicht, spricht man von einem partiell messwiederholten Design. Wir werden uns in Abschnitt 14.2 auf zwei Fälle beschränken: ein zweifaktorielles Design mit vollständiger Messwiederholung und ein zweifaktorielles Design mit Messwiederholung auf einem Faktor. Abschließend werden wir in diesem Kapitel noch einen nonparametrischen Test auf Unterschiede zwischen mehreren Medianen im Falle eines messwiederholten Faktors vorstellen (Abschn. 14.3) und auf Verfahren für kategoriale Variablen verweisen (Abschn. 14.4).

14.1 Einfaktorielle Varianzanalyse mit Messwiederholung

Wir haben in Abschnitt 13.1 darauf hingewiesen, dass die einfaktorielle Varianzanalyse *ohne* Messwiederholung eine Erweiterung (oder Verallgemeinerung) des *t*-Tests für unabhängige Stichproben ist (vgl. Abschn. 11.1.2). Dementsprechend ist die einfaktorielle Varianzanalyse *mit* Messwiederholung eine Erweiterung (oder Verallgemeinerung) des *t*-Tests für abhängige Stichproben (vgl. Abschn. 12.1.1). Sie wird verwendet, wenn die Nullhypothese getestet werden soll, dass sich die Mittelwerte mehrerer abhängiger Stichproben (z. B. experimenteller Bedingungen) nicht voneinander unterscheiden.

Der klassische Fall von abhängigen Stichproben ist die intraindividuelle Bedingungsvariation: Von allen Personen in der Stichprobe werden wiederholt Messwerte unter unterschiedlichen experimentellen Bedingungen erhoben. Beispielsweise könnte man die kognitive Leistungsfähigkeit von Personen dreimal erheben und miteinander vergleichen: (1) ohne Stimmungsinduktion, (2) nach positiver Stimmungsinduktion und (3) nach negativer Stimmungsinduktion. Mit einem solchen Design könnte die Hypothese überprüft werden, dass sich positive Stimmungförderlich und negative Stimmung hemmend auf die kognitive Leistungsfähigkeit auswirken. Da in allen drei experimentellen Bedingungen die gleichen Personen getestet werden, handelt es sich um abhängige Stichproben: Die Messwerte werden über die drei Bedingungen hinweg miteinander kovariieren, da die kognitive Leistungsfähigkeit eben nicht nur von der Stimmung, sondern auch von persongebundenen Variablen abhängt, die über die drei Stimmungsbedingungen hinweg stabil bleiben (Intelligenz, Teilnahmemotivation etc.).

Kovariation zwischen den Messungen. Ein Teil der Variation in den Messwerten ist also dadurch zu erklären, dass es sich um die gleichen Personen (und damit teilweise um die gleichen persongebundenen Einflüsse auf die Messwerte) handelt. Wir haben in Kapitel 12.1 die Kovarianz als einen statistischen Kennwert kennengelernt, der angibt, wie groß der Einfluss solcher persongebundenen Merkmale ist (s. hierzu ausführlich Abschn. 16.3.1). Genauer gesagt: Die Kovarianz gibt an, wie stabil Unterschiede zwischen Personen über zwei Messungen hinweg bleiben. Liegen mehr als zwei

Messungen vor, muss diese Stabilität anders quantifiziert werden. In Abschnitt 14.1.2 werden wir sehen, wie diese »Varianz zwischen Personen« bestimmt werden kann und welche Rolle sie für die Zerlegung der Gesamtvariation spielt. Durch die Kovariation wiederholter Messungen können stabile Störvariablen, die an die Person gebunden sind (z. B. Persönlichkeitsmerkmale), statistisch kontrolliert werden. Hierdurch reduziert sich der Anteil unerklärter Varianz. Folglich gewinnt man an Teststärke (Power) und benötigt eine geringere optimale Stichprobengröße als bei einer Varianzanalyse ohne Messwiederholung. Dies ist ein großer Vorteil von Messwiederholungsdesigns.

Sequenzeffekte. Ein Nachteil experimenteller Designs mit intraindividueller Bedingungsvariation besteht darin, dass die Messwerte nicht vor systematischen Verfälschungen gefeit sind. So könnte es sein, dass Versuchspersonen im Laufe der Untersuchung die Hypothese erraten, dass ihre Motivation, am Experiment teilzunehmen, zunehmend nachlässt oder dass sie Antwort- bzw. Lösungsstrategien von einer Messung auf die nächste übertragen bzw. sich an ihre früheren Antworten erinnern. All diese systematischen Verfälschungen, die typisch für Messwiederholungsdesigns sind, werden unter dem Begriff *Sequenzeffekte* zusammengefasst. Für eine Systematik solcher Sequenzeffekte und den Umgang mit ihnen (etwa die Berücksichtigung bestimmter Kontrollstrategien wie die intra- oder interindividuelle Ausbalancierung) verweisen wir auf einschlägige Lehrbücher zur Versuchsplanung (z. B. Huber, 2013; Hussy & Jain, 2002; Shadish et al., 2002).

Veränderungsmessung. In vielen psychologischen Fragestellungen geht es nicht um Unterschiede, die durch die Manipulation einer experimentellen Variablen hervorgerufen werden, sondern lediglich um Unterschiede im Laufe der Zeit. Anders gesagt: Oft ist die Zeit selbst die unabhängige Variable, und man interessiert sich für die Veränderung in den (durchschnittlichen) Messwerten über die Zeit hinweg. Beispielsweise interessiert man sich in der allgemeinen Entwicklungspsychologie dafür, ob und wie sich psychologische Merkmale mit zunehmendem Alter verändern. Dies erfordert die wiederholte Messung der fraglichen Merkmale. Werden über die Zeit hinweg immer wieder Messwerte von den gleichen Personen erhoben, handelt es sich um eine sog. Längsschnittuntersuchung. Auch hier kann man sich mithilfe

einer Varianzanalyse mit Messwiederholung der Frage widmen, ob die Unterschiede in den Mittelwerten zwischen den Messzeitpunkten signifikant variieren.

Evaluationsforschung. Interessiert man sich im Rahmen einer Untersuchung zur Wirksamkeit einer psychologischen Intervention dafür, ob und inwieweit sich durch die Intervention die Merkmalsausprägung über die Zeit hinweg verändert hat (z.B. inwieweit durch eine Psychotherapie das Ausmaß subjektiver Belastungen reduziert wurde oder inwieweit durch ein Lerntraining die Schulleistung von Kindern verbessert wurde), so sind wiederum Messwiederholungsdesigns angezeigt. Messwiederholungsdesigns sind typisch für die Evaluationsforschung, v.a. für die Prozess- und die Wirksamkeitsevaluation (vgl. Gollwitzer & Jäger, 2014).

Datenbeispiel

Um die Grundidee der einfaktoriellen Varianzanalyse mit Messwiederholung zu veranschaulichen, beginnen wir auch hier mit einem einfachen Datenbeispiel. Stellen wir uns vor, das in Abschnitt 13.1 geschilderte Experiment zum Modelllernen sei nicht auf der Basis einer interindividuellen Bedingungsvariation, sondern vielmehr auf der Basis einer *intraindividuellen* Bedingungsvariation durchgeführt worden. Insgesamt werden fünf Personen untersucht, und zwar jeweils unter drei Bedingungen. Zunächst sehen die Probanden einen Film, in dem eine Person ihres Alters für ein bestimmtes Verhalten (z.B. Aggression) belohnt wird. Im Anschluss daran wird mit einem Selbstberichtsmaß (z.B. einer visuellen Analogskala mit Ausprägungen zwischen 0 und 100) gemessen, wie stark die Personen dazu tendieren, das vom Modell gezeigte Verhalten nachzuahmen. Anschließend sehen die gleichen Personen einen zweiten Film, in dem die Modellperson für ihr Verhalten bestraft wird; wiederum wird die Nachahmungstendenz per Selbstauskunft gemessen. Schließlich sehen die Probanden einen dritten Film, in dem das Verhalten der Modellperson ohne Konsequenzen bleibt. Die Nachahmungstendenz wird hier ein drittes Mal gemessen.

Die unabhängige Variable (UV) ist also die Verhaltenskonsequenz (Belohnung, Bestrafung, keine Konsequenz). Die abhängige Variable (AV) ist die Nachahmungstendenz. Sollte die Hypothese, die aus der Theorie des Modelllernens von Bandura (1976, 1977) abgeleitet wurde, korrekt sein, müsste die Nachah-

mungstendenz nach dem ersten Film (Belohnungsbedingung) am stärksten und nach dem zweiten Film (Bestrafungsbedingung) am schwächsten sein. Dass es sich hier zugegebenermaßen um ein ziemlich problematisches Versuchsdesign handelt, da mit erheblichen Verzerrungen durch Sequenzeffekte zu rechnen ist, dürfte klar sein. Trotzdem werden wir mit diesem Beispiel weiterarbeiten.

Die experimentellen Bedingungen erhalten einen Laufindex von $j = 1, \dots, J$ (mit $J = 3$ in unserem Beispiel). Die Personen erhalten einen Laufindex von $m = 1, \dots, n$ (mit $n = 5$ in unserem Beispiel). Den Index j können wir bei n hier weglassen, da die Anzahl der beobachteten Werte pro Bedingung aufgrund der Messwiederholung der Gesamtanzahl aller Versuchspersonen entspricht. Die (fiktiven) Rohdaten des Versuchs stehen in Tabelle 14.1. Beachten Sie, dass es sich hier um die gleichen Rohdaten handelt wie in Tabelle 13.1 (in Abschn. 13.1). Wir haben bewusst die gleichen Daten gewählt, um den Unterschied zwischen der Varianzanalyse ohne Messwiederholung und der Varianzanalyse mit Messwiederholung (und die statistischen Implikationen dieses Unterschiedes) deutlich zu machen.

Personen als Stufen eines »zufälligen Faktors«

Der Unterschied zwischen Tabelle 13.1 und Tabelle 14.1 besteht darin, dass wir hier eine zusätzliche Spalte eingefügt haben, in der die *durchschnittliche* Nachahmungstendenz einer Person über die drei experimentellen Bedingungen hinweg angegeben ist. In Tabelle 13.1 wäre es aufgrund des interindividuellen Designs nicht sinnvoll gewesen, einen solchen Personmittelwert zu berechnen, da den unterschiedlichen experimentellen Bedingungen unterschiedliche Personen zugeordnet waren. Beim intraindividuellen Design hingegen stammen alle Messwerte innerhalb der gleichen Zeile des Versuchsplans von einer und derselben Person. Daher ist hier nicht nur der Bedingungsmittelwert $\bar{x}_{\cdot j}$ (Spaltenmittelwert; s. die untere Zeile in Tab. 14.1), sondern auch der Personmittelwert $\bar{x}_{m \cdot}$ (Zeilenmittelwert; s. die rechte Spalte in Tab. 14.1) sinnvoll interpretierbar.

In Abschnitt 13.1.11 hatten wir die Unterscheidung zwischen »festen« und »zufälligen« Effekten (bzw. Faktoren) kennengelernt. Bei den drei experimentellen Bedingungen in unserem Beispiel handelt es sich zweifelsfrei um Stufen eines festen Faktors, denn das variierte Merkmal kann nur eine bestimmte Anzahl möglicher

Ausprägungen haben, und die realisierten Faktorstufen entsprechen genau diesen Ausprägungen. Bei Messwiederholungsdesigns kann man nun auch die unterschiedlichen Personen als »Stufen« eines »Faktors« auffassen. Der Personfaktor umfasst dabei alle möglichen Unterschiede zwischen den Personen. Da die Personen eine Zufallsstichprobe darstellen und sich daher auch zufällig hinsichtlich der Personmerkmale unterscheiden, handelt es sich formal gesehen um einen zufälligen Faktor (s. Abschn. 13.1.11). Anders als der Bedingungsfaktor ist der Personfaktor von untergeordnetem Interesse für die Hypothesenprüfung, aber die Vorstellung, dass es sich bei den Personen um Stufen eines zufälligen Faktors handelt, hilft uns, die Logik der Quadratsummenzerlegung besser zu verstehen. Wir können also die einfaktorielle Varianzanalyse mit Messwiederholung als quasi-zweifaktorielles Design mit einem (festen) Faktor »Bedingung« und einem zweiten (zufälligen) Faktor »Person« auffassen.

14.1.1 Messwertzerlegung

14

In Tabelle 14.1 gibt es drei Quellen der Variation: (1) Unterschiede zwischen den experimentellen Bedingungen, (2) Unterschiede zwischen den Personen und (3) den Anteil der Variation, der weder auf Haupteffekte der Bedingungen noch auf Personeneffekte zurückgeführt werden kann. Dieser Restanteil (Residualanteil) geht auf Interaktionen zwischen den Personen und den Bedingungen und andere unsystematische Störeinflüsse wie z.B. den Messfehler zurück.

Der Messwert x_{mj} einer Person m in einer Bedingung a_j lässt sich daher wie folgt zerlegen:

$$x_{mj} = \bar{x} + t_j + p_m + e_{mj} \quad (\text{F 14.1})$$

In dieser Gleichung bezeichnet \bar{x} den Gesamtmittelwert aller Werte, $t_j = \bar{x}_{\cdot j} - \bar{x}$ ist der Haupteffekt der j -ten Stufe des Faktors A , d.h. die Abweichung des Bedingungsmittelwerts $\bar{x}_{\cdot j}$ vom Gesamtmittelwert \bar{x} . Mit $p_m = \bar{x}_{m\cdot} - \bar{x}$ wird der Haupteffekt der m -ten Person bezeichnet, d.h. die Abweichung des Mittelwerts $\bar{x}_{m\cdot}$ einer Person m (über alle J Bedingungen hinweg) vom Gesamtmittelwert \bar{x} . Der Residualwert $e_{mj} = x_{mj} - \bar{x} - t_j - p_m$ einer Person m in Bedingung a_j ist der Anteil am Wert der Person, der übrig bleibt, wenn man den Gesamtmittelwert und die Bedingungs- und Personeneffekte abzieht.

14.1.2 Quadratsummenzerlegung

Die Gesamtvariation (totale Quadratsumme) der 15 Messwerte berechnet sich analog zu dem in Abschnitt 13.1.4 beschriebenen Vorgehen. Sie ist definiert als die Summe der quadrierten Abweichungen der einzelnen Messwerte vom Gesamtmittelwert:

$$QS_{\text{tot}} = \sum_{j=1}^J \sum_{m=1}^n (x_{mj} - \bar{x})^2 \quad (\text{F 14.2})$$

Sie beträgt in unserem Beispiel $QS_{\text{tot}} = 4.532$.

Wie kann die totale Quadratsumme bei der einfaktoriellen Varianzanalyse mit Messwiederholung zerlegt werden? Um das zu veranschaulichen, machen wir von der Vorstellung Gebrauch, es handele sich bei den unterschiedlichen Personen um Stufen eines zufälligen

Tabelle 14.1 Fiktive Daten eines Experiments zum Modelllernen (mit intraindividueller Bedingungsvariation)

Person m	Stufe a_j des Faktors			Personmittelwert $\bar{x}_{m\cdot}$
	Belohnung (a_1)	Bestrafung (a_2)	Keine Konsequenz (a_3)	
1	57	18	36	37
2	45	15	27	29
3	49	13	43	35
4	69	37	29	45
5	70	37	55	54
Bedingungsmittelwert $\bar{x}_{\cdot j}$	58	24	38	$\bar{x} = 40$

Faktors. Genau wie jeder andere Faktor kann also auch jede Person einen Haupteffekt, d.h. einen von der experimentellen Bedingung unabhängigen (und insofern »unbedingten«) Effekt, haben. Die Haupteffekte der Personen sind darauf zurückzuführen, dass diese sich über die unterschiedlichen Messungen hinweg konsistent (d.h. gleichbleibend) in Personmerkmalen unterscheiden, welche die AV beeinflussen. Zusätzlich zu den Haupteffekten der Personen gibt es für jede Stufe des Faktors A einen Haupteffekt der experimentellen Bedingung, der von der jeweiligen Person unabhängig ist. Und schließlich kann es sein, dass der Effekt der experimentellen Manipulation bei unterschiedlichen Personen unterschiedlich ausfällt. Unklar bleibt jedoch, ob es sich hierbei um einen echten Interaktionseffekt zwischen Personmerkmalen und der experimentellen Bedingung handelt oder lediglich um unsystematische Effekte wie Messfehler.

Bei der einfaktoriellen Varianzanalyse mit Messwiederholung wird die totale Quadratsumme QS_{tot} also in drei Teile zerlegt:

- ▶ eine Quadratsumme, welche die Variation zwischen Personen angibt (QS_{zwP}),
- ▶ eine Quadratsumme, welche die Variation zwischen den Stufen des Faktors A angibt (QS_{zwA}), und
- ▶ eine Quadratsumme, welche diejenige Variation angibt, die weder durch Haupteffekte der Bedingungen noch durch Haupteffekte der Personen erklärt werden kann. Diese Quadratsumme werden wir im Folgenden als *Residualquadratsumme* (QS_{Res}) bezeichnen.

Variation zwischen Personen

Die Variation zwischen Personen ist derjenige Teil der Gesamtvariation, der auf Unterschiede zwischen den Personen – unabhängig von den Stufen des experimentellen Faktors – zurückgeht. Es handelt sich also um diejenigen Unterschiede zwischen Personen, die sich über alle Faktorstufen hinweg konsistent zeigen. Die Variation zwischen Personen ist ein unbedingter (d.h. von Faktor A unabhängiger) Effekt des »Faktors« Person.

Die Variation zwischen Personen entspricht im Falle von zwei Bedingungen der Kovarianz beim *t*-Test für abhängige Stichproben (s. Abschn. 12.1.1). Sie ist ein Maß für den Einfluss jener Merkmale, die über die experimentellen Bedingungen hinweg konsistente Unterschiede zwischen den Personen produzieren. Solche konsistenten Unterschiede zwischen Personen manifestieren sich in der Variation der Personmittelwerte (s. rechte Spalte in Tab. 14.1).

Die Zwischen-Personen-Quadratsumme QS_{zwP} basiert also auf den quadrierten Abweichungen der Personmittelwerte vom Gesamtmittelwert:

$$QS_{zwP} = J \cdot \sum_{m=1}^n (\bar{x}_{m\bullet} - \bar{x})^2 \quad (\text{F 14.3})$$

Sie beträgt in unserem Beispiel $QS_{zwP} = 1.128$.

Variation zwischen Faktorstufen

Die Variation zwischen den Faktorstufen ist derjenige Teil der Gesamtvariation, der auf systematische Unterschiede zwischen den experimentellen Bedingungen zurückgeführt werden kann. Diese Unterschiede manifestieren sich – genau wie bei der einfaktoriellen Varianzanalyse ohne Messwiederholung – in der Variation der Bedingungsmittelwerte (s. untere Zeile in Tab. 14.1). Die Zwischen-Quadratsumme des Faktors A (QS_{zwA}) basiert also auf den quadrierten Abweichungen der Bedingungsmittelwerte vom Gesamtmittelwert:

$$QS_{zwA} = n \cdot \sum_{j=1}^J (\bar{x}_{\bullet j} - \bar{x})^2 \quad (\text{F 14.4})$$

Sie beträgt in unserem Beispiel $QS_{zwA} = 2.920$.

Variation zwischen Personen in Bezug auf den Effekt des Faktors

Wenn wir die Personen und die Bedingungsvariation als Faktoren auffassen, dann besteht auch die Möglichkeit, dass diese miteinander interagieren. Inhaltlich bedeutet diese Interaktion, dass sich Personen darin unterscheiden, wie groß die Unterschiede in ihren Messwerten zwischen den drei Bedingungen sind. Grafisch kann man die Idee einer solchen Interaktion zwischen Person und Bedingung veranschaulichen, wenn man die Messwerte in einem Liniendiagramm abträgt, wobei die Personen als Ausprägungen auf der Abszisse und die experimentellen Bedingungen als drei unterschiedliche Linien dargestellt werden können (s. Abb. 14.1). Man sieht deutlich, dass die Linien nicht parallel verlaufen; es gibt also (Person-)Unterschiede in den (Bedingungs-)Unterschieden.

Machen wir uns die Idee einer Interaktion zwischen Person und Bedingung an Abbildung 14.1 klar. Wir sehen z.B., dass die Messwerte von Person 2 näher beieinanderliegen als die Messwerte von Person 1. Der Effekt der experimentellen Manipulation ist bei Per-

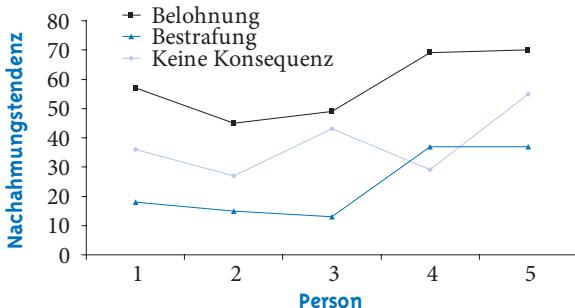


Abbildung 14.1 Grafische Darstellung der Interaktion zwischen Person und Bedingung (Datenbeispiel aus Tab. 14.1)

son 1 größer als bei Person 2. Möglicherweise ist dieser Unterschied zwischen den beiden Personen darauf zurückzuführen, dass Person 2 ihr Verhalten generell weniger stark an stellvertretenden Verhaltensbekräftigungen ausrichtet. Und möglicherweise ist der Effekt der experimentellen Manipulation bei Person 1 genau deshalb stärker, weil diese Person ihr Verhalten sehr stark daran ausrichtet, ob andere Menschen für das gleiche Verhalten belohnt oder bestraft werden. Diese Überlegung würde nahelegen, dass es eine echte Interaktion zwischen der experimentellen Manipulation (stellvertretende Verhaltenskonsequenzen) und Merkmalen der beobachtenden Personen (hier: Sensibilität für stellvertretende Verhaltenskonsequenzen) gibt. Das Problem ist: Wir können nicht testen, ob es sich um eine echte Interaktion handelt oder lediglich um zufällige Schwankungen bzw. Messfehler, die überhaupt nicht auf systematische Personunterschiede zurückzuführen sind. Der Grund dafür, dass wir systematische Person-Bedingungs-Interaktionen nicht testen können, ist der, dass wir in jeder Kombination von Person und Bedingung jeweils nur einen einzigen Messwert haben. Wie stark dieser Messwert von Messfehlern vs. von systematischen Person-Bedingungs-Interaktionseffekten beeinflusst ist, bleibt unbekannt. Insofern werden wir der Einfachheit halber die entsprechende Quadratsumme mit QS_{Res} bezeichnen.

Wie quantifiziert man die Residualquadratsumme QS_{Res} ? Grundsätzlich erfolgt die Berechnung nicht anders, als in Abschnitt 13.2.3 beschrieben. Es handelt sich um die Variation zwischen den Messwerten, die weder auf einen unbedingten Effekt der Person noch auf einen unbedingten Effekt der Bedingung zurückgeführt werden kann. Diese Variation manifestiert sich also in den Abweichungen der Messwerte vom Ge-

samtmittelwert, nachdem der Haupteffekt einer Person und der Haupteffekt der Bedingung a_j von diesen Abweichungen abgezogen wurden (vgl. auch Formel F 13.97a):

$$QS_{Res} = \sum_{j=1}^J \sum_{m=1}^n ((x_{mj} - \bar{x}) - (\bar{x}_{\cdot j} - \bar{x}) - (\bar{x}_{m\cdot} - \bar{x}))^2 \\ = \sum_{j=1}^J \sum_{m=1}^n (x_{mj} - \bar{x}_{\cdot j} - \bar{x}_{m\cdot} + \bar{x})^2 \quad (F 14.5)$$

Sie beträgt in unserem Beispiel $QS_{Res} = 484$.

Die Summe aus QS_{zwP} , QS_{zwA} und QS_{Res} entspricht der totalen Quadratsumme.



Additivität der Quadratsummen

Bei der einfaktoriellen Varianzanalyse mit Messwiederholung lässt sich die totale Quadratsumme QS_{tot} in drei Teile zerlegen:

- ▶ einen Teil, der die Variation zwischen Personen ausdrückt (»Haupteffekte« der Person; QS_{zwP}),
- ▶ einen Teil, der die Variation zwischen Bedingungen ausdrückt (Haupteffekte des Faktors A; QS_{zwA}), und
- ▶ einen unerklärten Teil (QS_{Res}):

$$QS_{tot} = QS_{zwP} + QS_{zwA} + QS_{Res} \quad (F 14.6)$$

Variation zwischen und innerhalb von Personen

Versuchen wir nun, die Quadratsummenzerlegung bei der einfaktoriellen Varianzanalyse mit Messwiederholung mit jener ohne Messwiederholung zu vergleichen. Bei der einfaktoriellen Varianzanalyse ohne Messwiederholung haben wir die totale Quadratsumme in zwei Teile zerlegt: einen, der Variation zwischen Bedingungen (QS_{zw}), und einen, der Variation innerhalb von Bedingungen (QS_{inn}). Bei der einfaktoriellen Varianzanalyse mit Messwiederholung haben wir die totale Quadratsumme in drei Teile zerlegt: einen, der Variation zwischen Bedingungen (QS_{zwA}), einen, der Variation zwischen Personen (QS_{zwP}), und einen dritten, der unerklärte Variation anzeigt (QS_{Res}).

Variation zwischen Bedingungen. Die Variation zwischen Bedingungen ist in beiden varianzanalytischen Modellen die gleiche, und sie wird auch gleich be-

rechnet (vgl. die Formeln F 13.6b und F 14.4). Da unser Datenbeispiel in Tabelle 14.1 exakt dem in Tabelle 13.1 entspricht, resultiert für die Variation zwischen den Bedingungen in beiden Fällen der gleiche Wert, nämlich $QS_{zwA} = 2.920$. Das bedeutet auch, dass der Anteil der Variation, der auf die experimentelle Manipulation zurückzuführen ist (also der Anteil der QS_{zwA} an der totalen Quadratsumme), für beide Modelle gleich ist. Darauf kommen wir später zurück, wenn es um die Bestimmung der Effektgröße geht. Der einzige formale Unterschied besteht darin, dass es sich bei der QS_{zw} der Varianzanalyse ohne Messwiederholung um eine Variation *zwischen Personen* handelt, während es sich bei der QS_{zwA} der Varianzanalyse mit Messwiederholung aufgrund der intraindividuellen Bedingungsvariation um eine Variation *innerhalb von Personen* handelt.

Variation innerhalb Bedingungen. Bei der einfaktoriellen Varianzanalyse ohne Messwiederholung haben wir die Innerhalb-Quadratsumme wie folgt hergeleitet: Wir hatten einen Fall konstruiert, in dem es nur Unterschiede *innerhalb* der, nicht aber *zwischen* den experimentellen Bedingungen gibt (s. Abschn. 13.1.4). Die Quadratsumme dieser an ihrem jeweiligen Bedingungsmittelwert zentrierten Messwerte betrug gemäß Formel F 13.7 $QS_{inn} = 1.612$. Ein Vergleich mit den Quadratsummen, die wir in diesem Kapitel berechnet haben, zeigt: Dieser Wert entspricht der Summe aus der Zwischen-Personen-Quadratsumme QS_{zwP} und der Residualquadratsumme QS_{Res} ($1.128 + 484 = 1.612$). Mit anderen Worten: Die QS_{inn} bei der Varianzanalyse ohne Messwiederholung entspricht der Summe aus QS_{zwP} und QS_{Res} bei der Varianzanalyse mit Messwiederholung:

$$QS_{inn} = QS_{zwP} + QS_{Res} \quad (\text{F 14.7})$$

Bei der QS_{zwP} handelt es sich um Variation *zwischen Personen*, während es sich bei der QS_{Res} um Variation *innerhalb von Personen* handelt.

QS_{zwP} und QS_{innP} . Während also bei der Varianzanalyse ohne Messwiederholung eine Unterscheidung in Variationsquellen zwischen den bzw. innerhalb der *Bedingungen* im Vordergrund steht, wird bei der Varianzanalyse mit Messwiederholung zunächst danach unterschieden, ob es sich um Variationsquellen zwi-

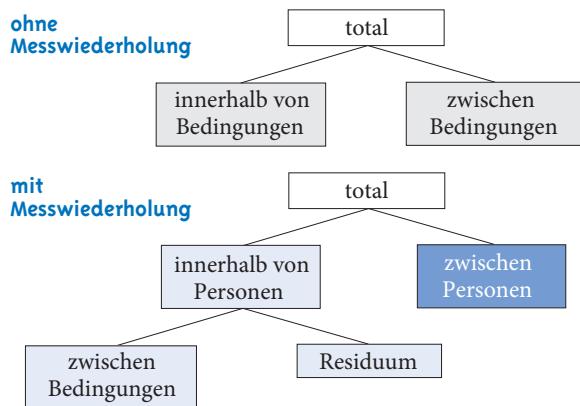


Abbildung 14.2 Quadratsummenzerlegung bei der einfaktoriellen Varianzanalyse mit und ohne Messwiederholung

schen oder innerhalb der *Personen* handelt. Die Variation zwischen Personen drückt sich in der Zwischen-Personen-Quadratsumme QS_{zwP} aus, die Variation innerhalb von Personen in der Zwischen-Bedingungen-Quadratsumme QS_{zwA} und der Quadratsumme QS_{Res} , deren Summe die Innerhalb-Personen-Quadratsumme QS_{innP} ergibt. Sie berechnet sich in unserem Beispiel zu $2.920 + 484 = 3.404$. Die Quadratsummenzerlegung bei der einfaktoriellen Varianzanalyse mit und ohne Messwiederholung ist in Abbildung 14.2 grafisch veranschaulicht.

14.1.3 Effektgrößenmaße

Als Effektgröße bietet sich auch bei der einfaktoriellen Varianzanalyse mit Messwiederholung an, denjenigen Anteil der Varianz der Messwerte zu quantifizieren, der auf den Effekt des Faktors A zurückzuführen ist. Insofern stimmt die Bedeutung des Konzepts »Effektgröße« bei Varianzanalysen mit Messwiederholung genau mit der Bedeutung bei Varianzanalysen ohne Messwiederholung überein. Bei der Schätzung dieser Effektgröße aus den Daten kann das Ausmaß der stabilen Personunterschiede entweder berücksichtigt werden oder nicht. Genau das macht den Unterschied zwischen dem partiellen Effektgrößenschätzer $\hat{\eta}_p^2$ und dem nicht-partiellen Effektgrößenschätzer $\hat{\eta}^2$ aus. Wir werden beide Maße im Folgenden behandeln und im Anschluss daran diskutieren, ob und wann es sinnvoll ist, die Abhängigkeit der Stichproben bei der Effektgrößenschätzung mit zu berücksichtigen oder nicht.

Nicht-partielles Effektgrößenmaß $\hat{\eta}^2$. Beim sog. »nicht-partiellen« Effektgrößenmaß $\hat{\eta}^2$ wird der Quotient aus der Quadratsumme QS_{zwA} und der totalen Quadratsumme gebildet:

$$\hat{\eta}^2 = \frac{QS_{zwA}}{QS_{\text{tot}}} = \frac{QS_{zwA}}{QS_{zwA} + QS_{zwp} + QS_{\text{Res}}} \quad (\text{F 14.8})$$

Partielles Effektgrößenmaß $\hat{\eta}_p^2$. Beim partiellen Effektgrößenmaß $\hat{\eta}_p^2$ wird der Anteil der Gesamtvarianz, der auf stabile Unterschiede zwischen den Personen zurückgeht, nicht mit berücksichtigt. Bei der Schätzung der Effektgröße geht die QS_{zwp} also nicht mehr mit in die Gleichung ein:

$$\hat{\eta}_p^2 = \frac{QS_{zwA}}{QS_{zwA} + QS_{\text{Res}}} \quad (\text{F 14.9})$$

Da beim partiellen Effektgrößenmaß der Nenner niemals einen größeren Wert annehmen kann als beim nicht-partiellen Effektgrößenmaß, kann $\hat{\eta}_p^2$ niemals kleiner werden als $\hat{\eta}^2$. Der Unterschied zwischen beiden wird umso größer, je größer der Anteil der Gesamtvariation ist, der auf stabile Personunterschiede zurückgeführt werden kann.

Welches Effektgrößenmaß ist informativer?

Stellen wir uns vor, das zu Beginn dieses Kapitels geschilderte Experiment zum Modellernen mit drei experimentellen Bedingungen wäre (1) einmal mit einer »echten« Messwiederholung (intraindividuelle Bedingungsvariation: alle Personen durchlaufen alle Bedingungen des Faktors A), (2) einmal mit einer »Quasi-Messwiederholung« aufgrund einer Parallelisierung der Versuchspersonen anhand eines Vortests (interindividuelle Bedingungsvariation, wobei Personen mit gleichen Ausprägungen auf der Vortestvariable jeweils einer der drei experimentellen Bedingungen zugewiesen werden) und (3) ein drittes Mal ohne Messwiederholung (interindividuelle Bedingungsvariation mit randomisierter Zuweisung der Personen zu einer der drei Bedingungen) durchgeführt worden. Im ersten Fall ist eine hohe Zwischen-Personen-Quadratsumme QS_{zwp} zu erwarten; im zweiten Fall dürfte die QS_{zwp} kleiner sein, da die Abhängigkeit zwischen den Messwerten hier nur noch auf eine einzige Variable (nämlich diejenige, die im Vortest gemessen wurde) zurückzuführen ist; und im dritten Fall ist die QS_{zwp} definitionsgemäß gleich 0. Die QS_{tot} wäre in allen drei Fällen die gleiche, und auch die Haupteffekte der Be-

dingungen wären in allen drei Fällen exakt identisch. Das nicht-partielle Effektgrößenmaß $\hat{\eta}^2$ würde in allen drei Fällen also den gleichen Wert annehmen. Das partielle Effektgrößenmaß $\hat{\eta}_p^2$ hingegen wäre im ersten Fall größer als im zweiten Fall und dort wiederum größer als im dritten Fall, weil eben die Varianz, die auf konsistente Personenunterschiede zurückgeht (QS_{zwp}), hier nicht mit in den Ausdruck im Nenner eingeht und der Ausdruck im Nenner dementsprechend im ersten Fall kleiner ist als im zweiten Fall und dort wiederum kleiner als im dritten Fall.

Das Beispiel zeigt, dass mit dem partiellen Effektgrößenmaß $\hat{\eta}_p^2$ ein Problem verbunden ist: Untersucht man den gleichen Effekt mit unterschiedlichen Designs (intraindividuelle vs. interindividuelle Bedingungsvariation), dann unterscheidet sich das partielle Effektgrößenmaß $\hat{\eta}_p^2$ zwischen diesen Designs auch dann, wenn die Unterschiede zwischen den Bedingungsmittelwerten in den Designs exakt identisch sind. Unterschiedliche Untersuchungen, die zwar das Gleiche untersuchen, aber unterschiedliche Designs verwenden, sind hinsichtlich ihrer Effektgrößen nicht mehr miteinander vergleichbar. Das ist v.a. dann ein Problem, wenn man im Rahmen einer sog. Metaanalyse versucht, die Effektgrößen, die in vielen unterschiedlichen Primärstudien gefunden wurden, zusammenzufassen (Dunlap et al., 1996). Und dabei sollen standardisierte Effektgrößen ja gerade so definiert sein, dass sie auch über unterschiedliche Studien hinweg miteinander verglichen werden können.

Welches der beiden Effektgrößenmaße informativer ist, hängt auch von der konzeptuellen Bedeutung des Faktors ab, den man untersuchen will. Handelt es sich bei dem Faktor A um eine experimentell manipulierte Variable, die man sowohl mit einem intraindividuellen als auch mit einem interindividuellen Design untersuchen kann, sollte man beide Effektgrößenmaße berichten. Das nicht-partielle Effektgrößenmaß $\hat{\eta}^2$ erlaubt dann einen besseren Vergleich mit Studien, die mit einem interindividuellen Design gearbeitet haben. Handelt es sich bei dem Faktor A hingegen um die Zeit und ist man daran interessiert, wie viel Varianz in den Messwerten durch Veränderungen in der Merkmalsausprägung über die Zeit hinweg aufgeklärt wird, so ist das partielle Effektgrößenmaß $\hat{\eta}_p^2$ informativer, da Unterschiede in der Größe der Personeneffekte für die Schätzung des Veränderungseffekts irrelevant sind.

Statistikprogramme wie SPSS geben nur das partielle Effektgrößenmaß an; will man zusätzlich das nicht-partielle Effektgrößenmaß berichten, so muss man es anhand von Formel F 14.8 selbst berechnen.

Beispiel

Effektgrößenmaße für das Beispiel zum Modellernen

Wie groß ist der Effekt der experimentellen Manipulation (Faktor A) in unserem Datenbeispiel aus Tabelle 14.1? Zunächst berechnen wir das geschätzte nicht-partielle Effektgrößenmaß nach Formel F 14.8:

$$\eta^2 = QS_{zwA} / QS_{tot} = 2.920 / 4.532 = 0,64$$

Das partielle Effektgrößenmaß beträgt nach Formel F 14.9:

$$\eta_p^2 = QS_{zwA} / (QS_{zwA} + QS_{Res}) = 2.920 / 3.404 = 0,86$$

14.1.4 Populationsmodell der einfaktoriellen Varianzanalyse mit Messwiederholung

Das Populationsmodell der einfaktoriellen Varianzanalyse mit Messwiederholung besagt, dass ein Messwert beeinflusst wird durch

- ▶ den unbedingten Populationsmittelwert (μ),
- ▶ den Populationseffekt derjenigen Bedingung, unter der der Wert erhoben wurde (τ_j),
- ▶ Effekte, die auf Eigenschaften der jeweiligen Person zurückgehen (π_m),
- ▶ den bedingten Effekt der Bedingung, gegeben eine spezifische Person (Interaktion Person \times Bedingung; $(\pi\tau)_{mj}$), und
- ▶ alle unsystematischen Einflüsse einschließlich des Messfehlers (ε_{mj}).

Formal:

$$x_{mj} = \mu + \tau_j + \pi_m + (\pi\tau)_{mj} + \varepsilon_{mj} \quad (F 14.10a)$$

Da sich die Einflussgrößen $(\pi\tau)_{mj}$ und ε_{mj} mit diesem Design empirisch nicht trennen lassen, gehen wir im Folgenden davon aus, dass alle Interaktionseffekte $(\pi\tau)_{mj}$ gleich 0 sind. Damit verkürzt sich Formel F 14.10a wie folgt:

$$x_{mj} = \mu + \tau_j + \pi_m + \varepsilon_{mj} \quad (F 14.10b)$$

Diese Annahme vereinfacht die weitere Ableitung des Modells.

Haupteffekt der Bedingung a_j . Der Koeffizient τ_j wird – genau wie bei der einfaktoriellen Varianzanalyse ohne Messwiederholung – als Haupteffekt einer Bedingung a_j bezeichnet. Er ist definiert als die Abweichung eines Bedingungsmittelwerts ($\mu_{\cdot j}$) vom Gesamtmittelwert:

$$\tau_j = \mu_{\cdot j} - \mu \quad (F 14.11)$$

Diese Definition impliziert, dass die Summe der Haupteffekte über alle J Bedingungen hinweg immer 0 ergeben muss:

$$\sum_{j=1}^J \tau_j = 0 \quad (F 14.12)$$

Die Haupteffekte τ_j variieren nicht über Personen hinweg; sie sind für alle Personen in der Population konstant. Insofern trägt ihre Varianz auch nichts zur Varianz der Messwerte innerhalb einer Bedingung bei.

Haupteffekt der Person m . Der Koeffizient π_m kennzeichnet den unbedingten Effekt einer Person m . Hierunter fallen alle Merkmale dieser Person, die einen Einfluss auf die abhängige Variable haben und von der experimentellen Manipulation unabhängig sind, also über die Messungen hinweg stabil bleiben. Der Haupteffekt π_m ist definiert als die Abweichung eines Personenmittelwerts ($\mu_{m\cdot}$) vom Gesamtmittelwert:

$$\pi_m = \mu_{m\cdot} - \mu \quad (F 14.13)$$

Auch hier gilt, dass die Summe der Personen-Haupteffekte über alle n Personen hinweg immer 0 ergeben muss:

$$\sum_{m=1}^n \pi_m = 0 \quad (F 14.14)$$

Da Haupteffekte der Personen zufällige Effekte sind, lässt sich ihre Varianz (σ_π^2) nicht von vornherein kontrollieren; vielmehr handelt es sich um einen Populationsparameter, der aus den Daten geschätzt werden muss. Wichtig ist, dass die Personen-Haupteffekte einen Teil der Unterschiede zwischen den Personen (und damit einen Teil der Varianz innerhalb einer Bedingung) erklären.

Residuum. Wie wir gesehen haben, setzt sich das Residuum ε_{mj} aus allen unsystematischen Einflüssen wie z. B. dem Messfehler zusammen. Es ist derjenige Teil in der Variation der Messwerte, der weder durch Bedingungseffekte noch durch Personeneffekte erklärt werden kann. Setzt man die Formeln F 14.11 und F 14.13 in Formel F 14.10b ein und löst nach ε_{mj} auf, ergibt sich:

$$\begin{aligned}\varepsilon_{mj} &= x_{mj} - \mu - \tau_j - \pi_m \\ &= x_{mj} - \mu - (\mu_{\cdot j} - \mu) - (\mu_{m\cdot} - \mu) \\ &= x_{mj} - \mu - \mu_{\cdot j} + \mu - \mu_{m\cdot} + \mu \\ &= x_{mj} - \mu_{\cdot j} - \mu_{m\cdot} + \mu\end{aligned}\quad (\text{F 14.15})$$

Das Residuum ist also nichts anderes als die Abweichung der einzelnen Messwerte vom Gesamtmittelwert, nachdem sowohl der Bedingungs- als auch der Personeneffekt kontrolliert (herausgerechnet) wurden. Die Varianz dieses Residuums (σ_ε^2) ist ein Populationsparameter, der aus den Daten geschätzt werden muss; sie ist ein Teil der Varianz der Messwerte innerhalb der Bedingungen.

Varianz der Messwertvariablen

Unter der Annahme, dass das Modell in Formel F 14.10b gültig ist, lässt sich die Varianz der Messwertvariablen X_{mj} anhand der siebten Rechenregel für Varianzen (vgl. Formel F 7.35) zerlegen. Die Variable X_{mj} repräsentiert die potenziellen Werte einer zufällig gezogenen Person m in einer Bedingung a_j . Der individuelle Wert x_{mj} ist die Realisierung der Variablen X_{mj} in einer konkreten Studie. Wenden wir Formel F 7.35 auf unser Problem an, erhalten wir:

$$\begin{aligned}\text{Var}(X_{mj}) &= \text{Var}(\mu + \tau_j + \pi_m + \varepsilon_{mj}) \\ &= \text{Var}(\mu) + \text{Var}(\tau_j) + \text{Var}(\pi_m) \\ &\quad + \text{Var}(\varepsilon_{mj}) + \text{Cov}(\mu, \tau_j) \\ &\quad + \text{Cov}(\mu, \pi_m) + \text{Cov}(\mu, \varepsilon_{mj}) \\ &\quad + \text{Cov}(\tau_j, \pi_m) + \text{Cov}(\tau_j, \varepsilon_{mj}) \\ &\quad + \text{Cov}(\pi_m, \varepsilon_{mj})\end{aligned}\quad (\text{F 14.16a})$$

Der Gesamtmittelwert μ und die Bedingungs-Haupteffekte τ_j variieren nicht zwischen Personen; der Effekt einer Bedingung a_j ist für alle Personen in der Population gleich. Innerhalb einer Bedingung a_j gibt es also keine Varianz von τ_j . Bei τ_j handelt es sich somit innerhalb von a_j um eine Konstante.

Die Feststellung, dass μ und τ_j Konstanten sind, hat verschiedene Implikationen:

- ▶ Die beiden Ausdrücke $\text{Var}(\mu)$ und $\text{Var}(\tau_j)$ sind beide gleich 0.
- ▶ Alle Kovarianzen, an denen μ und τ_j beteiligt sind, sind gleich 0; denn wenn eine Variable eine Konstante ist, kann sie auch nicht mit anderen Variablen kovariieren (s. Abschn. 16.3.1).

Damit verkürzt sich Formel F 14.16a wie folgt:

$$\text{Var}(X_{mj}) = \text{Var}(\pi_m) + \text{Var}(\varepsilon_{mj}) + \text{Cov}(\pi_m, \varepsilon_{mj}) \quad (\text{F 14.16b})$$

Annahmen

Bei der einfaktoriellen Varianzanalyse mit Messwiederholung werden in Bezug auf die drei Größen auf der rechten Seite von Formel F 14.16b die folgenden drei zusätzlichen Annahmen getroffen (Fahrmeir et al., 1996a):

- (1) Die zufälligen Personeneffekte π_m sind unabhängig und identisch normalverteilt mit $N(0, \sigma_\pi^2)$.
- (2) Die Residuen ε_{mj} sind unabhängig und identisch normalverteilt mit $N(0, \sigma_\varepsilon^2)$.
- (3) Die Kovarianz der Personeneffekte und der Residuen ist gleich 0: $\text{Cov}(\pi_m, \varepsilon_{mj}) = 0$.

Aus diesen Annahmen folgt, dass die Varianz von X_{mj} der Varianz $\sigma_{X_j}^2$ entspricht mit

$$\sigma_{X_j}^2 = \sigma_\pi^2 + \sigma_\varepsilon^2. \quad (\text{F 14.17})$$

Bei Gültigkeit der Modellannahmen in der Population muss die Varianz des Merkmals in allen Faktorstufen identisch sein, nämlich $\sigma_\pi^2 + \sigma_\varepsilon^2$. Aufgrund der Annahme, dass die Variablen unabhängig und identisch verteilt sind, lassen wir im Folgenden bei der weiteren Betrachtung der Variablen des Populationsmodells aus Gründen der Vereinfachung den Index m für die Person weg. Mit X_j bezeichnen wir die Variable, deren Werte die individuellen Messwerte x_{mj} in einer Bedingung (bzw. einem Messzeitpunkt) a_j sind, π bezeichnet die Personeneffektvariable und ε_j die Residualvariable in der Bedingung a_j .

Kovarianz zwischen den Faktorstufen

Bei Designs mit intraindividueller Bedingungsvariation (Messwiederholung) sind die Messwerte zwischen den Faktorstufen nicht unabhängig voneinander, da sie von den gleichen Personen stammen. Das Ausmaß dieser Abhängigkeit kann über die Kovarianz quantifiziert werden. Das hatten wir bereits in Abschnitt 12.1.1

im Rahmen des t -Tests für abhängige Stichproben festgestellt, und in Abschnitt 16.3.1 werden wir die Kovarianz im Detail behandeln. Die Kovarianz ist dann positiv, wenn Individuen, die in einer Bedingung überdurchschnittliche Werte haben, auch in einer anderen Bedingung überdurchschnittliche Werte aufweisen und umgekehrt. In einem Design mit intraindividueller Bedingungsvariation und $J = 3$ Faktorstufen können demnach drei Kovarianzen berechnet werden: $\text{Cov}(X_1, X_2)$, $\text{Cov}(X_1, X_3)$ und $\text{Cov}(X_2, X_3)$. Bei intraindividuellen Designs gibt es also nicht nur Varianzen innerhalb der Faktorstufen, sondern auch Kovarianzen zwischen den Faktorstufen. Varianzen und Kovarianzen werden in Form einer Matrix (Varianz-Kovarianz-Matrix oder einfache *Kovarianzmatrix*) dargestellt. Eine Kovarianzmatrix wird mit dem griechischen Großbuchstaben Σ (Sigma) symbolisiert. Um deutlich zu machen, dass es sich um eine Matrix handelt, wird Σ fett geschrieben. Für den Fall, dass es $J = 3$ Faktorstufen gibt, sieht die Kovarianzmatrix der Variablen X_j wie folgt aus:

$$\Sigma_X = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \text{Cov}(X_2, X_3) \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \text{Var}(X_3) \end{pmatrix} \quad (\text{F 14.18})$$

Eine Kovarianzmatrix ist immer quadratisch und hat J Zeilen und J Spalten. In der Hauptdiagonale der Matrix stehen die Varianzen innerhalb einer jeweiligen Faktorstufe. In den restlichen Zellen der Matrix stehen die Kovarianzen zwischen zwei der J Faktorstufen.

In der einfaktoriellen Varianzanalyse mit Messwiederholung wird angenommen, dass die Residuen unterschiedlicher Faktorstufen voneinander unabhängig sind und daher alle eine Kovarianz von 0 aufweisen. Unter den Modellannahmen der einfaktoriellen Varianzanalyse mit Messwiederholung gibt es daher nur einen Grund, aus dem die Messwerte über die Bedingungen hinweg kovariieren: nämlich dass es stabile Personunterschiede gibt. Diese sind auf die Personen-Haupteffekte π_m zurückzuführen; sie manifestieren sich also in der Varianz σ_π^2 . Da die Personen-Haupteffekte über alle Messzeitpunkte hinweg konstant sind, müssen alle Kovarianzen gleich sein und der Personvarianz σ_π^2 entsprechen. Mit dieser Feststellung und der Zerlegung der Varianzen in Formel F 14.17 lässt sich die Kovarianzmatrix in F 14.18 wie folgt reformulieren:

$$\Sigma_X = \begin{pmatrix} \sigma_\pi^2 + \sigma_\varepsilon^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_\varepsilon^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_\varepsilon^2 \end{pmatrix} \quad (\text{F 14.19})$$

! Kovarianzstruktur des Modells der einfaktoriellen Varianzanalyse mit Messwiederholung

Aus den genannten Annahmen der einfaktoriellen Varianzanalyse mit Messwiederholung folgt, dass alle Faktorstufen in der Population eine konstante Varianz aufweisen, die der Summe aus der Personvarianz und der Residualvarianz entspricht, und dass die Kovarianz der Messwerte zwischen zwei beliebigen Faktorstufen der Personvarianz entspricht.

14.1.5 Schätzung der Populationsparameter

Es gibt vier Populationsparameter, die wir aus den Stichprobendaten schätzen müssen: den unbedingten Populationsmittelwert (μ), den Effekt einer experimentellen Bedingung (τ_j), den Effekt einer Person (π_m) bzw. die Personvarianz (σ_π^2) und das Residuum ε_{mj} bzw. die Populationsresidualvarianz (σ_ε^2).

Unbedingter Populationsmittelwert μ . Den unbedingten Populationsmittelwert μ schätzen wir wieder aus dem Gesamtmittelwert \bar{x} (s. auch Formel F 13.23):

$$\hat{\mu} = \bar{x} = \frac{\sum_{j=1}^J \sum_{m=1}^n x_{mj}}{J \cdot n} \quad (\text{F 14.20})$$

Bedingungs-Haupteffekte τ_j . Die Bedingungs-Haupteffekte τ_j schätzen wir aus den Differenzen der Bedingungsmittelwerte $\bar{x}_{\cdot j}$ vom Gesamtmittelwert \bar{x} (vgl. Formel F 13.24):

$$\hat{\tau}_j = \bar{x}_{\cdot j} - \bar{x} \quad (\text{F 14.21})$$

Personen-Haupteffekte π_m . Die Personen-Haupteffekte π_m schätzen wir aus den Abweichungen der Personenmittelwerte $\bar{x}_{m \cdot}$ vom Gesamtmittelwert \bar{x} :

$$\hat{\pi}_m = \bar{x}_{m \cdot} - \bar{x} \quad (\text{F 14.22})$$

Residuum ε_{mj} . Nachdem die Abweichungen $\bar{x}_{\cdot j} - \bar{x}$ und $\bar{x}_{m \cdot} - \bar{x}$ zur Schätzung der Populationsparameter τ_j und π_m verwendet werden, bleibt nur noch eine Quelle der

Variation übrig, die zur Schätzung des Residuums ε_{mj} verwendet werden kann (vgl. Formel F 14.15):

$$\begin{aligned}\hat{\varepsilon}_{mj} &= e_{mj} = (x_{mj} - \bar{x}) - (\bar{x}_{\cdot j} - \bar{x}) - (\bar{x}_{m\cdot} - \bar{x}) \\ &= x_{mj} - \bar{x}_{\cdot j} - \bar{x}_{m\cdot} + \bar{x}\end{aligned}\quad (\text{F 14.23})$$

14.1.6 Inferenzstatistik der einfaktoriellen Varianzanalyse mit Messwiederholung

Statistisches Hypothesenpaar

Die Nullhypothese, die bei der einfaktoriellen Varianzanalyse mit Messwiederholung typischerweise getestet wird, besagt, dass es einen Effekt des Faktors A in der Population *nicht* gibt und dass Abweichungen der Bedingungsmittelwerte vom Gesamtmittelwert in der Stichprobe nur zufällig, d.h. aufgrund eines Stichprobenfehlers, von 0 verschieden sind. In unserem Beispiel würde die Nullhypothese annehmen, dass alle Populationsmittelwerte identisch sind:

14

$$H_0: \mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} \quad (\text{F 14.24a})$$

Oder alternativ:

$$H_0: \mu_{\cdot j} - \mu = 0 \quad \text{für alle } j \quad (\text{F 14.24b})$$

Die Alternativhypothese besagt, dass es zwischen mindestens zwei Populationsmittelwerten einen Unterschied gibt, also:

$$H_1: \mu_{\cdot j} - \mu \neq 0 \quad \text{für mindestens ein } j \quad (\text{F 14.25})$$

Um die Hypothesen überprüfen zu können, greifen wir wiederum auf eine *F*-verteilte Prüfgröße zurück, die zwei unabhängige Schätzungen der Residualvarianz zueinander ins Verhältnis setzt, deren Erwartungswerte unter der Gültigkeit der Nullhypothese gleich sind (Abschn. 13.1.7). Hierzu definieren wir zunächst wieder für jede Quadratsumme (QS) eine mittlere Quadratsumme (MQS), indem wir die Quadratsumme durch ihre Freiheitsgrade teilen. Wir werden daher zunächst die Zerlegung der Freiheitsgrade behandeln und darauf aufbauend die Definition der mittleren Quadratsummen.

Zerlegung der Freiheitsgrade

Die Freiheitsgrade zu den drei Quadratsummen bestimmen wir nach der gleichen Systematik wie bei der zweifaktoriellen Varianzanalyse ohne Messwiederholung (vgl. die Formeln F 13.120 bis F 13.122):

$$df_{zwP} = n - 1 \quad (\text{F 14.26})$$

$$df_{zwA} = J - 1 \quad (\text{F 14.27})$$

$$df_{\text{Res}} = (n - 1) \cdot (J - 1) \quad (\text{F 14.28})$$

Auch hier gilt, dass die totalen Freiheitsgrade der Summe dieser drei Freiheitsgrade entsprechen:

$$df_{\text{tot}} = df_{zwP} + df_{zwA} + df_{\text{Res}} \quad (\text{F 14.29})$$

Für unser Beispiel in Tabelle 14.1 ermitteln wir $df_{zwP} = 4$, $df_{zwA} = 2$ und $df_{\text{Res}} = 8$. Bei 15 Messwerten gibt es $df_{\text{tot}} = 14$ totale Freiheitsgrade; dies entspricht genau der Summe der drei Freiheitsgrade df_{zwP} , df_{zwA} und df_{Res} .

Mittlere Quadratsummen

Teilt man die Quadratsummen durch die Freiheitsgrade, erhält man mittlere Quadratsummen (MQS). Aufgrund der Tatsache, dass additive Zählerterme durch unterschiedlich große Nennerterme geteilt werden, sind diese Quotienten nicht länger additiv. Bei der einfaktoriellen Varianzanalyse mit Messwiederholung berechnen sich die mittleren Quadratsummen wie folgt:

$$MQS_{zwP} = \frac{QS_{zwP}}{df_{zwP}} = \frac{J \cdot \sum_{m=1}^n (\bar{x}_{m\cdot} - \bar{x})^2}{n - 1} \quad (\text{F 14.30})$$

$$MQS_{zwA} = \frac{QS_{zwA}}{df_{zwA}} = \frac{n \cdot \sum_{j=1}^J (\bar{x}_{\cdot j} - \bar{x})^2}{J - 1} \quad (\text{F 14.31})$$

$$MQS_{\text{Res}} = \frac{QS_{\text{Res}}}{df_{\text{Res}}} = \frac{\sum_{j=1}^J \sum_{m=1}^n (x_{mj} - \bar{x}_{\cdot j} - \bar{x}_{m\cdot} + \bar{x})^2}{(n - 1) \cdot (J - 1)} \quad (\text{F 14.32})$$

Für das Datenbeispiel in Tabelle 14.1 berechnen wir folgende mittlere Quadratsummen: $MQS_{zwP} = 1.128/4 = 282$; $MQS_{zwA} = 2.920/2 = 1.460$ und $MQS_{\text{Res}} = 484/8 = 60,5$.

Erwartungswerte der mittleren Quadratsummen

Um die Logik des *F*-Tests bei der einfaktoriellen Varianzanalyse mit Messwiederholung zu verstehen, müssen wir uns zunächst vergegenwärtigen, wie die Erwartungswerte der mittleren Quadratsummen bei Geltung der statistischen Null- bzw. der statistischen Alternativhypothese aussehen. Dabei werden wir nicht so detailliert vorgehen wie bei der einfaktoriellen Varianzana-

lyse ohne Messwiederholung (Abschn. 13.1.6), da die Herleitung prinzipiell identisch ist.

Erwartungswert der MQS_{Res} . Der Erwartungswert der MQS_{Res} entspricht der Populationsresidualvarianz:

$$E(MQS_{\text{Res}}) = \sigma_\varepsilon^2 \quad (\text{F 14.33})$$

Daher ist die MQS_{Res} ein erwartungstreuer Schätzer der Populationsresidualvarianz.

Erwartungswert der MQS_{ZWP} . Es lässt sich zeigen, dass die MQS_{ZWP} zwar die wahre Personvarianz beinhaltet, dass aber die Varianz der Personmittelwerte ihrerseits durch die Residualvarianz beeinflusst ist und insofern die MQS_{ZWP} neben der wahren Personvarianz auch noch Messfehler- und andere unsystematische Einflüsse in sich trägt. Der Erwartungswert der MQS_{ZWP} berechnet sich also wie folgt:

$$E(MQS_{\text{ZWP}}) = \sigma_\varepsilon^2 + J \cdot \sigma_\pi^2 \quad (\text{F 14.34})$$

Erwartungswert der MQS_{ZWA} . Auch für die MQS_{ZWA} gilt, dass sie von den wahren Bedingungseffekten abhängt, dass aber die Varianz der Bedingungsmittelwerte ihrerseits durch die Residualvarianz beeinflusst ist und insofern die MQS_{ZWA} neben den wahren Bedingungseffekten auch noch Messfehler- und andere unsystematische Einflüsse in sich trägt:

$$E(MQS_{\text{ZWA}}) = \sigma_\varepsilon^2 + \frac{n}{J-1} \cdot \sum_{j=1}^J \tau_j^2 \quad (\text{F 14.35})$$

Erwartungswert der MQS_{ZWA} bei Geltung der Nullhypothese. Wenn die Nullhypothese ($\mu_{\cdot j} - \mu = 0$ für alle j)

gilt, sind alle Effekte $\tau_j = 0$. In diesem Fall ist die Summe der quadrierten Bedingungseffekte gleich 0, und Gleichung F 14.35 würde sich wie folgt verkürzen:

$$E(MQS_{\text{ZWA}}) = \sigma_\varepsilon^2 \quad (\text{F 14.36})$$

Im Klartext: Bei Geltung der Nullhypothese hat die MQS_{ZWA} den gleichen Erwartungswert wie die MQS_{Res} . Beide wären dann erwartungstreue Schätzer der Populationsresidualvarianz.

Erwartungswert der MQS_{ZWA} bei Geltung der Alternativhypothese.

Wenn die Alternativhypothese ($\mu_{\cdot j} - \mu \neq 0$ für mind. ein j) gilt, dann ist bei mindestens einer Bedingung $\tau_j \neq 0$. In diesem Fall ist die Summe der quadrierten Bedingungseffekte größer 0, und der Erwartungswert der MQS_{ZWA} ist entsprechend größer als der Erwartungswert der MQS_{Res} .

F-Test

Die Prüfgröße zur Überprüfung der Nullhypothese ($\mu_{\cdot j} - \mu = 0$ für alle j) ist folgerichtig der Quotient aus MQS_{ZWA} und MQS_{Res} . Je größer die Bedingungs-Haupteffekte, desto größer wird die MQS_{ZWA} im Vergleich zur MQS_{Res} . Der Wert der Prüfgröße berechnet sich wie folgt:

$$F = \frac{MQS_{\text{ZWA}}}{MQS_{\text{Res}}} \quad (\text{F 14.37})$$

Dieser Quotient folgt einer F -Verteilung, die durch ihre Zähler- und Nennerfreiheitsgrade bestimmt wird. Die Freiheitsgrade der MQS_{ZWA} betragen gemäß Formel F 14.27 $df_1 = df_{\text{ZWA}} = J - 1$. Die Freiheitsgrade der MQS_{Res} betragen gemäß Formel F 14.28 $df_2 = df_{\text{Res}} = (n - 1) \cdot (J - 1)$. Ist der empirische F -Wert mindestens so groß wie der kritische F -Wert bei gegebenem Signifikanzniveau α , so wird die Nullhypothese verworfen.

Beispiel

F-Test im Beispiel zum Modellernen

Für unser Datenbeispiel in Tabelle 14.1 haben wir folgende mittlere Quadratsummen ermittelt: $MQS_{\text{ZWA}} = 1.460$ und $MQS_{\text{Res}} = 60,5$. Der Quotient aus diesen beiden mittleren Quadratsummen beträgt $F = 24,13$. Aus Tabelle A.8 im Anhang A oder aus einem Verteilungsrechner lesen wir ab, dass der kritische F -Wert bei $df_1 = df_{\text{ZWA}} = 2$ Zähler- und $df_2 = df_{\text{Res}} = 8$ Nennerfreiheitsgraden bei einem Signifikanzniveau

von $\alpha = 5\% F_{(0,95; 2; 8)} = 4,46$ beträgt. Unser Wert liegt darüber. Die Nullhypothese wird also abgelehnt. Zwischen mindestens zwei der drei experimentellen Bedingungen gibt es einen signifikanten Unterschied im Nachahmungsverhalten.

Das Ergebnis dieser Varianzanalyse kann auch hier in einer Tabelle zusammengefasst werden. Tabelle 14.2 zeigt, wie eine solche Zusammenfassung – bezogen

auf unser Datenbeispiel – aussieht. Die Tabelle sieht ähnlich aus wie Tabelle 13.4 (in Abschn. 13.1.7), mit zwei Unterschieden: Erstens ist hier eine weitere Zeile eingefügt, in der der Einfluss des Faktors »Person« ab-

getragen ist; zweitens ist eine weitere Spalte eingefügt, in der neben dem nicht-partiellen Effektgrößenschätzer $\hat{\eta}^2$ das partielle Effektgrößenmaß $\hat{\eta}_p^2$ abgetragen ist.

Tabelle 14.2 Ergebnistabelle einer einfaktoriellen Varianzanalyse mit Messwiederholung (Datenbeispiel aus Tab. 14.1)

Quelle der Variation	QS	df	MQS	F	p	$\hat{\eta}^2$	$\hat{\eta}_p^2$
Faktor A	2.920	2	1.460	24,13	0,0004	0,64	0,86
Person	1.128	4	282				
Residuum	484	8	60,5				
Total	4.532	14	323,71				

14.1.7 Sphärizität und Compound Symmetry

Wir haben die drei Annahmen der einfaktoriellen Varianzanalyse mit Messwiederholung bereits in Abschnitt 14.1.4 eingeführt und dabei gesehen, dass die Kovarianzmatrix eine spezifische Struktur aufweisen muss: Die Varianzen müssen in allen Faktorstufen homogen und die Kovarianzen für alle beliebigen Paare von Faktorstufen identisch sein. Eine Kovarianzmatrix, die diese Struktur aufweist, wird Compound-Symmetry-Matrix (oder CS-Matrix) genannt. Man kann sich vorstellen, dass eine solche CS-Struktur eine recht strenge Annahme darstellt, insbesondere wenn es sich bei den unabhängigen Variablen um den Faktor »Zeit« handelt und die Abstände zwischen den Messzeitpunkten groß sind. Die Kovarianz zwischen zwei Messzeitpunkten ist typischerweise umso größer, je näher die Messzeitpunkte beieinanderliegen. Je weiter sie auseinanderliegen, desto größer wird der Einfluss person-spezifischer Veränderungen und desto schwächer der Einfluss stabiler Personunterschiede.

Huynh und Feldt (1976) konnten zeigen, dass der F-Test der einfaktoriellen Varianzanalyse mit Messwiederholung auch dann zu robusten Ergebnissen kommt, wenn die Kovarianzmatrix keine CS-Struktur, sondern lediglich eine sphärische Struktur aufweist.

Sphärizität (oder Zirkularität)

Huynh und Feldt (1976) zufolge ist es ausreichend, wenn die Populations-Kovarianzmatrix der Faktorstufen sphärisch (oder zirkulär) ist. Sphärizität bedeutet,

dass die Varianzen aller möglichen Differenzvariablen gleich sind. Differenzwerte erhält man, wenn man zwei Messwerte der gleichen Person aus zwei der J Faktorstufen voneinander abzieht. Bei drei Faktorstufen lassen sich somit drei Differenzvariablen berechnen (bei vier Faktorstufen wären es bereits sechs Differenzvariablen, bei fünf Faktorstufen wären es zehn Differenzvariablen). Ist Sphärizität gegeben, haben alle Differenzvariablen die gleiche Varianz.

Die Sphärizitätsannahme ist weniger streng als die Annahme gleicher Varianzen und Kovarianzen; formal ist Compound Symmetry ein Spezialfall von Sphärizität. Aber auch die Sphärizitätsannahme ist schnell verletzt. Dies ist häufig der Fall, wenn Personen in größeren zeitlichen Abständen wiederholt untersucht werden. Ist Sphärizität nicht gegeben, wird der F-Test zu liberal, d.h., die Wahrscheinlichkeit eines α -Fehlers ist größer als das vorher festgelegte Signifikanzniveau (Box, 1954).

Box-Epsilon. Box (1954) hat einen Index entwickelt, der anzeigt, wie stark eine Kovarianzmatrix in der Population von der Sphärizität abweicht. Dieser Index wird mit ε_{Box} bezeichnet. Er kann Werte zwischen $1/(J - 1)$ und 1 annehmen. Eine Matrix, die die Sphärizitätsannahme perfekt erfüllt, weist ein $\varepsilon_{\text{Box}} = 1$ auf. Je stärker die Sphärizitätsannahme verletzt ist, umso mehr weicht der ε_{Box} -Wert von 1 ab. Ist Sphärizität nicht gegeben, erlaubt der ε_{Box} -Wert eine Korrektur der Zähler- und der Nennerfreiheitsgrade für die Bestimmung des kritischen F-Werts. Da der Wert für ε_{Box} in der Population üblicherweise nicht bekannt ist, muss er aus den

Daten geschätzt werden. Hierzu gibt es verschiedene Methoden, die zu unterschiedlichen Schätzungen und entsprechenden Korrekturen der Freiheitsgrade führen.

Greenhouse-Geisser-Epsilon. Eine Möglichkeit, ε_{Box} zu schätzen, haben Greenhouse und Geisser (1959) vorgestellt. Hierbei werden die Elemente der geschätzten Kovarianzmatrix $\hat{\Sigma}_X$ miteinander verrechnet. Der geschätzte $\hat{\varepsilon}_{\text{GG}}$ -Wert nach Greenhouse und Geisser (1959) kann ebenfalls zwischen $1/(J-1)$ und 1 variieren. Bei $J=3$ Faktorstufen beträgt die Untergrenze also 0,5. Je näher der $\hat{\varepsilon}_{\text{GG}}$ -Wert an dem Wert 1 liegt, desto eher ist die Sphärizitätsannahme erfüllt. Im Allgemeinen gilt die Sphärizitätsannahme als verletzt, wenn $\hat{\varepsilon}_{\text{GG}}$ kleiner ist als 0,75.

Mauchly-Test. Eine Möglichkeit, die Abweichung der empirischen Kovarianzmatrix von einer perfekt sphärischen Matrix zu testen, stellt der Mauchly-Test dar (Mauchly, 1940). Die Teststatistik basiert auf einem Vergleich zwischen der empirischen Kovarianzmatrix (bzw. ihrer Determinante und ihrer Spur, s. hierzu Anhang B) und einer sphärischen Matrix. Die Mauchly-Statistik ist approximativ χ^2 -verteilt mit $J-1$ Freiheitsgraden. Ist der empirische χ^2 -Wert mindestens so groß wie der kritische Wert auf einem vorher festgelegten Signifikanzniveau α , so muss die Annahme, dass es sich bei der empirischen Kovarianzmatrix um eine sphärische Matrix handelt, verworfen werden. Es hat sich allerdings gezeigt, dass der Mauchly-Test zu streng wird (d.h. zu oft zur Entscheidung führt, die Sphärizitätsannahme abzulehnen), wenn die Normalverteilungsannahme verletzt ist. Von daher sollte das Ergebnis des Mauchly-Tests nur mit Vorsicht interpretiert werden.

Beispiel

Compound Symmetry und Sphärizität im Beispiel zum Modelllernen

Wie sieht die geschätzte Kovarianzmatrix ($\hat{\Sigma}_X$) für unser Datenbeispiel in Tabelle 14.1 aus, und wie müsste eine modellkonforme Kovarianzmatrix in Anlehnung an Formel F 14.19 aussehen? Um die geschätzte Kovarianzmatrix vollständig angeben zu können, müssen wir zunächst die Varianzen innerhalb der Faktorstufen und die Kovarianzen zwischen den Faktorstufenpaaren berechnen. Die Stichprobenvarianzen können wir gemäß Formel F 8.11 berechnen; sie betragen hier $\hat{\sigma}_{X_1}^2 = 129$, $\hat{\sigma}_{X_2}^2 = 144$ und $\hat{\sigma}_{X_3}^2 = 130$. Die Kovarianz werden wir in Abschnitt 16.3.1 (Formel F 16.2) und Abschnitt 16.4.1 (Formel F 16.31) genauer kennenlernen; wir wollen sie hier trotzdem schon einmal vorwegnehmen. Für zwei Variablen X_1 und X_2 lautet die Formel für die Stichprobenkovarianz:

$$\hat{\sigma}_{X_1 X_2} = \frac{\sum_{m=1}^n (x_{m1} - \bar{x}_{\bullet 1}) \cdot (x_{m2} - \bar{x}_{\bullet 2})}{n-1} \quad (\text{F 14.38})$$

Die Kovarianzen betragen in unserem Beispiel $\hat{\sigma}_{X_1 X_2} = 130,25$; $\hat{\sigma}_{X_1 X_3} = 51,25$ und $\hat{\sigma}_{X_2 X_3} = 40$. In Matrixschreibweise lassen sich die Varianzen und Kovarianzen in unserem Datenbeispiel aus Tabelle 14.1 wie folgt darstellen:

$$\hat{\Sigma}_X = \begin{pmatrix} 129 & 130,25 & 51,25 \\ 130,25 & 144 & 40 \\ 51,25 & 40 & 130 \end{pmatrix}$$

Unter Geltung des Modells in Formel F 14.10b müssten die Varianzen sowie die Kovarianzen in der Population homogen sein. Wie müsste eine solche modellkonforme Matrix mit CS-Struktur aussehen? Aus Formel F 14.19 wissen wir, dass in der Hauptdiagonalen der Ausdruck $\sigma_{\pi}^2 + \sigma_{\varepsilon}^2$ und in allen anderen Zellen die Personvarianz σ_{π}^2 stehen müsste. Nun wissen wir aus Formel F 14.34, dass die MQS_{ZWP} ein erwartungstreuer Schätzer des Ausdrucks $\sigma_{\varepsilon}^2 + J \cdot \sigma_{\pi}^2$ ist. Setzen wir in Formel F 14.34 nun die Schätzer für die Populationsvarianzen ein und lösen nach $\hat{\sigma}_{\pi}^2$ auf, ergibt sich:

$$\hat{\sigma}_{\pi}^2 = \frac{MQS_{\text{ZWP}} - \hat{\sigma}_{\varepsilon}^2}{J} \quad (\text{F 14.39})$$

Ferner wissen wir aus Formel F 14.33, dass die MQS_{Res} ein erwartungstreuer Schätzer der Populationsresidualvarianz σ_{ε}^2 ist. Setzen wir also nun in Formel F 14.39 für $\hat{\sigma}_{\varepsilon}^2$ den Wert $MQS_{\text{Res}} = 60,5$ ein, ergibt sich bei $J=3$ für die geschätzte Personvarianz $\hat{\sigma}_{\pi}^2 = (282 - 60,5)/3 = 73,83$. Dies entspricht gleichzeitig der vom Modell geschätzten Kovarianz zwischen zwei beliebigen Faktorstufen. In die Hauptdiagonale setzen wir den Ausdruck $\hat{\sigma}_{\pi}^2 + \hat{\sigma}_{\varepsilon}^2 = \hat{\sigma}_{\pi}^2 + MQS_{\text{Res}} = 73,83 + 60,5 = 134,33$ ein. In Anlehnung an Formel F 14.19 erhalten wir die geschätzte CS-Matrix:

$$\hat{\Sigma}_X = \begin{pmatrix} 134,33 & 73,83 & 73,83 \\ 73,83 & 134,33 & 73,83 \\ 73,83 & 73,83 & 134,33 \end{pmatrix}$$

Hier handelt es sich um eine Matrix mit homogenen Varianzen und homogenen Kovarianzen, also eine CS-Struktur. Man sieht, dass unsere aus den Daten geschätzte Kovarianzmatrix von der geschätzten CS-Kovarianzmatrix abweicht: Sowohl die Varianzen in den drei experimentellen Bedingungen als auch die Kovarianzen zwischen zwei der drei Bedingungen unterscheiden sich relativ stark.

Als Nächstes testen wir, wie stark unsere aus den Daten geschätzte Kovarianzmatrix von einer sphärischen Matrix abweicht. Hierzu inspizieren wir den Epsilon-Index von Greenhouse und Geisser (1959) und das Ergebnis des Mauchly-Tests mithilfe eines Statistikprogramms (z. B. SPSS). Wir erhalten für

den Epsilon-Index nach Greenhouse-Geisser den Wert $\hat{\epsilon}_{GG} = 0,545$. Das ist schon recht nah an der Untergrenze, die $\hat{\epsilon}_{GG}$ bei $J = 3$ Faktorstufen haben kann, nämlich 0,5. Die Abweichung von der sphärischen Matrix ist also recht groß. Die Mauchly-Statistik beträgt 0,164. Für die Prüfgröße des Mauchly-Tests erhalten wir den Wert 5,425. Dieser Wert schneidet unter einer χ^2 -Verteilung mit $df = 2$ Freiheitsgraden einen Flächenanteil von $p = 0,07$ nach rechts ab. Die Nullhypothese, der zufolge es sich bei unserer Kovarianzmatrix um eine sphärische Matrix handelt, muss zwar auf einem Signifikanzniveau von $\alpha = 5\%$ nicht verworfen werden; es sollte allerdings hinzugefügt werden, dass der Mauchly-Test bei $n = 5$ Personen auch nur eine sehr geringe Power hat. Die Wahrscheinlichkeit, die Nullhypothese der Sphärizität selbst bei großen Abweichungen von einer sphärischen Matrix (wie sie in unserem Beispiel gegeben ist) abzulehnen, ist also gering.

14

Was tun, wenn die Sphärizitätsannahme verletzt ist?

Der F -Test ist gegenüber einer Verletzung der Sphärizitätsannahme nicht so robust wie gegenüber einer Verletzung der Normalverteilungsannahme (Box, 1954). Ist die Sphärizitätsannahme verletzt, sollte der F -Test nur mit Vorsicht interpretiert werden, da er dann zu liberal wird. Eine Möglichkeit besteht daher darin, den F -Test strenger zu machen, etwa indem man die Freiheitsgrade zur Bestimmung des kritischen F -Wertes reduziert und somit den kritischen Wert erhöht. Dies kann auf zwei Wegen geschehen.

Greenhouse-Geisser-Korrektur. Eine Möglichkeit, die Freiheitsgrade des kritischen F -Wertes zu korrigieren, besteht darin, sie mit dem $\hat{\epsilon}_{GG}$ -Index nach Greenhouse und Geisser (1959) zu multiplizieren. Da $\hat{\epsilon}_{GG}$ bei perfekter Sphärizität den Wert 1 annimmt, findet in diesem Fall keine Korrektur statt. Je weiter die Kovarianzmatrix von einer sphärischen Matrix abweicht, desto kleiner wird $\hat{\epsilon}_{GG}$. Das heißt für die Korrektur der Freiheitsgrade: Je weiter die Abweichung von der Sphärizität, desto mehr werden die Freiheitsgrade nach unten korrigiert.

Huynh-Feldt-Korrektur. In Simulationsstudien hat sich gezeigt, dass die Greenhouse-Geisser-Korrektur recht

streng ist und dazu führt, dass die Nullhypothese der einfaktoriellen Varianzanalyse mit Messwiederholung zu selten abgelehnt wird. Daher haben Huynh und Feldt (1976) eine alternative Schätzformel und eine darauf aufbauende Korrekturformel vorgeschlagen. Das nach Huynh-Feldt geschätzte Epsilon bezeichnen wir mit $\hat{\epsilon}_{HF}$. Die beiden Schätzungen von ϵ_{Box} stehen in folgender Beziehung zueinander:

$$\hat{\epsilon}_{HF} = \frac{n \cdot (J - 1) \cdot \hat{\epsilon}_{GG} - 2}{(J - 1) \cdot ((n - 1) - (J - 1) \cdot \hat{\epsilon}_{GG})} \quad (\text{F 14.40})$$

Der Wert für $\hat{\epsilon}_{HF}$ ist größer als der Wert für $\hat{\epsilon}_{GG}$. Multipliziert man die Freiheitsgrade des kritischen F -Wertes mit $\hat{\epsilon}_{HF}$ anstatt mit $\hat{\epsilon}_{GG}$, fällt die Korrektur also schwächer aus, was in den meisten Fällen dazu führt, dass der kritische F -Wert nicht so stark nach oben korrigiert wird wie bei der Greenhouse-Geisser-Korrektur.

Beispiel

Korrektur der Freiheitsgrade im Beispiel zum Modellernen

Zu welchem Ergebnis kommen wir, wenn wir in unserem Beispiel aus Tabelle 14.1 bzw. Tabelle 14.2 die Freiheitsgrade des kritischen F -Wertes nach Greenhouse-Geisser bzw. Huynh-Feldt korrigieren? Für den

geschätzten Epsilon-Index nach Greenhouse-Geisser erhalten wir mit SPSS einen Wert von $\hat{\epsilon}_{GG} = 0,545$. Multiplizieren wir beide Freiheitsgrade mit diesem Wert, so ergeben sich $df_{\text{Zähler_GG}} = 2 \cdot 0,545 = 1,09$ und $df_{\text{Nenner_GG}} = 8 \cdot 0,545 = 4,36$. Der kritische F -Wert bei einem Signifikanzniveau von $\alpha = 5\%$ beträgt in diesem Fall $F_{(0,95; 1,09; 4,36)} = 7,12$. Für den Epsilon-Index nach Huynh und Feldt ermitteln wir gemäß Formel F 14.40 einen Wert von $\hat{\epsilon}_{HF} = 0,593$. Multiplizieren wir die Freiheitsgrade mit $\hat{\epsilon}_{HF}$, so ergeben sich $df_{\text{Zähler_HF}} = 2 \cdot 0,593 = 1,19$ und $df_{\text{Nenner_HF}} = 8 \cdot 0,593 = 4,74$. Der kritische F -Wert bei einem Signifikanzniveau von $\alpha = 5\%$ beträgt in diesem Fall $F_{(0,95; 1,19; 4,74)} = 6,61$. Die Korrektur der Freiheitsgrade hat den F -Test (unabhängig von der gewählten Korrekturmethode) zwar strenger gemacht, aber die Nullhypothese ist in unserem Beispiel dennoch zu verwerfen, da der empirische F -Wert (24,13) immer noch größer ist als der korrigierte kritische Wert.

In den meisten Statistikprogrammen werden die Ergebnisse des F -Tests auf der Basis beider Korrekturmöglichkeiten automatisch mit ausgegeben, sodass eine händische Korrektur der Freiheitsgrade nicht nötig ist.

Der Vollständigkeit halber sei noch gesagt, dass es auch Auswertungsprozeduren gibt, die Sphärizität nicht voraussetzen. So werden der Kovarianzstruktur z. B. bei der multivariaten Varianzanalyse (MANOVA, s. Kap. 15) keine Restriktionen auferlegt, wenn man sie zur Auswertung eines einfaktoriellen Plans mit Messwiederholung einsetzt (Stevens, 2009). Ist die Sphärizitätsannahme erfüllt, führt die Anwendung der MANOVA zu einem Verlust an Teststärke, sodass in diesem Fall die einfaktorielle Varianzanalyse mit Messwiederholung verwendet werden sollte. Ist die Sphärizitätsannahme verletzt, kann auf die MANOVA oder die vorgestellten Korrekturen zurückgegriffen werden. Allerdings sollte die MANOVA nach Empfehlungen von Maxwell und Delaney (2004) nicht angewendet werden, wenn $n < J + 10$ ist. Eine weitere Möglichkeit besteht darin, hierarchische lineare Modelle (HLM; s. Abschn. 20.4) zur Auswertung längsschnittlicher Daten zu verwenden (z. B. Hosoya et al., 2014; Singer & Willett, 2003).

14.1.8 Effektgrößenmaße und Konfidenzintervalle

Wie wir bereits in Abschnitt 14.1.3 gesehen haben, gibt es bei der einfaktoriellen Varianzanalyse mit Messwiederholung zwei Effektgrößenmaße, das nicht-partielle η^2 und das partielle η_p^2 .

Effektgröße η^2

Die (nicht-partielle) Effektgröße η^2 lässt sich wie folgt definieren:

$$\eta^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_\pi^2 + \sigma_\varepsilon^2} \quad \text{mit} \quad \sigma_A^2 = \frac{1}{J} \cdot \sum_{j=1}^J \tau_{aj}^2 \quad (\text{F 14.41})$$

Diese Effektgrößen kann über $\hat{\eta}^2$ nach Gleichung F 14.8 aus den Daten geschätzt werden.

Wir haben bereits in Abschnitt 13.1.9 darauf hingewiesen, dass $\hat{\eta}^2$ kein erwartungstreuer Schätzer für η^2 ist, und daher den Effektgrößenschätzer $\hat{\omega}^2$ eingeführt. Dieser lautet für die einfaktorielle Varianzanalyse mit Messwiederholung (Olejnik & Algina, 2000; s. auch Formel F 13.52):

$$\hat{\omega}^2 = \frac{df_{zwA} \cdot (MQS_{zwA} - MQS_{Res})}{QS_{tot} + MQS_{zwP}} \quad (\text{F 14.42})$$

Effektgröße η_p^2

Die partielle Effektgröße η_p^2 lässt sich wie folgt definieren:

$$\eta_p^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_\varepsilon^2} \quad \text{mit} \quad \sigma_A^2 = \frac{1}{J} \cdot \sum_{j=1}^J \tau_{aj}^2 \quad (\text{F 14.43})$$

Ihr Wert lässt sich nach Gleichung F 14.9 schätzen. Der entsprechende erwartungstreue partielle Effektgrößenschätzer $\hat{\omega}_p^2$ lautet (Olejnik & Algina, 2000):

$$\hat{\omega}_p^2 = \frac{df_{zwA} \cdot (MQS_{zwA} - MQS_{Res})}{QS_{zwA} + MQS_{Res} \cdot (n - df_{zwA})} \quad (\text{F 14.44})$$

Intraklassen-Korrelation

Bei der einfaktoriellen Varianzanalyse mit Messwiederholung ist die Einteilung empirisch ermittelter Effekte in »große«, »mittlere« oder »kleine« etwas komplizierter als bei der Varianzanalyse ohne Messwiederholung. Der Grund ist der zuvor diskutierte: Wie groß ein Effekt ist, hängt auch davon ab, wie groß derjenige Anteil der Varianz in den Messwerten ist, der auf stabile

interindividuelle Unterschiede zurückgeht. Der Anteil der Personvarianz an der Gesamtvarianz wird auch als Intraklassen-Korrelation (abgekürzt ρ ; griech. Kleinbuchstabe »rho«) bezeichnet. Sie ist definiert als der Anteil der Personvarianz (σ_π^2) an der Summe der Personvarianz und der Residualvarianz (σ_ϵ^2):

$$\rho = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\epsilon^2} \quad (\text{F 14.45})$$

Wie wir anhand von Gleichung F 14.19 sehen können, entspricht die Intraklassen-Korrelation bei der einfaktoriellen Varianzanalyse mit Messwiederholung der Korrelation zwischen zwei Messzeitpunkten. Wir werden in Kapitel 20 näher auf die Intraklassen-Korrelation eingehen.

Wie kann man die Intraklassen-Korrelation empirisch bestimmen? Dazu müssen wir zunächst wissen, wie die Personvarianz und die Residualvarianz aus den Daten geschätzt werden können. Der Schätzer der Personvarianz wird nach Formel F 14.39 berechnet:

$$\hat{\sigma}_\pi^2 = \frac{MQS_{ZWP} - MQS_{Res}}{J}$$

14

Der Schätzer für die Residualvarianz ist MQS_{Res} (Formel F 14.33). Also ergibt sich für die geschätzte Intraklassen-Korrelation:

$$\hat{\rho} = \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\pi^2 + \hat{\sigma}_\epsilon^2} = \frac{MQS_{ZWP} - MQS_{Res}}{MQS_{ZWP} + (J-1) \cdot MQS_{Res}} \quad (\text{F 14.46})$$

Kleine, mittlere und große Effekte

Wie kann man nun diese Intraklassen-Korrelation zur Beurteilung der Effektgröße bei der einfaktoriellen Varianzanalyse mit Messwiederholung angemessen berücksichtigen? Eine Möglichkeit besteht darin, die Taxonomie zur Beurteilung von Effektgrößen für die einfaktorielle Varianzanalyse ohne Messwiederholung nach Cohen (1988) heranzuziehen und diese um den Einfluss der Intraklassen-Korrelation zu korrigieren. Ist die Annahme, dass alle Varianzen der J Bedingungen identisch sind und alle Kovarianzen zwischen den Bedingungspaaren identisch sind (die Varianz-Kovarianz-Matrix also eine CS-Struktur aufweist), erfüllt, dann hängt die Effektgröße ϕ'^2 einer Varianzanalyse mit Messwiederholung mit dem Effektgrößenmaß η^2 einer Varianzanalyse ohne Messwiederholung wie folgt zusammen:

$$\phi'^2 = \frac{\eta^2}{(1-\rho) \cdot (1-\eta^2)} \quad (\text{F 14.47})$$

Tabelle 14.3 Klassifikation von Effektgrößen (η'^2) bei Varianzanalysen mit Messwiederholung. Dem Vorschlag liegt die Taxonomie von Cohen (1988) für Effektgrößen bei der Varianzanalyse ohne Messwiederholung (η^2) zugrunde

Intraklassen-Korrelation	Kleiner Effekt ($\eta'^2 = 0,01$)	Mittlerer Effekt ($\eta'^2 = 0,06$)	Großer Effekt ($\eta'^2 = 0,14$)
$\rho = 0,20$	0,012	0,074	0,169
$\rho = 0,40$	0,017	0,096	0,213
$\rho = 0,60$	0,025	0,138	0,289
$\rho = 0,80$	0,048	0,242	0,449

Das Effektgrößenmaß ϕ'^2 lässt sich dann in die Effektgröße η'^2 umrechnen (s. auch Formel F 13.54):

$$\eta'^2 = \frac{\phi'^2}{1 + \phi'^2} \quad (\text{F 14.48})$$

Diesen Zusammenhang können wir nutzen, um Effektgrößen aus Varianzanalysen mit Messwiederholung anhand der Klassifikation von Cohen (1988) zu bewerten (vgl. Barcikowski & Robey, 1985). Tabelle 14.3 zeigt einen Klassifikationsvorschlag für große, mittlere und kleine Effektgrößen bei vier unterschiedlichen Ausprägungen der Intraklassen-Korrelation. Man sieht deutlich, dass ein Effekt einer bestimmten Größe als umso »größer« bewertet werden kann, je kleiner die Intraklassen-Korrelation (d.h. je geringer der Einfluss personenspezifischer Unterschiede) ist. Da Formel F 14.47 nur gilt, wenn alle Varianzen gleich sind und alle Kovarianzen gleich sind, sollten die Werte in Tabelle 14.3 auch nur verwendet werden, wenn diese Voraussetzung erfüllt bzw. nicht allzu stark verletzt ist.

Das Effektgrößenmaß η'^2 ist zwar nicht mit dem partiellen Effektgrößenmaß η_p^2 identisch, aber beide werden meist zu ähnlichen Ergebnissen kommen (jedenfalls dann, wenn die Annahme homogener Varianzen und homogener Kovarianzen erfüllt ist). Das liegt daran, dass sie im Grunde von der gleichen Ursache abhängen (nämlich dem Ausmaß stabiler Personunterschiede, also σ_π^2).

Konfidenzintervall für Effektgrößen

Um die Grenzen des zweiseitigen $(1-\alpha)$ -Konfidenzintervalls für eine Effektgröße bei der einfaktoriellen Varianzanalyse mit Messwiederholung zu bestimmen, benötigt man die Grenzen des entsprechenden Kon-

fidenzintervalls für den Nonzentralitätsparameter der nonzentralen F -Verteilung. Barcikowski und Robey (1985) haben vorgeschlagen, den Nonzentralitätsparameter anhand des Effektgrößenmaßes ϕ'^2 zu bestimmen:

$$\lambda' = n \cdot J \cdot \phi'^2 \quad (\text{F 14.49})$$

Nachdem Untergrenze und Obergrenze des Konfidenzintervalls für den Nonzentralitätsparameter (λ'_u und λ'_o) bestimmt sind, lassen sie sich wie folgt in die Unter- und Obergrenzen für das Effektgrößenmaß ϕ'^2 zurückrechnen:

$$\phi'^2 = \frac{\lambda'}{n \cdot J} \quad (\text{F 14.50})$$

Die Formeln F 14.49 und F 14.50 lassen sich nur dann zur Schätzung des Nonzentralitätsparameters aus den Daten verwenden, wenn die Sphärizitätsannahme erfüllt ist. Für den Fall, dass die Sphärizitätsannahme verletzt ist, muss der geschätzte Nonzentralitätsparameter korrigiert werden. Muller und Barton (1989) haben vorgeschlagen, für diese Korrektur den Koeffizienten ϵ_{Box} zu verwenden, den wir schon bei der Korrektur der Freiheitsgrade kennengelernt haben. Hierzu kann man entweder die Methode nach Greenhouse und Geisser (1959) oder nach Huynh und Feldt (1976) verwenden. Der geschätzte Nonzentralitätsparameter λ' wird dann einfach mit $\hat{\epsilon}_{\text{GG}}$ oder $\hat{\epsilon}_{\text{HF}}$ multipliziert.

Beispiel

Effektgröße und Konfidenzintervall im Beispiel zum Modelllernen

Wie groß ist der Effekt der experimentellen Manipulation (»Faktor A«) in unserem Datenbeispiel aus Tabelle 14.1, und wo liegen die Grenzen des zweiseitigen 90 %-Konfidenzintervall für den Populationseffekt? Um das Effektgrößenmaß ϕ'^2 schätzen zu können, benötigen wir zunächst die Intraklassen-Korrelation. Diese schätzen wir anhand von Formel F 14.46:

$$\hat{\rho} = \frac{282 - 60,5}{282 - 60,5 \cdot (1 - 3)} = 0,55$$

Damit kann das Effektgrößenmaß ϕ'^2 in Anlehnung an Formel F 14.47 wie folgt aus den Daten geschätzt werden

$$\phi'^2 = \frac{0,644}{(1 - 0,55) \cdot (1 - 0,644)} = 4,02$$

Mithilfe des Programms NDC (⬇️) bestimmen wir folgende Unter- bzw. Obergrenze des zweiseitigen 90 %-Konfidenzintervalls für den Nonzentralitätsparameter: $\lambda'_u = 11,898$ und $\lambda'_o = 100,688$. Diese beiden Werte können nach Formel F 14.50 in die Metrik des Effektgrößenmaßes ϕ'^2 zurücktransformiert werden. So ergibt sich als Untergrenze des Intervalls $\phi'^2_u = 0,793$ und als Obergrenze $\phi'^2_o = 6,713$. Der wahre Populationseffekt ϕ'^2 wird also mit einer Wahrscheinlichkeit von 90 % von einem Intervall mit den Grenzen [0,793; 6,713] überdeckt.

14.1.9 A-priori-Poweranalyse: Planung des optimalen Stichprobenumfangs

Mithilfe des Nonzentralitätsparameters sind wir in der Lage, den Stichprobenumfang einer einfaktoriellen Varianzanalyse mit Messwiederholung so zu planen, dass die Wahrscheinlichkeit, mit der unter der Annahme eines spezifischen Populationseffekts unter der Alternativhypothese (ϕ'^2_1) auf einem festgelegten Signifikanzniveau α die Nullhypothese abgelehnt werden kann, $1 - \beta$ beträgt. Hierzu legen wir einen spezifischen Populationseffekt – z. B. unter Zuhilfenahme von Tabelle 14.3 – fest. Dann lässt sich die optimale Stichprobengröße für den F -Test einfach über den Nonzentralitätsparameter λ'_1 bestimmen. Löst man Formel F 14.49 nach n auf, erhält man:

$$n = \frac{\lambda'_1}{J \cdot \phi'^2_1} \quad (\text{F 14.51})$$

Der Nonzentralitätsparameter λ'_1 ergibt sich aus der Teststärke, also dem Flächenanteil unter der nonzentralen F -Verteilung, den ein kritischer Wert F_{krit} nach rechts abschneidet. Der kritische Wert wiederum ist derjenige F -Wert, der unter der zentralen F -Verteilung einen Flächenanteil von α nach rechts abschneidet. Sind die Irrtumswahrscheinlichkeiten α und β festgelegt, steht also auch der Nonzentralitätsparameter λ'_1 fest. Komplizierter ist hingegen die Festlegung des hypothetischen Populationseffekts ϕ'^2_1 . Dieser Wert kann nur bestimmt werden, wenn man die Intraklassen-Korrelation in der Population kennt oder (z. B. auf der Basis von früheren Untersuchungen) schätzen kann (s. folgenden Kasten).

Möchte man die Teststärke post hoc für die empirischen Daten (d.h. unter Verwendung des empirisch ermittelten Effektgrößenschätzers) bestimmen und ist

die Sphärizitätsannahme verletzt, so muss der Nonzentralitätsparameter mit $\hat{\epsilon}_{GG}$ oder $\hat{\epsilon}_{HF}$ multipliziert werden.

Beispiel

Planung eines Messwiederholungs-Experiments

Eine Entwicklungspsychologin plant eine Längsschnittstudie mit $J = 4$ Messzeitpunkten. Sie möchte herausfinden, ob und wie sich die motorische Koordination bei dreijährigen Kindern im Laufe eines Jahres (im Abstand von jeweils drei Monaten) verändert. Der Test, mit dem sie motorische Fähigkeiten messen will, liefert intervallskalierte Messungen; die Testwerte sind in der Population normalverteilt. Die Forscherin weiß, dass die Stabilität motorischer Fähigkeiten innerhalb eines Jahres bei Kindern recht hoch ist; sie geht von einer Intraklassen-Korrelation von $\rho = 0,60$ aus. Ferner geht sie von einem kleinen Effekt in der Population aus. Das würde – im Falle eines Designs mit unkorrelierten Messungen – einer postulierten Populationseffektstärke von $\eta_1^2 = 0,01$ bzw. $\phi_1^2 = 0,01/(1 - 0,01) = 0,01$ entsprechen. Umgerechnet in das korrigierte Effektgrößenmaß $\phi_1'^2$ würde sich nach Formel F 14.47 dann ein Wert von $\phi_1'^2 = 0,01/((1 - 0,6) \cdot (1 - 0,01)) = 0,025$ bzw. $\eta_1'^2 = 0,025/(1 + 0,025) = 0,024$ ergeben. Wie viele Kinder muss ihre Stichprobe umfassen, damit die Nullhypothese auf einem Signifikanzniveau von $\alpha = 5\%$ mit einer Wahrscheinlichkeit von $1 - \beta = 90\%$ abgelehnt werden kann, falls es einen Effekt der postulierten Größe tatsächlich gibt?

Zur Berechnung des optimalen Stichprobenumfangs nehmen wir das Programm G*Power ( zu Hilfe.

G*Power verwendet als Effektgrößenmaß ϕ (bzw. ϕ_1 , also die positive Quadratwurzel aus ϕ_1^2). Wir geben in dem Programm das unkorrigierte Effektgrößenmaß ϕ_1 an, da die Intraklassen-Korrelation gesondert eingegeben werden muss und das Programm die Umrechnung in $\phi_1'^2$ selbst vornimmt. Die in G*Power einzugebenden Werte lauten also $\phi_1 = 0,1$; $\alpha = 0,05$; $1 - \beta = 0,90$; »number of groups« = 1 (da die Kinder hier nicht mehr in unterschiedliche Gruppen eingeteilt werden); »repetitions« (das entspricht der Anzahl der Stufen des messwiederholten Faktors, also J) = 4; »corr among repeated measures« (das entspricht der Intraklassen-Korrelation ρ) = 0,60 und »nonsphericity correction« (das entspricht dem Korrekturfaktor ϵ) = 1 (da wir hier der Einfachheit halber von einer perfekt sphärischen Matrix ausgehen). G*Power ermittelt einen Nonzentralitätsparameter von $\lambda_1' = 14,4$. Setzen wir diesen Wert in Formel F 14.51 ein, erhalten wir $n = 14,4/(4 \cdot 0,025) = 144$. Man würde also insgesamt 144 Personen benötigen, damit die Nullhypothese bei einem a priori festgelegten Populationseffekt von $\phi_1'^2 = 0,025$ auf einem α -Niveau von 5 % mit einer Wahrscheinlichkeit von $1 - \beta = 90\%$ abgelehnt werden kann (s. Abb. 14.3).

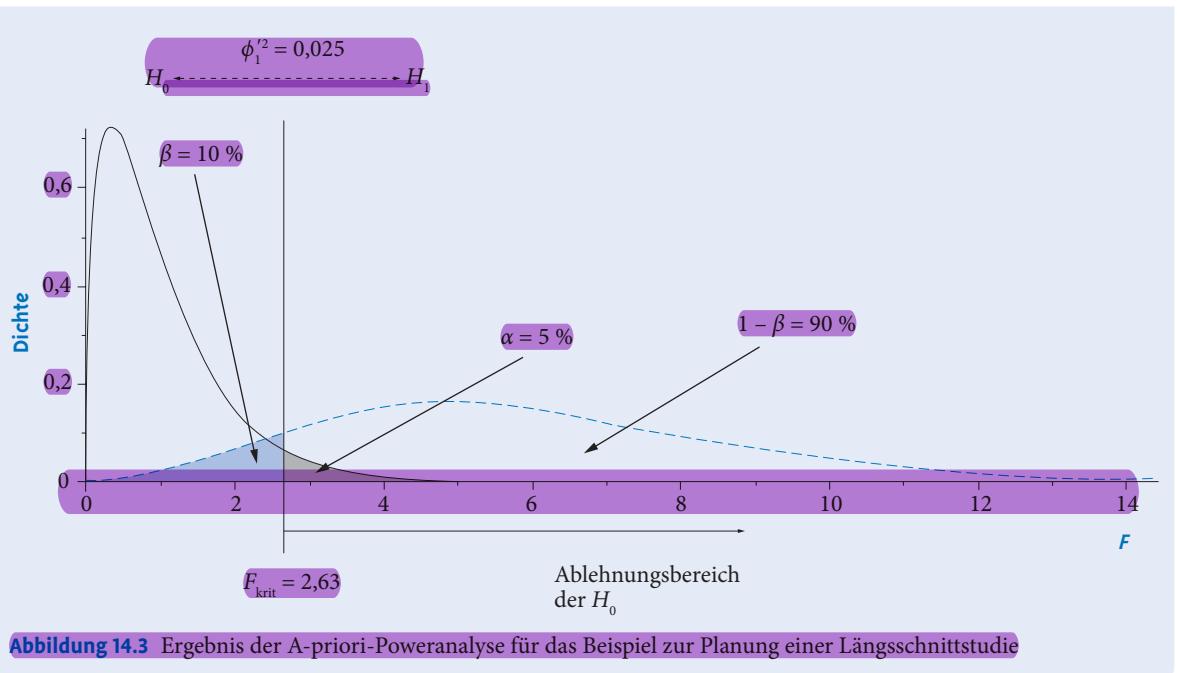


Abbildung 14.3 Ergebnis der A-priori-Poweranalyse für das Beispiel zur Planung einer Längsschnittstudie

14.1.10 Kontrastanalyse

Genau wie bei der einfaktoriellen Varianzanalyse ohne Messwiederholung lassen sich auch bei der einfaktoriellen Varianzanalyse mit Messwiederholung spezifische Kontrasthypthesen formulieren und testen. Solche Kontraste werden bei Messwiederholungsdesigns v.a. dann verwendet, wenn es sich bei der unabhängigen Variablen um den Faktor Zeit handelt und man einen bestimmten Verlauf der Mittelwerte über die J Messzeitpunkte annimmt. Der angenommene Verlauf der Mittelwerte impliziert eine Hypothese über die Form der Veränderung der Merkmalsausprägungen über die Zeit hinweg. Ein Beispiel: Man will die Annahme testen, dass die Lernmotivation von Schülerinnen und Schülern im Verlauf eines Schuljahres stetig absinkt. Diese Veränderungshypothese impliziert einen Abwärtstrend der durchschnittlichen Lernmotivationswerte über die Messzeitpunkte hinweg. Dieser Trend kann linear sein (d.h., mit jeder Zeiteinheit nimmt die Lernmotivation um einen konstanten Betrag ab), er kann aber auch polynomisch höherer Ordnung (also quadratisch, kubisch etc.) sein. Solche polynomischen Trends einfacher oder höherer Ordnung lassen sich relativ leicht modellieren, wie wir im Folgenden sehen werden. Nichtlineare Veränderungen, die sich nicht über einen polynomischen

Trend beschreiben lassen (etwa Veränderungen mit einem »Knick«, d.h. einem Punkt auf der Zeitachse, an dem sich die Stärke oder die Form eines Trends verändert), sind leichter mithilfe von multiplen Regressionsanalysen zu modellieren (Singer & Willett, 2003; s. auch Abschn. 20.4).

Linearer Trend

Unter einem Kontrast Λ versteht man eine Linearkombination (d.h. eine mit Kontrastkoeffizienten K gewichtete Summe) der J Mittelwerte des messwiederholten Faktors (vgl. Formel F 13.70):

$$\Lambda = K_1 \cdot \mu_{\cdot 1} + K_2 \cdot \mu_{\cdot 2} + K_3 \cdot \mu_{\cdot 3} + \dots + K_J \cdot \mu_{\cdot J},$$

wobei die Bedingung erfüllt sein muss, dass die Summe der Kontrastkoeffizienten über die Bedingungen hinweg 0 ergibt (vgl. Formel F 13.71):

$$\sum_{j=1}^J K_j = 0$$

Ein linearer Trend über die Zeit hinweg lässt sich einfach modellieren, indem man den Messzeitpunkten (d.h. den Faktorstufen) die Kontrastkoeffizienten so zuordnet, dass (1) ihre numerischen Unterschiede den tatsächlichen zeitlichen Abständen entsprechen (Äquidistanz) und dass (2) ihre Summe 0 ergibt. Sind

die Messungen ohnehin gleichabständig (d.h., liegt ein konstantes Intervall zwischen den Messzeitpunkten), so könnten die Kontrastkoeffizienten wie folgt festgelegt werden:

- ▶ bei $J = 3$ Messzeitpunkten:

$$K_1 = -1, K_2 = 0, K_3 = +1$$

- ▶ bei $J = 4$ Messzeitpunkten:

$$K_1 = -3, K_2 = -1, K_3 = +1, K_4 = +3$$

- ▶ bei $J = 5$ Messzeitpunkten:

$$K_1 = -2, K_2 = -1, K_3 = 0, K_4 = +1, K_5 = +2 \text{ usw.}$$

Das Ergebnis einer Kontrastanalyse ist von der Metrik der Kontrastkoeffizienten völlig unabhängig; lediglich die Summennormierung in Formel F 13.71 muss erfüllt sein. Die mit 0 codierte Messgelegenheit ist also immer die »mittlere«, d.h. der Durchschnitt auf der Zeitachse.

Die inferenzstatistische Absicherung eines Kontrasts erfolgt analog zu dem in Abschnitt 13.1.13 beschriebenen Vorgehen. Er wird über den Ausdruck

$$L = \sum_{j=1}^J K_j \cdot \bar{x}_{\cdot j}$$

geschätzt. Man erhält gemäß Formel F 13.73b und F 13.74 eine F -verteilte Prüfgröße, deren Werte wie folgt bestimmt werden:

$$F = \frac{n \cdot \left(\sum_{j=1}^J K_j \cdot \bar{x}_{\cdot j} \right)^2}{\sum_{j=1}^J K_j^2 \cdot \text{MQS}_{\text{Res}}} \quad (\text{F 14.52})$$

Der kritische Wert unter der F -Verteilung hat $df_1 = df_{\text{Kontrast}} = 1$ Zählerfreiheitsgrad und $df_2 = df_{\text{Res}} = (J - 1) \cdot (n - 1)$ Nennerfreiheitsgrade. Ist der Wert der Prüfgröße in F 14.52 mindestens so groß wie der kritische F -Wert auf einem vorher festgelegten Signifikanzniveau α , dann ist der Kontrast signifikant von 0 verschieden; die Nullhypothese, der zufolge es keinen Trend in der durchschnittlichen Merkmalsausprägung über die Messzeitpunkte hinweg gibt, kann abgelehnt werden. Für den Fall, dass man eine gerichtete Kontrasthypothese testen will, kann man den F -Test auf der Basis eines doppelten α -Niveaus durchführen (bzw. den p -Wert halbieren; vgl. Abschn. 13.1.13).

Polynomiale Trends höherer Ordnung

In Abschnitt 13.1.13 hatten wir polynomiale Trends höherer Ordnung besprochen und an einem Beispiel veranschaulicht. Zwei Einsichten sind uns aus Abschnitt 13.1.13 bereits bekannt und gelten auch bei der Varianzanalyse mit Messwiederholung: (1) Polynomiale Trends bilden ein Set orthogonaler Kontraste, d.h., die einzelnen Trendtests sind miteinander unkorreliert. (2) Die Anzahl der Kontraste in einem Set kann maximal $J - 1$ betragen. In Abbildung 14.4 sind die drei möglichen Trends (linear, quadratisch, kubisch) grafisch abgetragen.

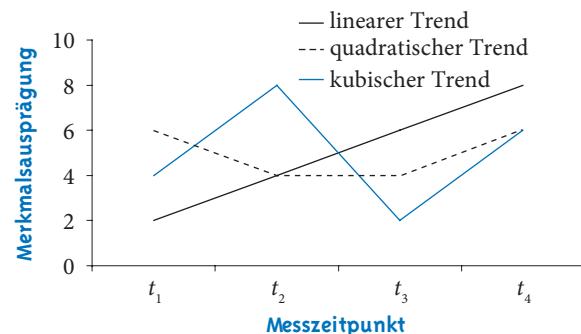


Abbildung 14.4 Set polynomialer Trends bei $J = 4$ Messzeitpunkten

Quadratischer Trend. Ein polomialer Trend zweiter Ordnung ist ein quadratischer Trend. Wenn es sich bei der unabhängigen Variablen um den Faktor »Zeit« handelt, dann würde ein quadratischer Trend besagen, dass die Merkmalsausprägung zunächst negativ beschleunigt abnimmt, um ab einem »mittleren« Messzeitpunkt wieder positiv beschleunigt anzusteigen (Abb. 14.4, gestrichelte graue Linie). Eine solche Trendhypothese lässt sich wie folgt in Kontrastkoeffizienten übersetzen:

- ▶ bei $J = 3$ Messzeitpunkten:

$$K_1 = +1, K_2 = -2, K_3 = +1$$

- ▶ bei $J = 4$ Messzeitpunkten:

$$K_1 = +1, K_2 = -1, K_3 = -1, K_4 = +1$$

- ▶ bei $J = 5$ Messzeitpunkten:

$$K_1 = +2, K_2 = -1, K_3 = -2, K_4 = -1, K_5 = +2 \text{ usw.}$$

Kubischer Trend. Ein polomialer Trend dritter Ordnung ist ein kubischer Trend. Wenn es sich bei der unabhängigen Variablen um den Faktor »Zeit« handelt, dann würde ein kubischer Trend besagen, dass die Merkmalsausprägung zunächst zunimmt, dann wieder abnimmt und

schließlich wieder zunimmt (s. Abb. 14.4, blaue Linie). Eine solche Trendhypothese lässt sich wie folgt in Kontrastkoeffizienten übersetzen (man beachte, dass ein kubischer Trend bei $J = 3$ Messzeitpunkten nicht testbar ist):

- ▶ bei $J = 4$ Messzeitpunkten:
 $K_1 = -1, K_2 = +3, K_3 = -3, K_4 = +1$
- ▶ bei $J = 5$ Messzeitpunkten:
 $K_1 = -1, K_2 = +2, K_3 = 0, K_4 = -2, K_5 = +1$ usw.

Beispiel

Wie sinkt die Lernmotivation von Schülerinnen und Schülern im Laufe des Schuljahres?

Eine Schulpsychologin misst die Lernmotivation von Schülerinnen und Schülern in einer Klasse im Laufe eines Schuljahres zu $J = 4$ Messzeitpunkten: das erste Mal zu Beginn des Schuljahres und anschließend noch drei weitere Male, jeweils im Abstand von genau drei Monaten. In Tabelle 14.4 sind die Daten von $n = 8$ Schülerinnen und Schülern abgetragen. Folgt der durchschnittliche Verlauf der Lernmotivation über die vier Messzeitpunkte hinweg einem linearen, einem quadratischen oder einem kubischen Trend? Um bei mehreren Kontrasten die einzelnen Kontraste L unterscheiden zu können, versehen wir sowohl die Kontraste als auch die dazugehörigen Kontrastkoeffizienten K mit einem zweiten Index i : K_{ij} ist somit der Kontrastkoeffizient der Bedingung a_j des i -ten Kontrastes. Die Kontrastkoeffizienten für die drei Kontraste lauten wie folgt:

- ▶ linearer Trend:
 $K_{11} = -3, K_{12} = -1, K_{13} = +1, K_{14} = +3$
- ▶ quadratischer Trend:
 $K_{21} = +1, K_{22} = -1, K_{23} = -1, K_{24} = +1$
- ▶ kubischer Trend:
 $K_{31} = -1, K_{32} = +3, K_{33} = -3, K_{34} = +1$

Für jede der drei Kontrasthypthesen kann nun eine gewichtete Summe der Bedingungsmittelwerte

$$\sum_{j=1}^J K_{ij} \cdot \bar{x}_{\bullet j}$$

berechnet werden. Für den linearen Trend beträgt sie $-10,75$; für den quadratischen Trend -1 und für den kubischen Trend $-2,75$. Rechnet man die gewichteten Summen nach Formel F 13.74 in Quadratsummen um, ergeben sich folgende Werte: $QS_1 = 46,225$; $QS_2 = 2$ und $QS_3 = 3,025$. Auch hier gilt, dass die Summe der drei Quadratsummen der QS_{ZWA} entspricht, d. h. der Quadratsumme für den Faktor »Zeit« in der herkömmlichen Varianzanalyse (s. Formel F 14.4); sie beträgt in diesem Beispiel $QS_{ZWA} = 51,25$. Die MQS_{Res} , die wir nach Formel F 14.32 berechnen, beträgt hier $3,37$. Setzen wir diese Werte in Formel F 14.52 ein, erhalten wir für die drei Kontraste die Werte $F_1 = 13,72$; $F_2 = 0,59$ und $F_3 = 0,90$. Da wir hier drei Kontraste testen, müssen wir die Kumulierung des α -Fehlers mit berücksichtigen.

Tabelle 14.4 Rohdaten für das Beispiel zur Veränderung der Lernmotivation

Person m	t ₁ : Beginn des Schuljahres	t ₂ : 3 Monate später	t ₃ : 6 Monate später	t ₄ : 9 Monate später
1	18	15	15	12
2	15	16	14	11
3	17	16	17	10
4	16	14	15	15
5	16	19	17	18
6	19	14	15	10
7	15	12	13	12
8	14	15	13	14
Mittelwert $\bar{x}_{\bullet j}$	16,25	15,125	14,875	12,75

Wir testen anhand der Bonferroni-Korrektur (s. Abschn. 13.1.12) jeden einzelnen Kontrast auf der Basis eines Signifikanzniveaus von $0,05/3 = 0,017$ (s. Formel F 13.67). Der kritische F -Wert für ein Signifikanzniveau von $\alpha = 1,7\%$ bei $df_{\text{Kontrast}} = 1$ Zählerfreiheitsgrad und $df_{\text{Res}} = 21$ Nennerfreiheitsgraden beträgt

$F_{(0,983; 1; 21)} = 6,72$. Der lineare Trend ist signifikant; der quadratische und der kubische Trend sind hingegen nicht signifikant. Die Veränderung der Lernmotivation der acht Schülerinnen und Schüler über die vier Messzeitpunkte hinweg folgt also einem linearen Trend.

Wie man bei der einfaktoriellen Varianzanalyse mit Messwiederholung paarweise Mittelwertsvergleiche post hoc durchführt und dabei die Inflation des α -Fehlers kontrolliert, werden wir hier aus Platzgründen nicht behandeln. Dazu sei auf das Lehrbuch von Hays (1994) verwiesen.

14.2 Zweifaktorielle Varianzanalyse mit Messwiederholung

Bei der zweifaktoriellen Varianzanalyse mit Messwiederholung unterscheiden wir zwei Fälle: Im einen Fall sind beide Faktoren (A und B) messwiederholt bzw. intraindividuell variiert (Abschn. 14.2.1), im zweiten Fall ist nur einer der Faktoren messwiederholt (Abschn. 14.2.2). Zentral für das Verständnis der zweifaktoriellen Varianzanalyse mit Messwiederholung ist das Konzept der Wechselwirkung zwischen den beiden Faktoren. Genau wie bei der zweifaktoriellen Varianzanalyse ohne Messwiederholung (Abschn. 13.2) können hier drei Typen von Effekten getestet werden – unabhängig davon, ob auf einem oder auf beiden Faktoren eine Messwiederholung vorliegt oder nicht: (1) Haupteffekte des Faktors A , (2) Haupteffekte des Faktors B und (3) Interaktionseffekte $A \times B$. Eine Interaktion zwischen zwei Faktoren liegt dann vor, wenn der Effekt eines Faktors von der Ausprägung auf dem jeweils anderen Faktor abhängt, d.h., wenn sich die einfachen Haupteffekte eines Faktors (d.h. Ausmaß und/oder Richtung der Unterschiede zwischen den Mittelwerten auf den Stufen dieses Faktors) zwischen den Stufen des jeweils anderen Faktors unterscheiden.

14.2.1 Zweifaktorielle Varianzanalyse mit Messwiederholung auf beiden Faktoren

Um die zweifaktorielle Varianzanalyse mit Messwiederholung auf beiden Faktoren zu veranschaulichen, greifen wir auf das Beispiel zum Modellernen in Ab-

schnitt 13.2 und auf die Rohdaten aus Tabelle 13.8 zurück. Variiert wurde zum einen, ob es sich bei der beobachteten Person um ein Kind (d.h. große Ähnlichkeit zur Versuchsperson) oder um einen Erwachsenen handelte (d.h. geringe Ähnlichkeit zur Versuchsperson; Faktor A). Zum anderen wurde variiert, ob die beobachtete Person für ihr Verhalten belohnt wurde oder nicht (Faktor B). Gemessen wurde die Nachahmungstendenz der Kinder, die an dem Experiment teilnahmen. Gehen wir nun davon aus, dass beide Faktoren intraindividuell variiert wurden. Das bedeutet: Jedes Kind hat vier Filme gesehen, die sich wie folgt unterschieden haben:

- (1) Die beobachtete Person war der Versuchsperson ähnlich und wurde für ihr Verhalten belohnt.
- (2) Die beobachtete Person war der Versuchsperson ähnlich und wurde für ihr Verhalten nicht belohnt.
- (3) Die beobachtete Person war der Versuchsperson unähnlich und wurde für ihr Verhalten belohnt.
- (4) Die beobachtete Person war der Versuchsperson unähnlich und wurde für ihr Verhalten nicht belohnt.

Nach jedem Film wurde die Nachahmungstendenz gemessen, d.h., jede Versuchsperson hat vier Werte abgegeben.

Man ahnt sofort die Vor- und Nachteile eines solchen Designs gegenüber einem komplett nicht-messwiederholten Design. Der Vorteil besteht darin, dass man Versuchspersonen spart: Wenn man beide Faktoren intraindividuell variiert und alle Versuchspersonen alle Bedingungskombinationen durchlaufen, dann entspricht die Anzahl der Messwerte pro Zelle der Anzahl aller benötigten Versuchspersonen. In unserem Beispiel nehmen an dem Versuch nur 5 Kinder teil, aber am Ende hat man trotzdem 20 Messwerte. Ein Nachteil besteht darin, dass es aufgrund der Messwiederholung zu Sequenzeffekten kommen kann: Falls es Unterschiede zwischen den Bedingungskombinationen gibt, ist nicht