.

Einführung in die Varianzanalyse

Uwe Mortensen
Westfälische Wilhelms-Universität
Fachbereich Psychologie und Sportwissenschaft

Letzte Bearbeitung 27. 12. 2015

Inhaltsverzeichnis

1 Einführung in die Varianzanalyse

Es wird eine elementare Einführung in die Varianzanalyse gegeben, d.h. es wird nicht nicht von der Vektor- und Matrixalgebra Gebrauch gemacht wird und es werden nur die Grundmodelle vorgestellt.

1.1 Allgemeine Überlegungen

Will man den Effekt einer bestimmten Bedingung auf eine Größe untersuchen, so erhebt man üblicherweise zwei Stichproben von Messungen: die Messungen der einen Stichprobe sind unter den zur Diskussion stehenden Bedingungen erhoben worden, während die der anderen unter "Normalbedingungen" bestimmt wurden. Sie sind die Kontrollmessungen, mit denen die unter Experimentalbedingungen erhobenen Messungen verglichen werden sollen. Statt eine Experimental- und eine Kontrollbedingung miteinander zu vergleichen kann man natürlich auch Personen aus verschiedenen Gruppen, etwa weibliche und männliche Personen, unter der gleichen Bedingung hinsichtlich einer interessierenden Variablen miteinader vergleichen. Der Vergleich der Messungen hängt von dem Modell ab, das für die Meßwerte aufgestellt wird. Ein beliebtes Modell, das Intervallskalenniveau der Messungen voraussetzt, besteht in dem Ansatz

$$\begin{aligned} x_{1i} &= \mu_1 + e_{1i}, \quad i = 1, \cdots, n_1 \\ x_{2j} &= \mu_2 + e_{2j}, \quad j = 1, \cdots, n_2 \end{aligned} \tag{1.1}$$

Dabei sind n_1 und n_2 die Stichprobenumfänge. μ_1 und μ_2 sind Konstante, die die Kontroll– bzw. die Experimentalbedingung charakterisieren (oder die die zu vergleichenden Gruppen charakterisieren), und die e_{1i}, e_{2j} sind die unvermeidlichen Meßfehler. Bekanntlich sind die Mittelwerte $\bar{x}_1$ und $\bar{x}_2$ Schätzungen für μ_1 und μ_2, und die Hypothese $H_0 : \mu_1 = \mu_2$ kann mit einem t-Test überprüft werden, falls die Annahme gleicher Varianzen für die beiden Stichproben gerechtfertigt ist.

Nun ist diese Art von Untersuchung ein einfacher Spezialfall eines allgemeineren Schemas. Denn oft will man nicht eine experimentelle Bedingung, sondern mehrere miteinander vergleichen. Solche Bedingungen können z.B. verschiedene Therapieformen sein, und man möchte differenziertere Aussagen über die Wirkung der Therapien machen in Abhängigkeit etwa vom (i) Geschlecht, (ii) Alter, (iii) Bildungsstand der Personen, die sich einer Therapie unterziehen. Es ergeben sich sofort viele "Experimentalgruppen", deren Mittelwerte miteinander verglichen werden müssen (hier wird stillschweigend und stark vereinfachend vorausgesetzt, daß man die Wirkung der Therapie durch eine Variable ausdrücken kann; es geht aber nur um die Etnwicklung der Fragestellung, nicht um ihre korrekte Beantwortung.) Man könnte nun alle Paare von Mittelwerten betrachten und sie mit dem t-Test auf Signifikanz prüfen. Angenommen, es gäbe gerade k Experimentalgruppen. Dann gibt es k Mittelwerte und dementsprechend $N_k = \binom{k}{2} = k(k-1)/2$

Paare. Die Anzahl N_k wächst einigermaßen schnell mit k, und man steht vor dem Problem, die Aussagen der vielen t-Tests zu beurteilen. So kann für ein bestimmtes Paar von Mittelwerten die Nullhypothese gelten, – aber mit der Wahrscheinlichkeit α entscheidet man sich für H_1. Machen wir für den Augenblick die vereinfachende Annahme, daß alle t-Tests stochastisch unabhängig sind, so kann man nach der Wahrscheinlichkeit mindestens eines Fehlers der ersten Art fragen. Bei N_k Tests beträgt die Wahrscheinlichkeit, keinen α-Fehler zu begehen, gerade $(1-\alpha)^{N_k}$, so daß die Wahrscheinlichkeit mindestens eines Fehlers durch

$$P(\text{mindestens eine Fehlentscheidung}) = 1 - (1-\alpha)^{N_k}$$

gegeben ist. Hat man z.B. $k = 5$ Gruppen, so kann man 10 paarweise t-Tests durchführen und die Wahrscheinlichkeit *mindestens* einer Fehlentscheidung beträgt bereits .40. Da man natürlich nicht weiß, bei welchem t-Test man sich falsch entschieden hat, resultiert ein Interpretationsproblem, denn es ist unklar, welche Unterschiede sich interpretieren lassen und welche nicht.

Wenn sich die inferenzstatistische Aussage über den Effekt eines der untersuchten Faktoren nur auf die tatsächlich untersuchten Stufen dieses Faktors bezieht, so handelt es sich um einen "festen Faktor" (fixed factor). Natürlich können alle Faktoren in einem Plan feste Faktoren sein. Gelegentlich liegt es aber nahe, die Stufen eines Faktors oder mehrer Faktoren als zufällige Auswahl aus der Menge der möglichen Stufen zu betrachten. Das inferenzstatistische Ziel ist gleichwohl eine Aussage über die Wirkung aller Stufen. Solche Faktoren heißen "zufällige Faktoren"(random factors). Ein Versuchsplan kann nur feste oder nur zufällige Faktoren enthalten; darüber hinaus kann man den gemischten Fall haben.

1.2 Allgemeine Voraussetzungen

1.2.1 Ein Hilfssatz

Es soll ein Test der Hypothese H_0 hergeleitet werden. Es seien s_j^2 die Varianzen der Meßwerte in den einzelnen Gruppen. Jede dieser Stichprobenvarianzen ist eine Schätzung für σ^2. Weiter kann man die Varianz $s_{\bar{x}}$ der Mittelwerte $\bar{x}_j$ berechnen. Angenommen, die Hypothese H_0 ist korrekt. Unter dieser Bedingung liefert $s_{\bar{x}}$ wegen der allgemeinen Beziehung $Var(\bar{x}) = \sigma^2/n$ ebenfalls eine Schätzung für σ^2; sind die n_j gleich groß, so ist $n_j s_{\bar{x}}$ diese Schätzung für σ^2. Man könnte also die s_j^2 mitteln, hätte damit eine zusammengefaßte (pooled) Schätzung für σ^2, und sie mit der Schätzung $n_j s_{\bar{x}}$ vergleichen. Der Vergleich kann dann ein F-Test sein, falls diese beiden Schätzungen stochastisch unabhängig voneinander sind. Tatsächlich läßt sich leicht zeigen, daß sie unabhängig sind.

Bevor die einzelnen Modelle betrachtet werden, wird ein allgemeiner Hilfssatz präsentiert; die Herleitung der Tests läßt sich im allgemeinen auf eine Anwendung dieses Satzes zurückführen. Die Idee, Die Quadratsummen anhand dieses Hilfsatzes herzuleiten, geht auf Mann (1949) zurück.

Hilfssatz: *Es seien $a_1, \cdots, a_n$ irgendwelche n Zahlen, und $\bar{a} = \sum_{i=1}^{n} a_i/n$ sei ihr arithmetisches Mittel. Weiter sei b irgendeine andere Zahl. Dann gilt*

$$\sum_{i=1}^{n}(a_i - \bar{a})^2 = \sum_{i=1}^{n}(a_i - b)^2 - n(\bar{a} - b)^2 \tag{1.2}$$

Beweis: Es ist

$$\begin{aligned}\sum_i (a_i - \bar{a})^2 &= \sum_i (a_i - b + b - \bar{a})^2 = \sum_i ((a_i - b) - (\bar{a} - b))^2 = \\ &= \sum_i (a_i - b)^2 + n(\bar{a} - b)^2 - 2(\bar{a} - b)\sum_i (a_i - b)\end{aligned}$$

Da $\sum_i (a_i - b) = n(\bar{a} - b)$, folgt (1.2). □

1.2.2 Verteilung von Quadratsummen normalverteilter Meßwerte

In den folgenden Abschnitten wird die Verteilung von Quadratsummen der Form $\sum_{i,j}(x_{ij} - \bar{x})^2$, $\sum_j n_j(\bar{x}_j - \bar{x})^2$ etc. benötigt, wobei die $x_{ij} \sim N(\mu, \sigma)$-verteilt[1] sind. Diese Quadratsummen χ^2-verteilt. In Abschnitt 9.3.2 insbesondere wird die Verteilung der Stichprobenvarianz $s^2 = \sum_i (x_i - \bar{x})^2/n$ bzw. $\hat{s}^2 = \sum_i (x_i - \bar{x})^2/(n-1)$ hergeleitet. Es zeigt sich (vergl. Satz 9.5), daß insbesondere

$$s^2 = \frac{\sigma^2}{n}\chi^2_{n-1}, \quad \text{bzw. } \hat{s}^2 = \frac{\sigma^2}{n-1}\chi^2_{n-1} \tag{1.3}$$

gilt, d.h. die Quadratsumme $\sum_{i=1}^{n}(x_i - \bar{x})^2$ ist χ^2_{n-1}-verteilt, d.h. χ^2-verteilt mit $(n-1)$ Freiheitsgraden. Nun ist

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu + \mu - \bar{x})^2 = \sum_{i=1}^{n}((x_i - \mu) - (\bar{x} - \mu))^2.$$

Es ergibt sich

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2.$$

Dividiert man durch σ^2, so erhält man

$$\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2 - \frac{(\bar{x} - \mu)^2}{\sigma^2/n}. \tag{1.4}$$

[1] $N(\mu, \sigma)$ steht für normalverteilt mit Erwartungswert (= "wahrer" Mittelwert, im Unterschied zum geschätzten Mittelwert $\bar{x}$) μ und der Streuung σ bzw. Varianz σ^2.

Die Summe auf der linken Seite ist χ^2_{n-1}-verteilt (vergl. Satz 9.5). $(\bar{x}-\mu)^2/(\sigma^2/n)$ ist sicherlich χ^2_1-verteilt, und da die erste Summe auf der rechten Seite ist χ^2_n-verteilt ist folgt, daß die erste Summe auf der rechten Seite und $(\bar{x}-\mu)^2/\sigma^2/n$ stochastisch unabhängig sind. In der Tat wurde in Abschnitt 9.3.2 gezeigt, daß bei normalverteilten Daten $\bar{x}$ und s^2, bzw. $\hat{s}^2$ stochastisch unabhängig voneinander verteilt sind. Dieser Sachverhalt wird in den Beweisen in den folgenden Abschnitten stillschweigend vorausgesetzt. Es werden also nur Zerlegungen der Form (1.4) hergeleitet und von diesen Zerlegungen wird auf die χ^2-Verteilungen mit den entsprechenden Freiheitsgranden geschlossen.

2 Der einfaktorielle Fall mit festen Effekten

Eine Variable X wird unter $k > 1$ verschiedenen Bedingungen gemessen. Der Mittelwert für die j-te Bedingung sei $\overline{x}_j$, $j = 1, ..., k$. Die Frage ist, ob sich die Mittelwerte statistisch bedeutsam unterscheiden. Man geht dabei aus von dem

2.1 Das Modell

Es wird angenommen, dass sich die x_{ij} aus Größen μ_j und e_{ij} zusammensetzen.

$$\begin{aligned} x_{ij} &= \mu_j + e_{ij}, \quad i = 1, \cdots, n_i; \quad j = 1, \cdots, k \\ e_{ij} &\sim N(0, \sigma^2), \quad \mathrm{Kov}(e_{ij}, e_{i'j'}) = 0 \end{aligned} \tag{2.1}$$

Weiter soll

$$E(x_{ij}) = \mu_j \tag{2.2}$$

gelten. $E(x_{ij})$ ist der Erwartungswert der zufälligen Veränderlichen x_{ij}, d.h. der Mittelwert über alle möglichen Werte, die x_{ij} annehmen kann. Da generell für Erwartungswerte gilt, dass der Erwartungswert einer Summe gleich der Summe der Erwartungswerte ist (wie man sich leicht am Beispiel eines Stichprobenmittelwerts verdeutlicht), folgt dann

$$E(x_{ij} = E(\mu_j + e_{ij}) = E(\mu_j) + E(e_{ij}) = \mu_j + E(e_{ij}) = \mu_j,$$

d.h.

$$E(e_{ij}) = 0. \tag{2.3}$$

Denn $E(\mu_j) = \mu_j$, da μ_j ja eine Konstante und keine zufällige Veränderliche ist. Man sagt, dass μ_j die Wirkung der j-ten Bedingung repräsentiert. Nach (2.1) wird postuliert, dass die Fehler normalverteilt sind mit dem Erwartungswert 0 (vergl. (2.3)) und alle dieselbe Varianz σ^2 haben; darüber hinaus werden sie als unkorreliert und damit als stochastisch unabhängig vorausgesetzt[2]

[2]Unkorreliertheit impliziert nicht bei allen Wahrscheinlichkeitsverteilungen stochastische Unabhängigkeit; diese Implikation ist ein Charakteristikum der Normalverteilung.

Es sei

$$\mu = \frac{1}{k} \sum_{j=1}^{k} \mu_j \tag{2.4}$$

Die Differenz

$$\alpha_j = \mu_j - \mu \tag{2.5}$$

heißt "Effekt" der j-ten Bedingung. Offenbar gilt wegen (2.4

$$\sum_{j=1}^{k} \alpha_j = \sum_{j=1}^{k} \mu_j - \sum_{j=1}^{k} \mu = k\mu - k\mu,$$

d.h. es ist

$$\sum_{j=1}^{k} \alpha_j = 0. \tag{2.6}$$

Weiter ist $\mu_j = \mu + \alpha_j$, so daß man statt (1.1) auch

$$x_{ij} = \mu + \alpha_j + e_{ij} \tag{2.7}$$

schreiben kann. Dies ist die *reparametrisierte Form* des Modells (1.1). Die Gleichung (2.7) heißt auch *Strukturgleichung* des Modells. μ und $\alpha_1, \ldots, \alpha_k$ sind die Parameter des Modells. Sie heißen auch *freie Parameter*, da sie unbekannt sind und aus den Daten geschätzt werden müssen.

Zur Schätzung der freien Parameter: Es gibt verschiedene Methoden der Parameterschätzung, und die Eigenschaften der geschätzten Parameter, d.h. ihre Verteilung, ihr Erwartungswert, ihre Varianz etc) hängen von der Methode ab (da die Parameter anhand einer Stichprobe geschätzt werden, sind sie selbst zufällige Veränderliche). Wünschenswert ist, dass die Schätzungen keinen systematischen Fehler ("Bias") und eine möglichst kleine Varianz haben. Es läßt sich zeigen, dass die *Methode der Kleinsten Quadrate* (MKQ) Schätzungen mit diesen Eigenschaften liefert. So soll nach (2.1) $x_{ij} = \mu_j + e_{ij}$ gelten, und in dieser Gleichung ist μ_j ein freier Parameter. Würde man μ_j kennen, so könnte man x_{ij} bis auf den Fehler e_{ij} "voraussagen". Man hat insgesamt $\sum_j n_j$ mal k Messwerte x_{ij}, ebensoviele Unbekannte Größen e_{ij} plus k unbekannte Parameter μ_j, so dass man insgesamt mehr Unbekannte als Gleichungen hat. Um die μ_j trotzdem schätzen zu können, wird bei der MKQ die Ansatz gemacht, die μ_j so zu bestimmen, dass die Summe

$$Q(\mu_1, \ldots, \mu_k) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} e_{ij}^2 \stackrel{!}{=} \min \tag{2.8}$$

einen minimalen Wert annimmt. Es soll also die Summe der Quadrate der Fehler e_{ij} minimalisiert werden. Man könnte auf die Idee kommen, einfach die Summe der e_{ij} zu minimalisieren, nur funktioniert dieser Ansatz nicht, denn die Summe der Fehler ist stets Null, da sie ja Abweichungen vom Mittelwert sind. Aber

$e_{ij}^2 \geq 0$, so dass $\sum_{i,j} e_{ij}^2 \geq 0$ sein muß, und die e_{ij}^2 sind monotone Funktionen der e_{ij}: werden die e_{ij}^2 im Mittel so klein wie möglich, werden auch die e_{ij} so klein wie möglich.

Man kann jetzt Q in (2.8) bezüglich der μ_j minimalisieren, oder bezüglich der Effekte α_j. Es ist üblich, die α_j zu schätzen, weil dies Vorteile bei der Interpretation der Daten haben kann, wie noch deutlich werden wird. (2.8) wird dann wegen $e_{ij} = x_{ij} - \mu - \alpha_j$ zu

$$Q(\mu_1, \ldots, \mu_k) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \mu - \alpha_j)^2 \tag{2.9}$$

Man bildet nun die Ableitungen nach μ und nach den α_j und setzt diese Ableitungen gleich Null. Man findet

$$\frac{\partial Q}{\partial \mu} = -2 \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \mu - \alpha_j) \tag{2.10}$$

$$\frac{\partial Q}{\partial \alpha_j} = -2 \sum_{i=1}^{n_j} (x_{ij} - \mu - \alpha_j), \tag{2.11}$$

denn es wird ja die Ableitung nach einem bestimmten α_j betrachtet[3]. Es seien $\hat{\mu}$ und $\hat{\alpha}_j$ diejenigen Werte, für die die Ausdrücke auf den rechten Seiten gleich Null werden. Aus (2.10) folgt zunächst

$$\sum_{i=1}^{n_j} (x_{ij} - \mu - \alpha_j) = \sum_{i=1}^{n_j} (x_{ij} - Nk\hat{\mu}) = 0,$$

$N = \sum_j n_j$, denn $\sum_j \alpha_j = 0$ wegen (2.6). Dann folgt aber

$$\hat{\mu} = \bar{x} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}, \tag{2.12}$$

d.h. die Kleinste-Quadrate-Schätzung (KQ-Schätzung) für μ ist einfach das Gesamtmittel der Messwerte. Aus (2.11) folgt dann

$$0 = \sum_{i=1}^{n_j} (x_{ij} - \bar{x} - \hat{\alpha}_j) = \sum_{i=1}^{n_j} x_{ij} - n_j \bar{x} - \hat{\alpha}_j,$$

woraus

$$\hat{\alpha}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} - \bar{x} = \bar{x}_{.j} - \bar{x}. \tag{2.13}$$

[3]Schreibt man die Summe (2.9) voll aus, so sieht man, dass alle $\alpha_1, \ldots, \alpha_k$ darin auftreten. Die Differentiation nach α_j bedeutet, das auch die Differentialquotienten $d\alpha_{j'}/d\alpha_j$ betrachtet werden müssen, die für $j' \neq j$ alle gleich Null sind.

folgt. Man sieht also, dass die KQ-Schätzungen für die freien Parameter einfach durch die entsprechenden Mittelwerte gegeben sind.

Im Folgenden Wird nur mit den Mittelwerten gearbeitet werden, ohne weiteren Bezug auf diese Werte als KQ-Schätzungen für die freien Parameter des Modells. Alle Betrachtungen übertragen sich auf mehrfaktorielle Versuchsanordnungen.

2.2 Hypothesen

Es soll zwischen den folgenden Hypothesen entschieden werden

H_0 : Die μ_i unterscheiden sich nicht, d.h. $\alpha_i = 0$ für alle i.

H_1: Mindestens ein Mittelwert unterscheidet sich signifikant von den anderen, d.h. es existiert ein μ_j, $\mu_j \neq \mu_i$ für alle $i \neq j$.

Das Ziel ist, H_0 zunächst mit *nur einem* Test zu prüfen, also nicht etwa paarweise Tests der Hypothesen $\mu_i \stackrel{?}{=} \mu_j$ durchzuführen. Wie in Abschnitt 1.2 angedeutet wurde kann man einen solchen Test konstruieren, indem man zwei Schätzungen für die Varianz σ^2 miteinander vergleicht, wobei eine Schätzung auf der Schätzung der Varianz der Mittelwerte beruht und die andere auf einer Schätzung der Varianz der Meßwerte selbst. Diese letztere Schätzung ergibt sich, indem man den Gesamtmittelwert $\overline{x}$ der Daten berechnet und dann die Gesamtsumme der Quadrate der Abweichungen bestimmt, also etwa

$$s^2 = \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2 \tag{2.14}$$

berechnet: Hierbei ist N ist die Gesamtzahl der Meßwerte; warum durch $n-k$ geteilt wird und nicht etwa durch $N-1$ wird später deutlich gemacht. Es zeigt sich, daß die hier auftretende Quadratsumme so zerlegt werden kann, daß man aus den Teilsummen den gewünschten Test konstruieren kann. Dies wird im folgenden Abschnitt gezeigt.

2.3 Zerlegung der Quadratsumme

Die Quadratsumme, über die die Gesamtvarianz definiert ist, kann in der folgenden Weise zerlegt werden.

Satz 2.1 *Es sei n_j der Stichprobenumfang der j-ten Gruppe und es sei weiter $n = n_1 + \cdots + n_k$, und*

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}, \quad \bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}, \quad QS_{ges} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2$$

QS_{ges} kann gemäß

$$QS_{ges} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2 \tag{2.15}$$

zerlegt werden.

Beweis: Man betrachte die j-te Teilsumme

$$\sum_{i=1}^{n_i}(x_{ij} - \bar{x}_j)^2.$$

Anhand des Hilfssatzes, d.h. der Aussage (1.2), erhält man mit $b = \bar{x}_j$ sofort

$$\sum_i (x_{ij} - \bar{x})^2 = \sum_i (x_{ij} - \bar{x}_j)^2 - n_j(\bar{x}_j - \bar{x})^2.$$

Summiert man noch über j, so erhält man (2.15). □

Die erste Summe auf der rechten Seite von (2.15) heißt auch *Quadratsumme innerhalb* (QS_{inn}),

$$QS_{inn} = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2,$$

und die zweite heißt *Quadratsumme zwischen* (QS_{zw})

$$QS_{zw} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2,$$

so daß man symbolisch

$$QS_{ges} = QS_{inn} + QS_{zw} \tag{2.16}$$

schreiben kann. Offenbar liefert QS_{inn} eine Schätzung für σ^2, wenn man QS_{inn} durch eine geeignete Anzahl von "Freiheitsgraden" teilt, und QS_{zw} liefert auf analoge Weise eine Schätzung der Varianz der Mittelwerte.

Die Frage ist nun, wie diese Quadratsummen in einen F-Test eingebracht werden können, um H_0 zu testen. Dazu braucht man die Stichprobenverteilung dieser Quadratsummen unter der Bedingung, dass H_0 gilt. Da die F-Verteilung die Verteilung eines Quotienten von χ^2-verteilten Variablen ist, muß die Beziehung von QS_{inn} und QS_{zw} und damit von QS_{ges} zur χ^2-Verteilung hergestellt werden.

Satz 2.2 *Es gelte H_0. Dann gelten die Beziehungen*

$$\frac{QS_{ges}}{\sigma^2} = \chi^2_{n-1}, \quad \frac{QS_{inn}}{\sigma^2} = \chi^2_{n-k}, \quad \frac{QS_{zw}}{\sigma^2} = \chi^2_{k-1}, \tag{2.17}$$

d.h. QS_{ges}/σ^2 ist χ^2-verteilt mit $n-1$ Freiheitsgraden, QS_{inn}/σ^2 ist χ^2-verteilt mit $n-k$ Freiheitsgraden und QS_{zw}/σ^2 ist χ^2-verteilt mit $k-1$ Freiheitsgraden.

Beweis: Mit $b = \mu$ liefert der Hilfsatz

$$QS_{ges} = \sum_{ij}(x_{ij} - \bar{x})^2 = \sum_{ij}(x_{ij} - \mu)^2 - n(\bar{x} - \mu)^2$$

Dividiert man diese Gleichung für QS_{ges} durch σ^2, so erhält man

$$\frac{QS_{ges}}{\sigma^2} = \sum_{ij}\left(\frac{x_{ij} - \mu}{\sigma}\right)^2 - \left(\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}\right)^2 = \chi_n^2 - \chi_1^2 = \chi_{n-1}^2. \tag{2.18}$$

Auf die j-te Teilsumme QS_{inn} kann der Hilfssatz (1.2) mit $b = \mu$ angewendet werden. Demnach ist

$$\sum_i(x_{ij} - \bar{x}_j)^2 = \sum_i(x_{ij} - \mu)^2 - n_j(\bar{x}_j - \mu)^2.$$

Summiert man über j und dividiert man durch σ^2, so erhält man

$$\frac{QS_{inn}}{\sigma^2} = \sum_{ij}\left(\frac{x_{ij} - \mu}{\sigma}\right)^2 - \sum_j\left(\frac{\bar{x}_j - \mu}{\sigma/\sqrt{n_j}}\right)^2 = \chi_n^2 - \chi_k^2 = \chi_{n-k}^2.$$

Es ist

$$\sum_j n_j(\bar{x}_j - \bar{x})^2 = \sum_{ij}(\bar{x}_j - \bar{x})^2.$$

Auf die j-te Teilsumme wendet man den Hilfssatz (1.2) mit $b = \mu$ an:

$$QS_{zw} = \sum_j\left(\sum_i(\bar{x}_j - \mu)^2 - n_j(\bar{x} - \mu)^2\right) = \sum_{ij}(\bar{x}_j - \mu)^2 - n(\overline{x} - \mu)^2.$$

Division durch σ^2 liefert wieder

$$\frac{QS_{zw}}{\sigma^2} = \sum_j\left(\frac{\overline{x}_j - \mu}{\sigma/\sqrt{n_j}}\right)^2 - \left(\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}\right)^2$$

QS_{zw}/σ^2 ist also χ^2-verteilt mit $k - 1$ Freiheitsgraden. □

Nach Satz 2.1 kann die Gesamtvarianz in einen Anteil "innerhalb" der Gruppen und in einen Anteil "zwischen" den Gruppen zerlegt werden, und aus dem Beweis dieser Aussage folgt, daß die Kovarianz zwischen diesen Varianzkomponenten verschwindet, d.h. die Komponenten sind stochastisch unabhängig. Deshalb wird es möglich, die Komponenten in einen F-Test einzubringen (der ja unabhängige Komponenten voraussetzt!).

Satz 2.3 *Es gelte H_0, d.h. die Faktorstufen haben keinen unterschiedlichen Einfluß auf die abhängige Variable. Dann ist der Quotient*

$$F = \frac{QS_{zw}/(k-1)}{QS_{inn}/(n-k)} \tag{2.19}$$

$F_{k-1,n-k}$-verteilt, d.h. F-verteilt mit $k-1$ und $n-k$ Freiheitsgraden.

Beweis: Die Aussage folgt sofort aus der stochastischen Unabhängigkeit der Quadratsummen QS_{inn} und QS_{zw} sowie aus den Aussagen des Satzes (2.2). □

Ist $F_{k-1,n-k}$ größer als der für die gewählte Irrtumswahrscheinlichkeit α kritische F-Wert, so muß die Nullhypothese H_0 verworfen werden; mindestens einer der Mittelwerte weicht signifikant von den anderen ab. Für den Spezialfall von $k = 2$ Gruppen ergibt sich gerade der 2-seitige Test für den Vergleich zweier Mittelwerte, $t_{n-1}^2 = F_{1,n-1}$.

In Satz 2.3 wurde die Voraussetzung genannt, daß die Faktorstufen keinen unterschiedlichen Einfluß auf die abhängige Variable haben sollen, d.h. es muß H_0 gelten, damit der Quotient (2.19) F-verteilt ist. Gilt H_0 so sind alle $\alpha_i = 0$. Man macht sich leicht klar, wie QS_{zw} verteilt ist, wenn für mindestens ein $\alpha_i \neq 0$ gilt. Denn aus $x_{ij} = \mu + \alpha_i + e_{ij}$ folgt $\bar{x}_j = \mu + \alpha_j + e_{\cdot j}$, wobei $e_{\cdot j}$ der mittlere Fehler in der j-ten Gruppe ist. Um den Effekt von $\alpha_i \neq 0$ zu sehen, kann man den Erwartungswert vom QS_{zw} betrachten. Es ist

$$E(QS_{zw}) = E\left(\sum_{j=1}^{k} n_j(\bar{x}_j - \overline{x})^2\right)$$

Umformen und Einsetzen von (2.7) ergibt

$$\begin{aligned} E(QS_{zw}) &= \sum_{j=1}^{k} n_j E(\alpha_i - \mu + e_{\cdot j} + \mu - e_{\cdot\cdot})^2 \\ &= \sum_{j=1}^{k} n_j E(\alpha_i - (e_{\cdot j} - e_{\cdot\cdot}))^2 \\ &= \sum_{j=1}^{k} n_j \alpha_i^2 + \sum_{j=1}^{k} n_j E(e_{\cdot j} - e_{\cdot\cdot})^2 \end{aligned} \tag{2.20}$$

(die Summe $\sum_j n_j(e_{\cdot j} - e_{\cdot\cdot})^2$ ist gleich Null, wie man leicht nachprüft!). Die zweite Summe auf der rechten Seite ist aber unabhängig von den α_i und liefert einen Ausdruck für σ^2. Die erste Summe auf der rechten Seite zeigt, daß die Effekte α_i in QS_{zw} quadratisch, also stets in positiver Form eingehen. QS_{zw} wird also größer als unter H_0 zu erwarten ist, wenn auch nur ein α_i ungleich Null ist.

Die Ergebnisse einer Varianzanalyse werden oft in einer Tabelle zusammengefaßt. Tabelle 1 zeigt die allgemeine Form einer solchen Tabelle für den 1-faktoriellen Fall.

Tabelle 1: Varianzanalyse, 1-faktoriell

Quelle d. Var.	Quadratsumme	Freiheitsgrade	F-Test
A	QS_{zw}	$k-1$	$\frac{QS_{zw}/(k-1)}{QS_{inn}/(n-k)}$
Fehler	QS_{inn}	$n-k$	
Gesamt	QS_{ges}	$n-1$	

2.4 A posteriori-Hypothesen

Ein signifikanter F-Wert legt nahe, dass mindestens eine Differenz $\mu_j - \mu_k$ signifikant von Null verschieden ist. Man kann annehmen, dass die größte Mittelwertdifferenz die Signifikanz erzeugt hat. Es ist aber auch möglich, dass weitere Mittelwertdifferenzen signifikant von Null verschieden sind.

Man kann die Mittelwerte $\bar{x}_j$ in eine Rangreihe bringen; nach geeigneter Umnummerierung hat man etwa

$$\bar{x}_1, \geq \bar{x}_2 \geq \cdots \geq \bar{x}_p.$$

Es ist möglich, dass alle Differenzen $\bar{x}_j - \bar{x}_{j_1}$ signifikant sind, oder nur ein Teil von ihnen, etwa die ersten k. Vielleichtz ist man auch daran interessiert, Gruppen von Mittelwerten zu testen, z.b.

$$\frac{1}{3}(\bar{x}_1 + \bar{x}_2 + \bar{x}_3) - \frac{1}{2}(\bar{x}_4 + \bar{x}_5),$$

etc.

Tests für spezielle Mittelwerte oder Gruppen von Mittelwerte beziehen sich auf *Kontraste*, deren Definition zuerst gegeben wird:

Definition 2.1 *Ein* Kontrast *zwischen Parametern $\mu_1, \ldots, \mu_p$ ist eine lineare Funktion der Form*

$$\Psi = c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p, \quad \sum_{j=1}^{p} c_j = 0. \tag{2.21}$$

wobei die c_j vorgegeben werden und den speziell gewünschten Kontrast definiere.

Man möchte zum Beispiel wissen, ob der kleinste Unterschied $\bar{x}_p - \bar{x}_{p-1}$ signifikant von Null verschieden ist. Dann ist Π mit $c_1 = \cdots = c_{p-2} = 0$, $c_p = 1$, $c_{p-1} = -1$ durch

$$\Psi = \sum_{j=1}^{p} c_j \mu_j = \mu_p - \mu_{p-1}$$

gegeben.

Ein anderes Beispiel ist der Kontrast $(\bar{x}_1 + \bar{x}_2)/2 - (\bar{x}_3 + \bar{x}_4 + \bar{x}_5)/3$: man erhält

$$\Psi = c_1\mu_1 + c_2\mu_2 + c_3\mu_3 + c_4\mu_4 + c_5\mu_5 + c_6\mu_6 + \cdots c_p\mu_p.$$

Da nur die ersten 5 Mittelwerte interessieren, kann man gleich $c_6 = \cdots = c_p = 0$ setzen. Generell muß dann auch

$$c_1 + \cdots + c_5 = 0$$

gelten. Sollen alle Mittelwerte mund es folgt $2c_1 + 3c_3 = 0$. Setzt man $c_1 = 1$ so folgt $c_3 = -2c_1/3 = -2/3$. In der Tat ist dann

$$1 + 1 - 2/3 - 2/3 - 2/3 = 2 - 2 = 0.$$

Man hat dann den geschätzten Kontrast

$$\hat{\Psi} = \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \frac{1}{3}(\bar{x}_3 + \bar{x}_4 + \bar{x}_5).$$

In der Praxis werden Kontraste dieser Art, bei der die einzelnen Mittelwerte gleichgewichtig eingehen, am häufigksten vorkommen, aber die allgemeine Definition des Kontrastes läßt im Prinzip auch unterschiedliche Gewichtungen der Mittelwerte zu, – die Hauptsache ist, dass $\sum_j c_j = 0$ gilt.

Anmerkung: Die allgemeine Definition des Kontrastbegriffs hat den Hintergrund, dass ganz allgemein gezeigt werden kann, dass die Funktionen $\hat{\Psi}$ biasfreie (verzerrungsfreie) Schätzungen der Ψ sind, was eine sehr wünschenswerte Eigenschaft ist, da sie dazu beiträgt, die Wahrscheinlichkeit Fehlentscheidungen zu minimieren. □

Für diese Anmerkung sowie für die folgenden Aussagen werden in diesem Skript kene Beweise gegeben, da sie ohne Anwendung der Vektor- udn Matrixrechung sehr unübersichtlich werden; man findet sie z.B. in Scheffé (1959).

Die Varianz von $\hat{\Psi}$ ist durch

$$Var(\hat{\Psi}) = s^2_{\hat{\Psi}} = \sum_{j=1}^{p} c_j^2 Var(\hat{x}_j) \tag{2.22}$$

gegeben. Nimmt man an, dass

$$Var(\bar{x}_j) = \frac{\sigma_e^2}{n_j} \tag{2.23}$$

für alle j gilt (σ_e^2 in dieser Formel reflektiert die Annahme gleicher Varianzen für alle Bedingungen). Dann folgt

$$\sigma_\Psi^2 = \sigma_e^2 \sum_{j=1}^p \frac{c_j^2}{n_j} \tag{2.24}$$

$$\hat{\sigma}_{\hat{\Psi}}^2 = s_e^2 \sum_{j=1}^p \frac{c_j^2}{n_j} \tag{2.25}$$

Satz 2.4 *Es gilt*

$$P(\hat{\Psi} - S\hat{\sigma}_{\hat{\Psi}}^2 \leq \Psi \leq \hat{\Psi} + S\hat{\sigma}_{\hat{\Psi}}^2) = 1 - \alpha, \tag{2.26}$$

simultan für alle $\hat{\Psi}$, *wobei*

$$S^2 = (p-1)F_{\alpha;p-1,N-1}. \tag{2.27}$$

.

Beweis: Scheffé (1959), p. 68.

Satz 2.5 *Es sei*

$$D_c = \left(\frac{2(p-1)\sigma_e^2 F_{p-1,m,1-\alpha}}{N}\right)^{1/2}. \tag{2.28}$$

Ist eine beliebige Mittelwertsdifferenz größer als D_c, *so ist sie signifikant auf dem* α*-Niveau,* m *die Anzahl der Freiheitsgrade der Fehlervarianz.*

2.5 Lineare Regression und Varianzzerlegung

Es seien $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_n$ Messwerte für zwei Variablen und es werde die Regression

$$Y_i = aX_i + b + e_i, \quad i = 1, \ldots, n \tag{2.29}$$

betrachtet. Zu schätzen sind die freien Parameter a und b. Die Methode der Kleinsten Quadrate führt auf den Ansatz

$$Q(a,b) = \sum_{i=1}^n (Y_i - aX_i - b)^2, \tag{2.30}$$

und die partiellen Ableitungen nach a bzw b sind

$$\frac{\partial Q}{\partial a} = -2\sum_{i=1}^n (Y_i - aX_i - b)X_i = -2\sum_{i=1}^n e_i x_i \tag{2.31}$$

$$\frac{\partial Q}{\partial b} = -2\sum_{i=1}^n (Y_i - aX_i - b) = -2\sum_{i=1}^n e_i \tag{2.32}$$

Es sei $\partial Q/\partial a = 0$ und $\partial Q/\partial b = 0$ für $a = \hat{a}$, $b = \hat{b}$, woraus sich insbesondere

$$\sum_{i=1}^{n} \hat{e}_i x_i = 0, \quad \sum_{i=1}^{n} \hat{e}_i = 0 \tag{2.33}$$

ergibt, mit $\hat{e}_i = Y_i - \hat{a}x_i - \hat{b}$. Es folgt sofort

$$\sum_{i=1}^{n} \hat{y}_i \hat{e}_i = 0, \tag{2.34}$$

denn

$$\sum_{i=1}^{n} \hat{y}_i \hat{e}_i = \sum_{i=1}^{n} (\hat{a}x_i + \hat{b})\hat{e}_i = \hat{a}\sum_{i=1}^{n} x_i \hat{e}_i + \hat{b}\sum_{i=1}^{n} \hat{e}_i = 0$$

wegen (2.31) und (2.31). (2.34) bedeutet, dass die Kovarianz zwischen den vorhergesagten Werten $\hat{y}_i$ und den Fehlern $\hat{e}_i$ gleich Null ist, d.h. die $\hat{y}_i$ und die $\hat{e}_i$ sind unabhängig voneinander.

Man rechnet leicht nach, dass dann

$$\hat{a} = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(Y_i - \bar{y})}{\sum_{i=1}^{n} X_i^2 - \bar{x}^2} = \frac{Kov(X,Y)}{s_x^2} \tag{2.35}$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \tag{2.36}$$

(die rechte Seite von (2.35) ergibt sich, wenn man Zähler und Nenner mit $1/n$ bzw. $1/(n-1)$ multipliziert). Bekanntlich ist

$$r_{xy} = \hat{a}\frac{s_x}{s_y} = \frac{Kov(X,Y)}{s_x s_y} \tag{2.37}$$

der (Stichproben-)Korrelationskoeffizient mit $-1 \leq r_{xy} \leq 1$.

Setzt man $\hat{Y}_i = \hat{a}X_i + \hat{b}$, so wird (2.29) zu $Y_i = \hat{Y}_i + e_i$. Es seien $x_i = X_i - \bar{x}$, $y_i = Y_i - \bar{y}$, $\hat{y}_i = \hat{Y}_i - \bar{\hat{y}}$, $\varepsilon_i = e_i - \bar{e}$. Dann folgt

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i}{s_x^2}, \quad \hat{y}_i = \hat{a}x_i. \tag{2.38}$$

Weiter ist $y_i = \hat{y}_i + \varepsilon_i$ und wegen (2.34) folgt

$$s_y^2 = s_{\hat{y}}^2 + s_\varepsilon^2, \tag{2.39}$$

so dass

$$\frac{s_{\hat{y}}^2}{s_y^2} = 1 = \frac{s_{\hat{y}}^2}{s_y^2} + \frac{S_\varepsilon^2}{s_y^2}. \tag{2.40}$$

Wegen (2.37) folgt

$$\frac{s_{\hat{y}}^2}{s_y^2} = \hat{a}^2\frac{s_x^2}{s_y^2} = r_{xy}^2 \tag{2.41}$$

r_{xy}^2 heißt *Determinationskoeffizient.* Er ist gleich dem Anteil der Varianz der vorhergesagten Werte $\hat{y}$ and der VArianz der y-Werte, und offenbar gilt

$$r_{xy}^2 = 1 - \frac{s_\varepsilon^2}{s_y^2} \Rightarrow \frac{s_\varepsilon^2}{s_y^2} = 1 - r_{xy}^2. \tag{2.42}$$

r_{xy}^2 ist ein Maß für die Güte der Vorhersage der y-Werte durch die x-Werte. Aus der rechten Seite von (2.42) ergibt sich

$$s_\varepsilon = s_y^2\sqrt{1 - r_{xy}^2}. \tag{2.43}$$

s_ε ist der *Standardmessfehler.*

Varianzzerlegung Die Frage ist nun, ob die Vorhersage besser als erwartet ist, wenn der wahre Wert ρ_{xy} gleich Null ist, d.h. ob r_{xy} signifikant ist. Dazu kann eine Zerlegung der Varianz der Y-Werte durchgeführt werden.

Die Varianz der y-Werte kann wieder wie beim einfaktoriellen Fall in Varianzkomp0onenten zerlegt werden: es sei n_j die Anzahl der Fälle, die für einen gegebenen x_j-Wert beobachtet werden, und $N = \sum_{j=1}^p n_j$. Weiter sei

$$\bar{y} = \frac{1}{N}\sum_{j=1}^p\sum_{i=1}^{n_j} y_{ij}.$$

Dann ist

$$s_y^2 = \frac{1}{N}\sum_{j=1}^p\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2 = \frac{QS_{ges}(y)}{N}. \tag{2.44}$$

Wie im einfaktoriellen Fall gilt

$$QS_{ges}(y) = QS_{inn}(y) + QS_{zw}(y) \tag{2.45}$$

mit

$$QS_{inn}(y) = \sum_{j=1}^p\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_i)^2, \quad QS_{zw}(y) = \sum_{j=1}^p n_j(\bar{y}_j - \bar{y})^2 \tag{2.46}$$

Diese Zerlegung gilt unabhängig davon, ob die lineare Regression $Y = aX + b + e$ gilt oder nicht. Um dieser Frage nachzugehen, kann man $QS_{zw}(y)$ in eine Komponente $QS_{Abw.lin.Reg}$ und in eine Komponente $QS_{lin.Reg.}$ zerlegt werden:

$$QS_{inn} = QS_{Abw.lin.Reg} + QS_{lin.Reg.}, \tag{2.47}$$

mit

$$QS_{lin.Reg} = \sum_{j=1}^p n_j(\hat{y}_j - \bar{y})^2 \tag{2.48}$$

$$QS_{Abw.lin.Reg} = \sum_{j=1}^p n_j(\bar{y}_j - \hat{y}_j)^2 \tag{2.49}$$

Es ist $y_{ij} - \hat{y}_j = e_{ij}$; Mittelung liefert $\bar{y}_j - \hat{y}_j = e_{j.} = \bar{e}_j$. Weiter ist $\hat{y}_j - \bar{y} = \hat{a}(x_j - \bar{x})$. Dann folgt

$$\begin{aligned}
\sum_{j=1}^{p} n_j(\bar{y}_j - \hat{y}_j)(\hat{y}_j - \bar{y}) &= \hat{a} \sum_{j=1}^{p} n_j \bar{e}_j (x_j - \bar{x}) \\
&= \hat{a} \left(\sum_{j=1}^{p} n_j (\bar{e}_j x_j - \bar{x} \sum_{j=1}^{p} n_j \bar{e}_j \right),
\end{aligned}$$

Es ist

$$\sum_{j=1}^{p} \sum_{i=1}^{n_j} e_{ij} x_j = \sum_{j=1}^{p} x_j \sum_{i=1}^{n_j} e_{ij} = \sum_{j=1}^{p} n_j \bar{e}_j,$$

und $\sum_i e_{ij} = n_j \bar{e}_j$. Aber dann ist $\sum_{j=1}^{p} \sum_{i=1}^{n_j} e_{ij} x_i = 0$ und $\sum_{j=1}^{p} \sum_{i=1}^{n_j} e_{ij} = 0$ und das Kreuzprodukt der ersten Gleichung verschwindet, womit (2.47) gezeigt ist. □

3 Mehrfaktorielle Designs mit festen Effekten

3.1 Das 2-faktorielle Design mit festen Effekten

Die Varianz σ^2 ist um so größer, je weniger Variable, die auf die interessierende abhängige Variable einwirken können, explizit kontrolliert werden. Dies kann bewirken, daß die Unterschiede zwischen den Stufen eines Faktors nicht bemerkt werden, weil σ^2 relativ zur Größe der Unterschiede zu groß ist. So können Unterschiede zwischen Therapieformen unerkannt bleiben, weil systematische Unterschiede zwischen weiblichen und männlichen Klienten in den Stichproben die Varianz der Meßwerte groß werden lassen im Vergleich zu den durchschnittlichen, also mittleren Unterschieden zwischen den Therapien. Untersucht man die Therapien für weibliche und männliche Klienten getrennt, so können die Unterschiede allerdings sichtbar werden. Man sagt, daß die Variable (der "Faktor") Geschlecht nun "kontrolliert" wird. Diese Kontrolle bedeutet, daß man auch die Art der Unterschiede zwischen weiblichen und männlichen Klienten erfassen kann: so könnte z.B. die Therapie bei männlichen Klienten einen positiven Effekt, bei weiblichen aber einen negativen Effekt haben; man spricht dabei von einer *Wechselwirkung* (Interaktion) zwischen den Geschlechtern. Mittelt man über weibliche und männliche Klienten, kann es dabei geschehen, daß *im Mittel* kein Unterschied zwischen den Geschlechtern sichtbar wird. In jedem Fall hat man einen 2-dimensionalen oder 2-faktoriellen Versuchsplan. Natürlich kann man noch weitere Variable kontrollieren, etwa das Alter, den Bildungsstand, etc. Ein wesentliches Merkmal mehrfaktorieller Versuchspläne ("Designs"), nämlich die Möglichkeit, Hypothesen über Wechselwirkungen zwischen unabhängigen Variablen überprüfbar zu machen, kann bereits an einem 2-faktoriellen Versuchsplan erläutert werden.

Es sei x_{ijk} ein Meßwert der betrachteten abhängigen Variablen (etwa: Befindlichkeit nach einer Therapie).Der Index i zeigt die i-te Stufe des Faktors A (etwa: Geschlecht) an, wobei $i = 1, \cdots, I$, I die Anzahl der Stufen des Faktors (Geschlecht: $I = 2$). j steht für die j-te Stufe des Faktors B (z.B. Art der Therapie), $J = 1, \cdots, J$. k indiziert die Meßwerte in einer Stichprobe. Es soll angenommen werden, daß $k = 1, \cdots, n$ für alle Faktorkombinationen (i, j). Betrachtet wird das

Modell: Es wird angenommen, daß sich die Messungen gemäß

$$\begin{aligned} x_{ijk} &= \mu_{ij} + e_{ijk}, \quad i = 1, \cdots, n_i; \quad j = 1, \cdots, k, \quad k = 1, \cdots, n \\ e_{ijk} &\sim N(0, \sigma^2), \quad \text{Kov}(e_{ijk}, e_{i'j'k'}) = 0 \end{aligned} \tag{3.1}$$

repräsentieren lassen. Es wird also

$$E(x_{ijk}) = \mu_{ij}$$

angenommen. μ_{ij} repräsentiert die Wirkung der Kombination der i-ten und der j-ten Bedingung. Die Fehler sollen normalverteilt sein mit dem Erwartungswert

0 und der für alle Bedingungen gleichen Varianz σ^2; darüber hinaus werden sie als unkorreliert und damit als unabhängig vorausgesetzt. Diese Annahmen entsprechen denen für den einfaktoriellen Fall.

Weiter sei

$$\mu = \frac{1}{IJ}\sum_{ij}\mu_{ij}, \quad \mu_{i.} = \frac{1}{J}\sum_{j}\mu_{ij}, \quad \mu_{.j} = \frac{1}{I}\sum_{i}\mu_{ij}. \tag{3.2}$$

$\mu_{i.}$ und $\mu_{.j}$ sind die Erwartungswerte der $x_{i.}$ und der $x_{.j}$, die sich ergeben, wenn man über die Werte eines Faktors mittelt. Die $x_{.j}$ etwa sind die über die Stufen des Faktors A gemittelten Werte; diese Werte würde man analysieren, würde man eine 1-faktorielle Varianzanalyse bezüglich B rechnen. Die Bedeutung der $x_{i.}$ ist analog; ihre Analyse entspräche einer 1-faktoriellen Analyse des Faktors A.

Es sei nun

$$\alpha_i = \mu_{i.} - \mu, \quad \beta_j = \mu_{.j} - \mu, \quad \gamma_{ij} = \mu_{ij} - \alpha_i - \beta_j - \mu \tag{3.3}$$

Offenbar gilt dann

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = \sum_{ij} \gamma_{ij} = 0 \tag{3.4}$$

Dann folgt

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \tag{3.5}$$

bzw.

$$\mu_{ij} = \mu_{i.} - \mu + \mu_{j.} - \mu + \gamma_{ij} + \mu \tag{3.6}$$

woraus

$$\gamma_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu. \tag{3.7}$$

folgt. Für die messung x_{ijk} ergibt sich der Ausdruck

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \tag{3.8}$$

(3.8) ist die *Strukturgleichung* der 2-faktoriellen Varianzanalyse. α_i ist ein Effekt des Faktors A, β_j ist ein Effekt des Faktors B. Die einfaktoriellen Analysen bezüglich des Faktors A bzw. B beziehen sich auf diese Effekte. α_i und β_j werden als *Haupteffekte* bezeichnet. Neu ist die Größe γ_{ij}. Sie kennzeichnet die Wechselwirkung oder Interaktion der Faktoren A und B. Interaktionen sind Mittelwertsunterschiede, die nicht aufgrund eines Faktors allein auftreten. So kann es sein, daß Frauen nach einer Therapie der Art 1, Männer aber nach einer Therapie der Art 2 eine bessere Befindlichkeit zeigen. Insbesondere ist es möglich, daß *im Durchschnitt* Männer und Frauen gleichermaßen eine bessere Befindlicheit zeigen, wenn man über die Therapien mittelt (bzw. gar nicht erst nach unterschiedlichen Therapien aufschlüsselt); dann existiert kein Haupteffekt für A. Ebenso kann es sein, daß kein Haupteffekt für B existiert: gemittelt über weibliche und männliche Klienten sind dann alle Therapien gleich gut. Einfaktorielle Analysen können die Unterschiede, die in Form von Wechselwirkungen existieren, nicht erfassen. Neben der Reduktion der Fehlervarianz ist also die Möglichkeit, Unterschiede zwischen der Wirkung der Faktoren, die nur als Interaktionen existieren, eine weitere, wesentliche Motivation für mehrfaktorielle Versuchspläne.

Zur überprüfung der Haupt- und Wechselwirkungseffekte werden wieder F-Tests konstruiert, die wieder auf einer Zerlegung der Gesamtvarianz der Daten beruhen.

3.1.1 Zerlegung der Gesamtvarianz

Es werden zunächst die folgenden Mittelwerte definiert

$$\bar{x} = \frac{1}{nIJ}\sum_{ijk} x_{ijk}, \quad \bar{x}_{ij} = \frac{1}{n}\sum_{k} x_{ijk} \tag{3.9}$$

$$\bar{x}_{i.} = \frac{1}{nJ}\sum_{jk} x_{ijk}, \quad \bar{x}_{.j} = \frac{1}{nI}\sum_{ik} x_{ijk} \tag{3.10}$$

Diese Mittelwerte können als Schätzungen für die Parameter μ, μ_{ij}, etc aufgefaßt werden.

Satz 3.1 *Es sei*

$$QS_{ges} = \sum_{ijk}(x_{ijk} - \bar{x})^2 \tag{3.11}$$

Dann gilt die Zerlegung

$$\begin{aligned} QS_{ges} &= \sum_{ik}(x_{ijk} - \bar{x}_{ij})^2 + nJ\sum_{i}(\bar{x}_{i.} - \bar{x})^2 \\ &\quad + nI\sum_{j}(\bar{x}_{.j} - \bar{x})^2 + n\sum_{ij}(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2, \end{aligned} \tag{3.12}$$

d.h.

$$QS_{ges} = QS_{inn} + QS_{zw.A} + QS_{zw.B} + QS_{Rest} \tag{3.13}$$

Beweis: Mit $b = \bar{x}_{ij}$ liefert der Hilfssatz (1.2), angewendet auf QS_{ges},

$$\sum_{ijk}(x_{ijk} - \bar{x})^2 = \sum_{ijk}(x_{ijk} - \bar{x}_{ij})^2 + n\sum_{ij}(\bar{x}_{ij} - \bar{x})^2$$

Setzt man nun $b = \bar{x}_i$, so erhält man aus der Anwendung des Hilfssatzes (1.2) es auf die zweite Summe auf der rechten Seite

$$n\sum_{ij}(\bar{x}_{ij} - \bar{x})^2 = n\sum_{ij}(\bar{x}_{ij} - \bar{x}_i)^2 + nJ\sum_{i}(\bar{x}_i - \bar{x})^2$$

Schließlich sei $b = \bar{x}_i + \bar{x}_j - \bar{x}$. Dann findet man aufgrund des Hilfsatzes

$$n\sum_{ij}(\bar{x}_{ij} - \bar{x}_i)^2 = n\sum_{ij}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = n\sum_{ij}(\bar{x}_i - (\bar{x}_i + \bar{x}_j - \bar{x}))^2$$

Aber es ist $\bar{x}_i - (\bar{x}_i + \bar{x}_j - \bar{x}) = \bar{x} - \bar{x}_j$, so daß schließlich

$$n\sum_{ij}(\bar{x}_{ij} - \bar{x}_i)^2 = n\sum_{ij}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

Setzt man diese Ausdrücke ein, so ergibt sich die Behauptung. (Hier wird von der Tatsache Gebrauch gemacht, daß $(-(\bar{x} - \bar{x}_j))^2 = (\bar{x}_j - \bar{x})^2$. □

Die Zerlegung ist so gewählt worden, daß die Gesamtvariation in Komponenten aufgespalten wird, die auf die Wirkung der Faktoren und deren Wechselwirkung zurückzuführen ist. Die Wechselwirkung oder Interaktion wird dabei duch diejenige Varianzkomponente erklärt, die durch die additive Wirkung der Faktoren *nicht* erklärt werden kann; sie ist durch QS_{Rest} in (3.13) gegeben.

Es sei noch darauf hingewiesen, daß alle Kreuzprodukte, d.h.

$$\sum_{ij}(x_{ijk} - \bar{x}_{ij})(\bar{x}_{i.} - \bar{x})$$

etc, verschwinden; dies bedeutet, daß die Abweichungen $x_{ijk} - \bar{x}_{ij}$, $\bar{x}_{i.} - \bar{x}$ etc. unabhängig voneinander variieren. Dieser Sachverhalt ermöglicht wieder die Konstruktion von F-Tests.

Die erste Summe auf der rechten Seite ist die Quadratsumme "innerhalb", QS_{inn}. Sie repräsentiert die gemittelten ("pooled") Stichprobenvarianzen und kann deshalb zur Schätzung der Fehlervarianz σ^2 benutzt werden.

Die zweite Quadratsumme repräsentiert die Variation der Stufen des Faktors A um den Gesamtmittelwert. Sie korrespondiert zu QS_{zw} in einer einfaktoriellen Analyse des Faktors A, sie soll hier QS_A genannt werden. Die dritte Quadratsumme hat die analoge Bedeutung für den Faktor B und soll deswegen mit QS_B bezeichnet werden.

Die letzte Quadratsumme repräsentiert offenbar die Effekte der Wechselwirkung der Faktoren A und B (vergl. (3.7). Es ist

$$(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}) = (\bar{x}_{ij} - \bar{x}) - (\bar{x}_{i.} - \bar{x}) - (\bar{x}_{.j} - \bar{x});$$

man sieht, daß die Summe den "Rest" an Variation angibt, der *nicht* durch das additive Zusammenwirken der Faktoren erklärt werden kann. Die Quadratsumme wird deshalb gelegentlich auch mit QS_{Rest} bezeichnet, hier wird sie $QS_{A\times B}$ genannt. Die Gleichung (3.12) kann symbolisch in der Form

$$QS_{ges} = QS_{inn} + QS_A + QS_B + QS_{A\times B} \tag{3.14}$$

geschrieben werden. Ebenso kann man schreiben

$$\sum_{ijk}(x_{ijk}-\mu)^2 = nJ\sum_i \alpha_i^2 + nI\sum_j \beta_j^2 + n\sum_{ij}\gamma_{ij}^2 + \sum_{ijk} e_{ijk}^2 \tag{3.15}$$

Aus den Quadratsummen können F-Tests zur überprüfung der Signifikanz der Haupt- und Wechselwirkungseffekte gebildet werden. Dazu muß wieder die Stichprobenverteilung der Summen gefunden werden. Es gilt der

Satz 3.2 *Es gelte $H_0^A : \alpha_i = 0$ für alle i, $H_0^B : \beta_j = 0$ für alle j, und $H_0^{A\times B} : \gamma_{ij} = 0$ für alle i, j. Dann haben die Quadratsummen die folgenden Verteilungen:*

$$\begin{aligned}
\frac{QS_{ges}}{\sigma^2} &= \chi^2_{IJn-1} & (3.16)\\
\frac{QS_{inn}}{\sigma^2} &= \chi^2_{IJ(n-1)} & (3.17)\\
\frac{QS_A}{\sigma^2} &= \chi^2_{I-1} & (3.18)\\
\frac{QS_B}{\sigma^2} &= \chi^2_{J-1} & (3.19)\\
\frac{QS_{A\times B}}{\sigma^2} &= \chi^2_{(I-1)(J-1)} & (3.20)
\end{aligned}$$

Beweis: (3.16): Es wird der Hilfssatz (1.2) mit $b = \mu$ auf die Quadratsumme QS_{ges} angewandt. Man erhält sofort

$$QS_{ges} = \sum_{ijk}(x_{ijk}-\mu)^2 - IJn(\bar{x}-\mu)^2$$

Nach Division durch σ^2 hat man

$$\frac{QS_{ges}}{\sigma^2} = \sum_{ijk}\left(\frac{x_{ijk}-\mu}{\sigma}\right)^2 - \left(\frac{\bar{x}-\mu}{\sigma/\sqrt{IJn}}\right)^2$$

Die erste Summe geht über alle IJn Werte, so daß $QS_{ges}/\sigma^2 = \chi^2_{IJn} - \chi^2_1 = \chi^2_{IJn-1}$, und dies ist behauptet worden.

(3.17): Mit $b = \mu_{ij}$ erhält man über den Hilfssatz (1.2)

$$\begin{aligned}
QS_{inn} &= \sum_{i,j}\left(\sum_k (x_{ijk}-\mu_{ij})^2 - n(\bar{x}_{ij}-\mu_{ij})^2\right)\\
&= \sum_{ijk}(x_{i,j,k}-\mu_{ij})^2 - n\sum_{i,j}(\bar{x}_{ij}-\mu)^2 \qquad (3.21)
\end{aligned}$$

Nach Division durch σ^2 sieht man, daß QS_{inn}/σ^2 wie $\chi^2_{nIJ} - \chi_{IJ} = \chi^2_{IJ(n-1)}$-verteilt ist.

(3.18): Mit $b = \mu$ erhält man über den Hilfssatz (1.2)

$$\sum_i (\bar{x}_i - \bar{x})^2 = \sum_i (\bar{x}_i - \mu)^2 - n(\bar{x} - \mu)^2,$$

so daß

$$\frac{QS_A}{\sigma^2} = I \sum_i \left(\frac{\bar{x}_i - \mu}{\sigma/\sqrt{nJ}} \right)^2 - \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{nIJ}} \right)^2,$$

Die rechte Seite ist demnach wie $\chi_I^2 - \chi_1^2 = \chi_{I-1}^2$-verteilt.

(3.19): Analog zu (3.18)

(3.20): Der Beweis kann analog wie die vorangegangenen geführt werden, was sich aber als verhältnismäßig länglich erweist. Andererseits folgt aus (3.14)

$$QS_{A\times B} = QS_{ges} - QS_A - QS_B - QS_{inn}.$$

Nach Division durch σ^2 sieht man, daß die rechte Seite eine Summe von χ^2-verteilten Variablen ist. Mithin muß $QS_{A\times B}/\sigma^2$ ebenfalls χ^2-verteilt mit ν Freiheitsgraden sein. Die Freiheitsgrade ν ergeben sich aus denen der Größen auf der rechten Seite: $\nu = nIJ - 1 - (I-1) - (J-1) - IJ(n-1)$; also ist die gesuchte Anzahl $\nu = (I-1)(J-1)$. $\square$.

Die F-Tests für die drei möglichen Nullhypothesen können nun direkt angeschrieben werden. Da stets getestet werden soll, ob sich die Quadratsummen, die zu den Effekten korrespondieren, signifikant von der Quadratsumme unterscheidet, die die Fehlervarianz repräsentiert, und die letztere durch QS_{inn} gegeben ist, sollte sie im Nenner des F-Bruchs stehen. Es ergeben sich die folgenden F-Brüche:

Satz 3.3 *Gilt die Nullhypothese H_0^A, so ist der Quotient*

$$F^A = \frac{QS_A/(I-1)}{QS_{inn}/IJ(n-1)} \tag{3.22}$$

verteilt wie $F_{I-1,IJ(n-1)}$. Gilt H_0^B, so ist der Quotient

$$F^B = \frac{QS_B/(J-1)}{QS_{inn}/IJ(n-1)} \tag{3.23}$$

wie $F_{J-1,IJ(n-1)}$ verteilt. Gilt schließlich $H_0^{A\times B}$, so ist der Quotient

$$F^{A\times B} = \frac{QS_{A\times B}/(I-1)(J-1)}{QS_{inn}/IJ(n-1)} \tag{3.24}$$

$F_{(I-1)(J-1),IJ(n-1)}$-verteilt.

Anmerkung: Der Fall $n = 1$ ist offenbar nicht abgedeckt, da dann die F-Brüche F^A, F^B und $F^{A\times B}$ nicht mehr definiert sind. Zweifaktorielle Versuchsanordnungen für den Fall $n = 1$ müssen deshalb gesondert betrachtet werden.

Beweis: Der Beweis folgt aus der Unabhängigkeit der Varianzkomponenten sowie aus den Aussagen des Satzes 3.2. □

Die Resultate werden wieder in einer Tabelle der Form Tab. 2 zusammengefaßt.

Tabelle 2: Varianzanalyse, 2-faktoriell

Quelle d. Var.	Quadrats.	Freiheitsgr.	F-Test
A	QS_A	$I-1$	$\frac{QS_A/(I-1)}{QS_{inn}/(IJ(n-1))}$
B	QS_B	$J-1$	$\frac{QS_B/(J-1)}{QS_{inn}/(IJ(n-1))}$
$A\times B$	$QS_{A\times B}$	$(I-1)(J-1)$	$\frac{QS_{A\times B}/(I-1)(J-1)}{QS_{inn}/IJ(n-1)}$
Fehler	QS_{inn}	$IJ(n-1)$	
Gesamt	QS_{ges}	$IJn-1$	

über die Existenz von Interaktionen kann man sich bereits auf graphischem Wege Informationen verschaffen. Denn angenommen, es gäbe keine Wechselwirkungen. Dann gilt ja für alle x_{ijk} das Modell

$$x_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$

bzw. für die Mittelwerte

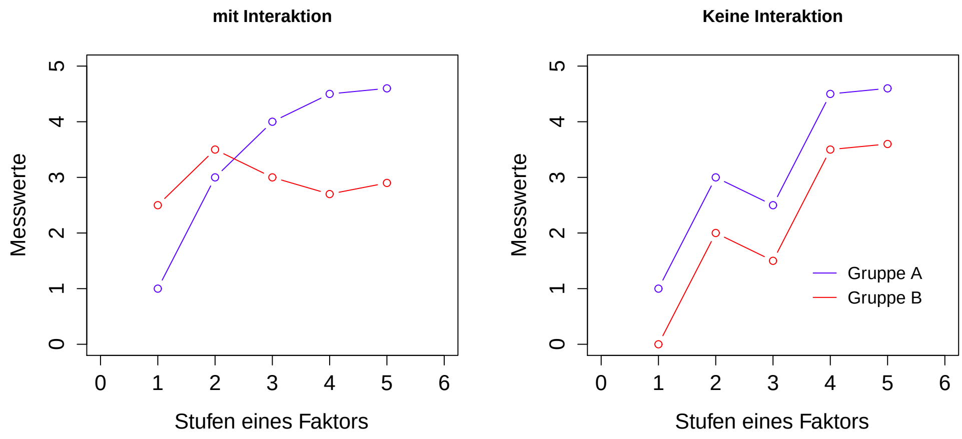
$$\bar{x}_{ij} = \mu + \alpha_i + \beta_j + e_{ij\cdot} \tag{3.25}$$

wobei $e_{ij\cdot}$ der mittlere (Stichproben-) Fehler unter der Bedingung (A_i, B_j) ist. Man kann nun die Mittelwerte $\overline{x}_{ij}$ für festgehaltene Stufe A_i in ein Koordinatensystem eintragen, also etwa für A_1 die Mittelwerte $\bar{x}_{1j}$ gegen $j = 1, \cdots, J$ abtragen. Das gleiche macht man mit den Mittelwerten $\bar{x}_{2j}$, bei denen also die Stufe A_2 festgehalten wird. Die Kurven für A_1, A_2 etc. sollten dann parallel sein, denn nach (3.25) unterscheiden sich die Kurven nur aufgrund der additiven Konstanten α_i und den zufälligen Fehler $e_{ij\cdot}$. Existiert aber eine Interaktion zwischen den Faktoren A und B, so tritt ja noch ein Term γ_{ij} auf,

$$\bar{x}_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij\cdot} \tag{3.26}$$

Da dieser Term von i und j abhängt, ist er keine Konstante. Er bewirkt deshalb, daß die Kurven *nicht* parallel sein werden. Ein Beispiel für solche Abbildungen wird in Abb. 1 gegeben.

Abbildung 1: Der Effekt von Interaktionen



3.1.2 Der Fall $n = 1$

Im vorangegangenen Abschnitt war $n > 1$ vorausgesetzt worden; der Fall $n = 1$ kann nicht aus den Betrachtungen für $n > 1$ abgeleitet werden und wird deswegen gesondert behandelt.

Für $n = 1$ ist $\bar{x}_{ij} = x_{ij}$ und $QS_{inn} = 0$; damit können die F-Tests nicht mehr berechnet werden, da die Nenner gleich Null sind. Der Fall $n = 1$ kann also nicht einfach als Spezialfall aus den Betrachtungen für $n > 1$ gewonnen werden. Ist $QS_{inn} = 0$, so wird QS_{ges} nur noch in drei Variationsanteile zerlegt: zwei zu Lasten der Haupteffekte und einen, der zur überprüfung der Signifikanz von Wechselwirkungen benutzt wurde. Will man nun die Haupteffekte testen, so kann man sie nur noch gegen diesen Interaktionsterm testen (es wird gleich gezeigt werden, daß alle Quadratsummen wieder χ^2-verteilt sind), denn unter der Bedingung, daß eine Interaktion und der jeweils betrachtete Haupteffekt nicht existieren sind die entsprechenden Quadratsummen unabhängig χ^2-verteilt und die Quotienten F-verteilt. Die Annahme, daß keine Interaktionen existieren, ist also notwendig, damit die Analyse durchgeführt werden kann.

Man geht wieder von dem allgemeinen Ansatz

$$x_{ij} = \mu_{ij} + e_{ij} \tag{3.27}$$

aus.

Es werde wieder $\mu = \sum_{ij} \mu_{ij}/IJ$, $\mu_{i.} = \sum_j \mu_{ij}/J$, $\mu_{.j} = \sum_i \mu_{ij}/I$ gesetzt. Unter $H_0(A)$ ist $\alpha_i = 0$ für alle i, und unter $H_0(B)$ ist $\beta_j = 0$ für alle j. Für die

Gesamtvarianz findet man die folgende Quadratsumme:

$$QS_{ges} = \sum_{ij} (x_{ij} - \bar{x})^2. \tag{3.28}$$

Diese Quadratsumme läßt sich so zerlegen, daß sich zumindest die Haupteffekte testen lassen. Dazu betrachte man die folgende Zerlegung der Differenz $x_{ij} - \bar{x}$:

$$x_{ij} - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.j} - \bar{x}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})$$

Quadriert man $x_{ij} - \bar{x}$ und summiert über i und j, so ergibt sich

$$\begin{aligned} QS_{ges} &= \sum_{ij} (x_{ij} - \bar{x})^2 = \\ &= \sum_{ij} (\bar{x}_{i.} - \bar{x})^2 + \sum_{ij} (\bar{x}_{.j} - \bar{x})^2 + \\ &\quad + \sum_{ij} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 \end{aligned} \tag{3.29}$$

Man zeigt jetzt wieder, daß die Quadratsummen - geteilt durch σ^2 - unter der Nullhypothese χ^2-verteilt sind. Auf $\sum_{ij}(x_{ij} - \bar{x})^2$ läßt sich sofort der Hilfssatz (1.2) mit $b = \mu$ anwenden:

$$QS_{ges} = \sum_{ij} (x_{ij} - \mu)^2 - IJ(\bar{x} - \mu)^2$$

Division durch σ^2 impliziert, daß $\sum_{ij}((x_{ij}-\mu)/\sigma)^2$ χ^2_{IJ}-verteilt ist, d.h. χ^2-verteilt mit IJ Freiheitsgraden. Für den zweiten Term ist $IJ(\bar{x}-\mu)^2/\sigma^2 = (\bar{x}-\mu)/\sigma\sqrt{IJ})^2$ gerade χ^2_1-verteilt, so daß QS_{ges} $\chi^2_{(IJ-1)}$-verteilt ist.

Für $QS_A = \sum_{ij}(\bar{x}_{i.} - \bar{x})^2$ findet man, nach Anwendung des Hilfssatzes (1.2) mit $b = \mu$ und nach Division durch σ^2,

$$\frac{QS_A}{\sigma^2} = \sum_{ij} \left(\frac{\bar{x}_{i.} - \mu}{\sigma/\sqrt{J}} \right)^2 - \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{I}} \right)^2 \tag{3.30}$$

so daß QS_A/σ^2 wie $\chi^2_{(I-1)}$-verteilt ist. Analog findet man, daß QS_B wie $\chi^2_{(J-1)}$ verteilt ist. Die dritte Summe werde mit QS_{Rest} bezeichnet. Da

$$QS_{Rest} = QS_{ges} - QS_A - QS_B, \tag{3.31}$$

folgt, daß QS_{Rest} χ^2-verteilt mit $IJ - (I-1) - (J-1) = (I-1)(J-1)$ Freiheitsgraden ist.

Es wird jetzt deutlich, daß man im Falle $n = 1$ tatsächlich keine Interaktion testen kann. Zwar erinnert die dritte Quadratsumme in (3.29) an die Quadratsumme, die im Falle $n > 1$ die Wechselwirkung definierte, aber es gibt eben keine

Quadratsumme QS_{inn}, "gegen" die man die Wechselwirkung testen könnte. In der Tat ist QS_{Rest} die einzige Quadratsumme, gegen die man die durch QS_A und QS_B definierten Haupteffekte testen kann. Denn man kann ja die F-Brüche

$$F_{I-1,(I-1)(J-1)} = \frac{QS_A/(I-1)}{QS_{Rest}/(I-1)(J-1)} \tag{3.32}$$

und

$$F_{J-1,(I-1)(J-1)} = \frac{QS_B/(J-1)}{QS_{Rest}/(I-1)(J-1)} \tag{3.33}$$

definieren; unter den Nullhypothesen $H_0(A)$ und $H_0(B)$ sind sie mit den indizierten Freiheitsgraden F-verteilt.

3.2 3- und mehrfaktorielle Versuchspläne

Es ist im Prinzip ohne weiteres möglich, nicht nur zwei, sondern drei und mehr Faktoren zu berücksichtigen. Wird ein dritter Faktor C mit den Stufen $C_1, \cdots, C_k$ eingeführt, so erhält man die allgemeine Strukturgleichung

$$x_{ijks} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + e_{ijks} \tag{3.34}$$

Hier bezeichnet nun γ_k den Haupteffekt der k-ten Stufe von C. Es gibt drei Interaktionen $A \times B$, $A \times C$ und $B \times C$ sowie eine Interaktion $A \times B \times C$, die die Wechselwirkung zwischen allen drei Faktoren repräsentiert.

Für vier Faktoren mag man sich die Strukturgleichung zur übung selbst ausdenken. Die Rechnungen folgen immer dem gleichen Schema und sollen hier nicht weiter ausgeführt werden. Man sich sich leicht klar, daß die Anzahl der Beobachtungen (Vpn) sehr schnell mit der Anzahl der Faktoren steigt. Hat jeder Faktor etwa 3 Stufen, so benötigt man bei einem 2-faktoriellen Plan bereits $3 \times 3 = 9$ Stichproben (Mindestumfang der Stichproben: $n = 2$, da sonst keine Inteaktionen getestet werden können.) Bei drei Faktoren hat man $3 \times 3 \times 3 = 27$ Stichproben, bei vier Faktoren hat man bereits 81 Stichproben.

3.3 Zufällige Faktoren (Random Factors)

3.3.1 Der 1-faktorielle Fall

Vielfach ist es nicht möglich, den Wertebereich möglicher Faktorabstufungen vollständig in einem Versuch zu untersuchen. So kann man fragen, ob die Reihenfolge der Darbietung der Testkarten beim Rohrschach-Test einen Einfluß auf die Art der Antworten hat, die Testpersonen geben. Umfaßt der Test 10 Karten, so gibt es $10! = 3628800$ mögliche Reihenfolgen. Offenbar kann man nicht alle Reihenfolgen als Stufen eines Faktors "Reihenfolge"in einen Versuchsplan aufnehmen. Vielmehr wird man eine Stichprobe aus der Menge der möglichen Reihenfolgen auswählen und dann versuchen, von den Resultaten (Effekt oder kein Effekt) auf die Menge aller Reihenfolge zu schließen.

Der Effekt einer zufällig gewählten Stufe werde für die i-te Stufe mit a_i bezeichnet. a_i ist jetzt keine feste Größe mehr, sondern muß als zufällige Veränderliche angesehen werden. Denn der Effekt kann ja von Stufe zu Stufe variieren, und in dem Sinne, in dem die die Stufen zufällig ausgewählt wurden, müssen auch die Werte von a_i zufällig sein. So ist es auch möglich, daß bestimmte a_i gleich Null sind, obwohl nicht alle Effekte gleich Null sind. Man muß demnach die Formulierung der Nullhypothese so verallgemeinern, daß sie sich auf die gesamte Population bezieht. Dazu kann man die Varianz σ_A^2 der a_i betrachten: sind alle $a_i = 0$, so muß auch $\sigma_A^2 = 0$ sein. Demnach definiert man nun

$$H_0 : \sigma_A^2 = 0. \tag{3.35}$$

Es wird angenommen, daß die a_i normalverteilt und von den Fehlern e_{ij} stochastisch unabhängig sind.

Bei einem Versuch mit festen Effekten gilt stets $\sum_i a_i = 0$, unabhängig davon, ob die individuellen a_i ungleich Null sind oder nicht. Analog dazu läßt sich sagen, daß diese Summe gleich Null sein wird, wenn man über *alle* möglichen Effekte der Population möglicher Effekte summiert. Für die gewählte Stichprobe von Stufen muß dies allerdings nicht gelten, so daß für die betrachtete Stichprobe von Stufen $\sum_i a_i \neq 0$ sein kann (und im allgemeinen sein wird). Für die Meßwerte läßt sich nun die Strukturgleichung

$$x_{ij} = \mu + a_i + e_{ij} \tag{3.36}$$

schreiben; weiter sei $\bar{a} = \sum_i a_i/I$, I die Anzahl der untersuchten Stufen des Faktors. Dann ist

$$\bar{x} = \frac{1}{n}\sum_{ij} x_{ij} = \mu + \bar{a} + bare$$

wobei *bare* der durchschhnittliche Fehler ist. Dann ist

$$\begin{aligned} QS_{inn} &= \sum_{ij}(x_{ij} - \bar{x})^2 \\ &= \sum_{ij}(\mu + a_i + e_{ij} - \mu - \bar{a} - \overline{e})^2 \\ &= \sum_{ij}(a_i - \bar{a} + e_{ij} - bare)^2 \\ &= \sum_{ij}(a_i - \bar{a})^2 + \sum_{ij}(e_{ij} - \overline{e})^2 \end{aligned} \tag{3.37}$$

Die "Kreuzprodukte"verschwinden wieder wegen der (postulierten) Unabhängigkeit der a_i und der e_{ij}.

Ebenso läßt sich QS_{zw} aufschhlüsseln. Da $\bar{x}_i = \sum_j x_{ij}/n_i = \sum_j(\mu + a_i + e_{ij}) = \mu + a_i + bare_i$, hat man

$$\begin{aligned}
QS_{zw} &= \sum_{ij}(\bar{x}_i - \bar{x})^2 \\
&= \sum_{ij}(\mu + a_i + e_{ij} - \mu - \bar{a} - bare_i)^2 \\
&= \sum_i n_i(a_i - \bar{a} + bare_i - \bar{e})^2 \\
&= \sum_i n_i(a_i - \bar{a})^2 + \sum_i n_i(bare_i - bare)^2 \qquad (3.38)
\end{aligned}$$

Man sieht, daß die Variation der a_i in QS_{zw} in der gleichen Weise eingeht wie in QS_{inn}; H_0 kann deshalb in der gleichen Weise mit

$$F_{I-1,n-I} = \frac{QS_{zw}/(I-1)}{QS_{inn}/(n-I)} \qquad (3.39)$$

getestet werden wie im Modell mit festen Effekten.

3.3.2 Der 2-faktorielle Fall

Hier gilt

$$x_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk} \qquad (3.40)$$

Wieder wird angenommen, daß die a_i und die b_j von den Fehlern stochastisch unabhängig sind, außerdem sollen sie in bezug aufeinander unabhängig sein.

Die a_i, b_j und c_{ij} sind zufällige Variable, deren Summe, über die Stufen, nicht notwendig Null wird. Weiter gilt

$$\begin{aligned}
\bar{x}_i &= \mu + a_i + \bar{b} + \bar{c}_{i.} + bare_{i.} \qquad (3.41) \\
\bar{x}_j &= \mu + \bar{a} + b_j + \bar{c}_{.j} + bare_{j.} \qquad (3.42) \\
\bar{x} &= \mu + \bar{a} + \bar{b} + \bar{c} + bare \qquad (3.43)
\end{aligned}$$

Dann kann man QS_{inn}, QS_A, QS_B, und $QS_{A\times B}$ wie folgt darstellen:

$$\begin{aligned}
QS_{inn} &= \sum_{ijk}(\mu + a_i + b_j + c_{ij} + e_{ijk} - a_i - b_j - c_{ij} - bare_{ij})^2 = \\
&= \sum_{ijk}(e_{ijk} - bare_{ij.})^2 \qquad (3.44)
\end{aligned}$$

QS_{inn} ist also wie im Falle fester Effekte definiert. Weiter ist

$$\begin{aligned}
QS_A &= n\sum_{ij}(\bar{x}_i - \bar{x})^2 \\
&= n\left(\sum_{ij}(a_i - \bar{a})^2 + \sum_{ij}(\bar{c}_i - \bar{c})^2 + \sum_{ij}(bare_{i.} - bare)^2\right) \qquad (3.45)
\end{aligned}$$

Auf analoge Weise findet man

$$QS_B = n\left(\sum_{ij}(b_j - \bar{b})^2 + \sum_{ij}(\bar{c}_{.j} - \bar{c})^2 + \sum_{ij}(bare_{.j.} - bare)^2\right) \tag{3.46}$$

und

$$QS_{A\times B} = n\sum_{ij}(c_{ij} - \bar{c}_{i.} - \bar{c}_{j.} + bare_{ij.} - bare_{i..} - \bar{e}_{.j.} + bare)^2 \tag{3.47}$$

Nicht nur in $QS_{A\times B}$, sondern auch in QS_A und in QS_B geht die Wechselwirkung ein!. Würde man also wie bei den festen Effekten $H_0(A)$ und $H_0(B)$ in bezug auf QS_{inn} testen, so könnte man im Falle eines signifikanten F-Bruchs nicht sagen, ob dieser zu Lasten signifikanter Haupteffekte oder "nur"zu Lasten existierender Interaktionseffekte geht. Deshalb tested man die Haupteffekte nicht gegen QS_{inn}, sondern gegen $QS_{A\times B}$, denn in diese Quadratsumme geht, wie in QS_A und QS_B, eine mögliche Interaktion ein; gleich, ob letztere vorhanden ist oder nicht, ein F-Bruch

$$F_{I-1,(I-1)(J-1)} = \frac{QS_A/(I-1)}{QS_{A\times B}/(I-1)(J-1)} \tag{3.48}$$

$$F_{J-1,(I-1)(J-1)} = \frac{QS_B/(J-1)}{QS_{A\times B}/(I-1)(J-1)} \tag{3.49}$$

wird nur signifikant, wenn der entsprechende Haupteffekt signifikant wird. Die Wechselwirkung wird, wie im Falle fester Effekte auch, durch

$$F_{(I-1)(J-1),IJ(n-1)} = \frac{QS_{A\times B}/(I-1)(J-1)}{QS_{inn}/IJ(n-1)} \tag{3.50}$$

getestet.

> *Im 2-faktoriellen Design ergeben sich also Unterschiede zwischen einem Modell mit festen und einem mit zufälligen Faktoren. Bis auf die F-Brüche sind die Verfahren im übrigen identisch. Gelegentlich kann es eine Frage der speziellen Hypothese sein, ob man die Stufen eines Faktors als fest oder als zufällig auffaßt. So kann man bei der eingangs erwähnten Frage, ob die Reihenfolge der Stimuluskarten einen Effekt auf die Generierung von Antworten hat, sich auf die speziell gewählten und untersuchten Reihenfolgen beschränken, weil man eben nur etwas über diese Reihenfolgen aussagen möchte; dann wird man das Modell mit festen Effekten wählen. Nur wenn man auf alle möglichen Reihenfolgen verallgemeinern möchte, wird man zum Modell mit zufälligen Effekten übergehen.*

3.3.3 Gemischte Modelle

Oft implizieren die psychologischen Fragestellungen, daß ein Faktor fest, der andere zufällig ist. So kann man die Hypothese aufstellen, daß der Reihenfolgeeffekt beim Rorschach-Test wohl bei weiblichen Testpersonen auftritt, weil sie ganzheitlich denken, nicht aber bei männlichen, denn die denken zergliederend und können das Ganze nicht so gut erfassen. Man hat dann ein 2-faktorielles, gemischtes Design vorliegen: der Faktor "Geschlecht"ist fest, die Reihenfolgen sind zufällig.

Eine ausführliche Diskussion dieses Modelles ist hier nicht möglich (sie wäre relativ zum Rahmen dieses Skripts zu lang). Das Grundmodell kann aber als Strukturgleichung angegeben werden. Die festen Effekte werden dabei mit griechischen, die zufälligen mit lateinischen Buchstaben gekennzeichnet. Es ist dann

$$x_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk} \tag{3.51}$$

Also ist A ein fester Faktor, und B sei zufällig. Die durch c_{ij} angezeigte Interaktion zwischen den beiden Faktoren ist wieder zufällig, da ja die mit der i-ten Stufe von A kombinierte j-te Stufe von B zufällig ist, und damit muß die Wechselwirkung ein zufälliger Effekt sein.

Es wird wieder angenommen, daß die zufälligen Variablen b_j und c_{ij} gemeinsam normalverteilt sind, mit der Varianz σ_B^2 für die b_j und der Varianz $\sigma_{A\times B}^2$ für die Interaktionsterme. Die Fehler e_{ijk} seien unabhängig von den b_j und den c_{ijk}.

Die Zerlegung der Quadratsumme QS_{ges} für die Gesamtvariation geschieht wieder wie bei einer Analyse mit festen Effekten. Die Teilsummen QS_{inn}, QS_A und $QS_{A\times B}$ setzen sich aber anders zusammen, wie aufgrund der Darstellung des Modelles mit zufälligen Effekten zu erwarten ist, was Auswirkungen auf die Konstruktion der F-Tests hat.

Für QS_A ergibt sich

$$\begin{aligned} QS_A &= \sum_{ijk}(\bar{x}_{i..} - \bar{x})^2 \\ &= \sum_{ijk}(\alpha_i + \overline{c}_{i.} - bare + bare_{i.} + bare)^2 \end{aligned} \tag{3.52}$$

Man sieht, daß die Interaktion zwischen dem festen Faktor und dem zufälligen Faktor wieder den Haupteffekt des Faktors A beeinflußt. Für den zufälligen Faktor ergibt sich

$$QS_B = \sum_{ijk}(b_j + bare_{.j} - \overline{b} - bare)^2 \tag{3.53}$$

Hier erscheinen also die Interaktionseffekte *nicht*!

Für die Interaktion ergibt sich

$$QS_{A\times B} = \sum_{ijk}(c_{ij} - \overline{c}_{.j} - bare_{ij.} - bare_{i..} + bare)^2 \tag{3.54}$$

Diese Quadratsumme hängt also nur von den Interaktionstermen und den Fehlertermen ab. Man rechnet leicht nach, daß QS_{inn} frei von Interaktions- und Haupteffekten ist und wie beim Modell mit festen Effekten definiert ist. Damit ergeben sich die folgenden F-Brüche:

$$F_{I-1,(I-1)(J-1)} = \frac{QS_A/(I-1)}{QS_{A\times B}/(I-1)(J-1)} \tag{3.55}$$

$$F_{J-1,IJ(n-1)} = \frac{QS_B/(J-1)}{QS_{inn}/IJ(n-1)} \tag{3.56}$$

$$F_{(I-1)(J-1),IJ(n-1)} = \frac{QS_{A\times B}/(I-1)(J-1)}{QS_{inn}/IJ(n-1)} \tag{3.57}$$

3.4 Meßwiederholungen (Repeated Measurements)

In vielen psychologischen Untersuchungen werden von ein und derselben Gruppe von Personen Messungen unter verschiedenen Bedingungen gewonnen:

- Therapieverlaufsstudien: die Wirkung einer Therapie wird anhand einer Folge von Messungen zu Zeitpunkten t_1, t_2, t_3, $\cdots$ bei den gleichen Personen erhoben,
- Die Wirkung eines Medikaments oder einer Droge (Alkohol) in verschiedenen Dosierungen nauf die Reaktionszeit wird bei den gleichen Personen bestimmt,
- die emotionale Stabilität von Jugendlichen wird jeweils im Abstand von einem halben Jahr untersucht, etc.

Die Vorteile, nur eine Gruppe von Vpn unter den verschiedenen Bedingungen (Zeitpunkte gelten hier als Bedingung) zu beobachten,liegen auf der Hand: zum einen ist es ökonomischer, nur eine Stichprobe erheben zu müssen, zum anderen kann es möglich sein, die Reaktion relativ zum individuellen Standard einer Person zu bestimmen. Die Unterschiede zwischen den Personen gehen sonst in die Fehlervarianz ein und verwischen auf diese Weise die Wirkung der Effekte.

Andererseits wird bei der Varianzanalyse vorausgesetzt, daß die Messungen stochastisch unabhängig sind. Werden die Messungen bei verschiedenen Personen erhoben, ist diese Unabhängigkeit im allgemeinen gegeben. Stammen die Messungen aber stets von den gleichen Personen, so sind die Messungen nicht mehr notwendig auch unabhängig.

Um die die Struktur der Messungen zu sehen, muß man sich zunächst klar machen, daß jede Person eine spezifische Art hat, überhaupt auf die in Frage stehenden Bedingungen zu reagieren. So haben manche Personen im allgemeinen kürzere Reaktionszeiten als andere. Die Wirkung eines Medikaments wirkt zwar auch bei ihnen verlangsamend oder beschleunigend, aber dieser Effekt wirkt gewissermaßen auf die Basisgeschwindigkeit, mit der sie reagieren. Für ein 1-faktorielles Design kann man dann die Strukturgleichung

$$x_{ij} = \mu + \eta_i + \alpha_j + (\eta\alpha)_{ij} + e_{ij}, \quad j = 1, \ldots, J; i = 1, \ldots, n_j \tag{3.58}$$

ansetzen. Hierin ist α_j der Effekt der j-ten Stufe des Faktors – etwa des j-ten Zeitpunkts, an dem die Wirksamkeit einer Therapie untersucht wird, oder der j-ten Dosis des untersuchten Medikaments. η_i ist der "Effekt" der i-ten Person; er repräsentiert etwa eine Grundempfindlichkeit für Alkohol, wenn die Wirkung von Alkohol auf die Reaktionsgeschwindigkeit getestet werden soll, etc. $(\eta\alpha)_{ij}$ ist die Interaktion zwischen Person und Bedingung. So kann bei an Alkohol gewöhnten Personen beobachtet werden, daß sie bei geringen Mengen von Alkohol sogar schneller reagieren als gewöhnlich, bei etwas größer werdenden Mengen aber schneller langsamer werden als weniger an Alkohol gewöhnte Personen.

Der Form nach repräsentiert (3.58) ein 2-faktorielles Design. Der eigentlich interessierende Faktor A ist etwa ein fester Faktor; der zweite Faktor wird aber von der Gruppe der Vpn gebildet. Da diese Gruppe im allgemeinen eine Zufallsstichprobe ist, ist dieser Faktor zufällig. Damit ist der Plan ein gemischter Plan. Aber für jede Bedingung A_j gibt es nur eine Messung, und dies bedeutet, daß die Interaktion $(\eta\alpha)_{ij}$ nicht getestet werden kann! in anderen Worten: die Analyse des Plans (3.58) setzt voraus, daß die Interaktion zwischen Personen und Stufen des Faktors A identisch Null ist. Darüber hinaus wird vorausgesetzt, daß die Fehler $e_{ij}, e_{ij'}$ nicht miteinander korrelieren.

3.4.1 Der allgemeine Ansatz

Diese Annahmen müssen natürlich keineswegs gelten. Ein erster Ansatz besteht darin, den Ansatz (3.58) umzuschreiben. Die Terme $(\eta\alpha)_{ij} + e_{ij}$ können zusammengefaßt werden zu

$$\xi_{ij} = (\eta\alpha)_{ij} + e_{ij}$$

und (3.58) geht über in

$$x_{ij} = \mu + \eta_i + \alpha_j + \xi_{ij} \tag{3.59}$$

Hier sind die ξ_{ij} neue "Fehler"terme, die aber im allgemeinen nicht mehr unkorreliert sind. Im übrigen zeigt der Ansatz, worauf die Analyse tatsächlich zielt, nämlich auf die mögliche Wirksamkeit von A. Um also zu testen, ob es mindestens ein α_i gibt, das von Null verschieden ist, benötigt man Schätzungen für die α_j sowie eine Testgröße. Es sei daran erinnert, daß im "normalen" Design (keine *repeated measurements*, sondern *replicated measurements*, d.h. es gibt zwar wiederholte Messungen, aber an *verschiedenen* Personen) die Fehler e_{ij} und die Effekte α_j, β_k etc. unabhängig voneinander sind. Aber in (3.59) sind die Fehler ξ_{ij} und die Effekte α_j nicht unabhängig voneinander, denn in den ξ_{ij} sind ja die α_j enthalten.

4 Hotelling's T^2

Man kann nun die Differenzen

$$\begin{aligned} y_{ij} &= x_{ij} - x_{i,j+1} \\ &= \mu + \eta_i + \alpha_j - \mu - \eta_i - \alpha_{j+1} + \xi_{ij} - \xi_{i,j+1} \\ &\quad +\alpha_j - \alpha_{j+1} + \xi_{ij} - \xi_{i,j+1} \end{aligned} \tag{4.1}$$

für $j = 1, \cdots, k-1$ bilden. In die y_{ij} gehen nur noch die von den Vpn unabhängigen Größen $\alpha_j - \alpha_{j+1}$ ein sowie die Fehlerdifferenzen $\xi_{ij} - \xi_{i,j+1}$. Weiter ist

$$\bar{y}_j = \frac{1}{m}\sum_{i=1}^{m} y_{ij} = \alpha_j - \alpha_{j+1} + \frac{1}{m}\sum_{i=1}^{m}(\xi_{ij} - \xi_{i,j+1})$$

Dann ist

$$E(\bar{y}_j) = \alpha_j - \alpha_{j+1}$$

unter der Bedingung $E(\xi_{ij} - \xi_{i,j+1}) = 0$, und diese Forderung ist vernünftig; sie entspricht der üblichen Forderung $E(e_{ij}) = 0$. Für gegebene Vp werden die ξ_{ij} allerdings korreliert sein (schon weil die Wechselwirkungen $(\eta\alpha)_{ij}$ eingehen). Die Korrelationen zwischen den ξ_{ij} werden in einer Matrix S zusammengefaßt: $S = (r_{jj'})$. Weiter sei

$$\bar{\mathbf{y}} = (\bar{y}_1, \cdots, \bar{y}_k)'$$

Die Nullhypothese lautet:

$$H_0 : \mu_1 = \cdots = \mu_k$$

Die äquivalente Forderung ist

$$H_0 : \mu_j - \mu_{j+1} = \alpha_j - \alpha_{j+1} = 0, \quad \text{für alle } j = 1, \cdots, k-1 \tag{4.2}$$

Ein Test von H_0 besteht darin, *Hotelling's* T^2 zu berechnen:

$$T^2 = \bar{\mathbf{y}}' S^{-1} \bar{\mathbf{y}} \tag{4.3}$$

Gilt H_0, so ist

$$\frac{m-k+1}{(m-1)(k-1)} T^2 \sim F_{k-1;m-k+1} \tag{4.4}$$

Dies ist der F-Test für eine 1-faktorielle Varianzanalyse mit wiederholten Messungen. Er kann als mehrdimensionale Verallgemeinerung des t-Tests für abhängige Stichproben betrachtet werden.

In komplexeren Versuchsanordnungen wird nicht nur ein Faktor kontrolliert. In diesem Fall kann (3.58) der flexiblere Ansatz sein. Allerdings muß die Matrix S der Kovarianzen geschätzt werden. Wie die Anwendung von Hotelling's T^2 kann diese Aufgabe nur mit dem Rechner gelöst werden. Eine Darstellung des allgemeinen Ansatzes findet man etwa in Christensen (1987).

Literatur

[1] Christensen, R.: Plane Answers to Complex Questions. The Theory of Linear Models. Springer-Verlag, New York, Berlin, Heidelberg etc. 1987

[2] Mann, H.B.: Analysis and design of experiments. Dover Publications, New York 1949

Index