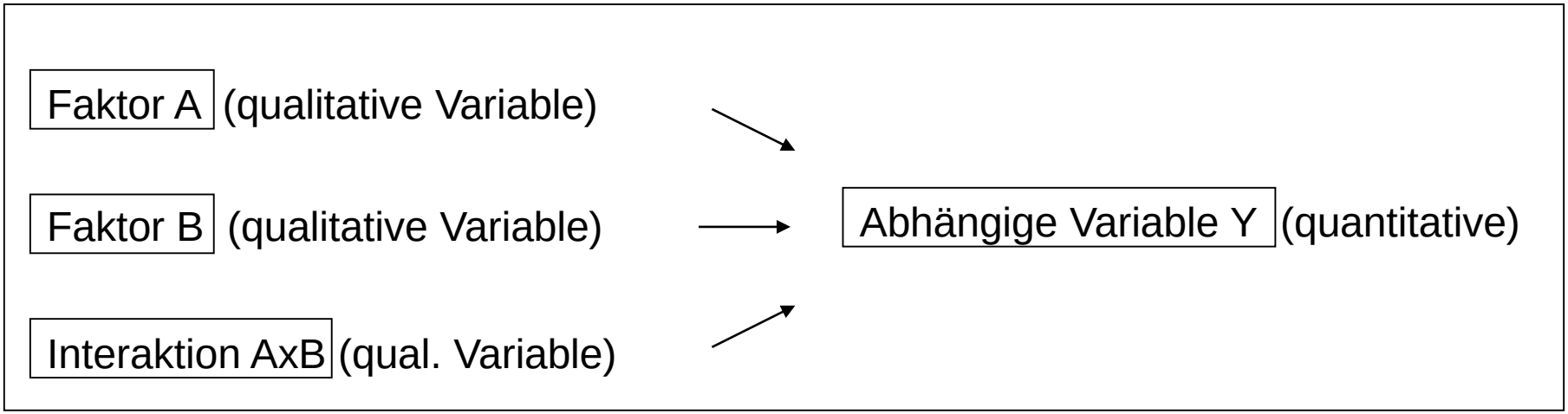
B. Modell der zweifaktoriellen Varianzanalyse mit Wechselwirkungen

Eine **Wechselwirkung (Interaktion) zwischen** den beiden **Faktoren A und B** liegt vor, wenn über die einzelnen Faktoreffekte hinaus bestimmte Kombinationen der Stufen des Faktors A mit dem Faktor B die abhängige Variable Y beeinflussen. So könnte in unserem Beispiel der Produktionsleistung je Schicht ein bestimmtes Produktionsverfahren besonders gut auf eine bestimmte Arbeitsgruppe zugeschnitten sein, so dass in dieser Kombination eine Produktivität erbracht wird, die separat durch die Faktoren A und B nicht erklärt werden kann. Wir kennzeichnen die Interaktionsvariable mit AxB und berücksichtigen ihren Effekt zusätzlich in dem Modell der zweifaktoriellen Varianzanalyse.

Abhängigkeitsstruktur bei der zweifaktoriellen Varianzanalyse mit Wechselwirkung



Das **varianzanalytische Modell mit Wechselwirkung** enthält alle möglichen Einflüsse auf die abhängige Variable Y, die direkt oder indirekt durch die beiden Faktoren A und B hervorgerufen werden können. Es wird daher als gesättigtes (saturiertes) Modell bezeichnet. Bezeichnet man den Interaktionseffekt der Faktoren A und B mit $(\alpha\beta)_{jk}$, dann nimmt das **vollständige Modell (saturiertes Modell) der zweifachen Varianzanalyse** die Form

(3.32) $$y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

an. Im Hinblick auf die Interaktion AxB lautet die Nullhypothese

(3.33) $$H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{12} = ... = (\alpha\beta)_{pq} = 0$$

Wir hatten μ als Gesamtmittelwert der Grundgesamtheit definiert. Die Mittelwerte der Stufen des Faktors A werden mit $\mu_{j\bullet}$ und die Mittelwerte der Stufen des Faktors B mit $\mu_{\bullet k}$ bezeichnet. Konsequent ist dann μ_{jk} der Mittelwert der Faktorkombination A_j und B_k in der Grundgesamtheit. Es zeigt sich dann, dass der Interaktionseffekt $(\alpha\beta)_{jk}$ durch

(3.34) $$(\alpha\beta)_{jk} = \mu_{jk} - \mu_{j\bullet} - \mu_{\bullet k} + \mu$$

gegeben ist. Die Bestimmtheit des Interaktionseffekts in der Form (3.34) durch die Mittelwerte μ_{jk}, $\mu_{j\bullet}$, $\mu_{\bullet k}$ und μ soll jetzt begründet werden. Ihr Verständnis ist von Bedeutung, da sie auch bei der Berechnung der durch die Interaktion AxB erklärten Abweichungsquadratsumme benötigt wird.

Exkurs: Herleitung des Interaktionseffekts (3.34)

Da die Störgröße ε_{ijk} keinen systematischen Einfluss auf die abhängige Variable hat, ist ihr Erwartungswert gleich 0: $E(\varepsilon_{ijk}) = 0$. Bildet man den Erwartungswert von y_{ijk} im varianzanalytischen Modell (3.32), dann erhält man wegen $E(\varepsilon_{ijk})=0$

(E1) $$E(y_{ijk}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

Was gibt aber der Erwartungswert von y_{ijk} an?

- Erwartungswert von Y ohne irgendeinen Faktoreffekt: μ,
- Erwartungswert von Y, wenn nur der Faktor A wirksam ist und der y-Wert zur Stufe A_j dieses Faktors gehört: $\mu_{j\bullet}$,
- Erwartungswert von Y, wenn nur der Faktor B wirksam ist und der y-Wert zur Stufe B_k dieses Faktors gehört: $\mu_{\bullet k}$,
- Erwartungswert von Y bei Wirksamkeit beider Faktoren, wenn der y-Wert zur Stufe A_j des Faktors A und zur Stufe B_k des Faktors B gehört: μ_{jk}.

Der Erwartungswert von y_{ijk}, $E(y_{ijk})$, gibt den Wert an, der bei einer großen Zahl von Realisationen von y-Werten aus der Stufe A_j des Faktors A und der Stufe B_k des Faktors B eintreten wird. Das ist aber genau der Mittelwert μ_{jk} dieser Faktorstufen-kombination: $E(y_{ijk}) = \mu_{jk}$. Gleichung (E1) kann daher gleichwertig in der Form

(E2) $$\mu_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

geschrieben werden. Setzen wir in Gleichung (E2)

$\alpha_j = \mu_{j\bullet} - \mu$ aus Gleichung (3.14)

und

$\beta_k = \mu_{\bullet k} - \mu$ aus Gleichung (3.15)

ein, so erhalten wir

$$\mu_{jk} = \mu + (\mu_{j\bullet} - \mu) + (\mu_{\bullet k} - \mu) + (\alpha\beta)_{jk}$$

Durch Auflösen nach dem Interaktionseffekt $(\alpha\beta)_{jk}$ erhalten wir zunächst die Form

(E3) $(\alpha\beta)_{jk} = \mu_{jk} - \mu - (\mu_{j\bullet} - \mu) - (\mu_{\bullet k} - \mu)$

und schließlich die Darstellung (3.34):

(E4) $(\alpha\beta)_{jk} = \mu_{jk} - \mu_{j\bullet} - \mu_{\bullet k} + \mu$ ♦

Bei den varianzanalytischen Auswertungen werden die theoretischen Mittelwerte (Mittelwerte der Grundgesamtheit) μ_{jk}, $\mu_{j\bullet}$, $\mu_{\bullet k}$ und μ stets durch die empirischen Mittelwerte $\bar{y}_{jk}$, $\bar{y}_{j\bullet}$, $\bar{y}_{\bullet k}$ und $\bar{y}$ ersetzt, die aus den jeweils vorliegenden Daten einer Stichprobe berechnet werden können. Die Stichprobenäquivalente [=berechnete Interaktionseffekte $(\alpha\beta)_{jk}$] zu (E3) und (E4) bzw. (3.34) lauten

(3.35) $AB_{jk} = \bar{y}_{jk} - (\bar{y}_{j\bullet} - \bar{y}) - (\bar{y}_{\bullet k} - \bar{y}) - \bar{y}$

und

(3.36) $AB_{jk} = \bar{y}_{jk} - \bar{y}_{j\bullet} - \bar{y}_{\bullet k} + \bar{y}$

Anschaulich lässt sich der Interaktionseffekt AB_{jk} anhand von Gleichung (3.35) interpretieren: Man rechnet aus einem berechneten Zellenmittelwert den Gesamtmittelwert sowie die separaten Effekte der Faktoren A und B heraus. Was danach noch vom Zellenmittelwert $\bar{y}_{jk}$ verbleibt, lässt sich der Interaktion der beiden Faktoren A und B zuschreiben. Zur Interpretation des Interaktionseffekts AB_{jk} ist es daher zweckmäßig, die Gleichung (3.35) heranzuziehen. Seine Berechung kann aber ohne weiteres unter Verwendung der Formel (3.36) erfolgen.

Während die Mittelwerte der Faktorstufen und der Gesamtmittelwert bereits definiert worden sind, ist hier noch die Berechnung des **Mittelwertes** $\bar{y}_{jk}$ **der Kombination der Faktorstufen A_j und B_k (Zellenmittelwerte)** anzugeben:

$$(3.37)\quad \bar{y}_{jk} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_{ijk}$$

Beispiel 3.9:

Es könnte sein, dass bestimmten Arbeitsgruppen einzelne Produktionsverfahren ganz besonders liegen, so dass sie dadurch eine höhere Produktionsleistung je Schicht erzielen. Um zu prüfen, ob eine solche Wechselwirkung AxB zwischen den Faktoren A und B tatsächlich vorhanden ist, ordnen wird die y-Werte zunächst einmal in einer Kreuztabelle den durch die Faktorkombinationen A_j und B_k entstehende Zellen zu:

Faktor A (Prod.meth.)	Faktor B (Arbeitsgruppe)		
	1	2	3
1	$y_{111} = 69$, $y_{211} = 77$	$y_{112} = 61$, $y_{212} = 67$	$y_{113} = 65$, $y_{213} = 69$
2	$y_{121} = 79$, $y_{221} = 83$	$y_{122} = 62$, $y_{222} = 64$	$y_{123} = 72$, $y_{223} = 78$
3	$y_{131} = 80$, $y_{231} = 78$	$y_{132} = 74$, $y_{232} = 76$	$y_{133} = 80$, $y_{233} = 74$
4	$y_{141} = 75$, $y_{241} = 67$	$y_{142} = 70$, $y_{242} = 62$	$y_{143} = 70$, $y_{243} = 76$

Anschließend bestimmen wir die Zellenmittelwerte:

Faktor A (Prod.meth.)	Faktor B (Arbeitsgruppe)		
	1	2	3
1	$\bar{y}_{11} = \frac{1}{2} \cdot (69 + 77) = 73$	$\bar{y}_{12} = \frac{1}{2} \cdot (61 + 67) = 64$	$\bar{y}_{13} = \frac{1}{2} \cdot (65 + 69) = 67$
2	$\bar{y}_{21} = 81$	$\bar{y}_{22} = 63$	$\bar{y}_{23} = 75$
3	$\bar{y}_{31} = 79$	$\bar{y}_{32} = 75$	$\bar{y}_{33} = 77$
4	$\bar{y}_{41} = 71$	$\bar{y}_{42} = 66$	$\bar{y}_{43} = 73$

Zellenmittelwerte mit SPSS

Mit **SPSS** erhält man die **Zellenmittelwerte** unter dem Menüpunkt

Analysieren
    Mittelwerte vergleichen

in dem man in dem Fenster „Mittelwerte“ das Produktionsergebnis je Schicht dem Feld „Abhängige Variable“ zuweist sowie die Schicht 1 mit dem Faktor A (Produktionsmethode) und Schicht 2 mit dem Faktor B (Arbeitsgruppe) belegt.

Bericht

Produktionsergebnis je Arbeitsschicht

Produktionsmethode	Arbeitsgruppe	Mittelwert	N	Standardabweichung
1	1	73.00	2	5.657
	2	64.00	2	4.243
	3	67.00	2	2.828
	Insgesamt	68.00	6	5.329
2	1	81.00	2	2.828
	2	63.00	2	1.414
	3	75.00	2	4.243
	Insgesamt	73.00	6	8.532
3	1	79.00	2	1.414
	2	75.00	2	1.414
	3	77.00	2	4.243
	Insgesamt	77.00	6	2.757
4	1	71.00	2	5.657
	2	66.00	2	5.657
	3	73.00	2	4.243
	Insgesamt	70.00	6	5.177
Insgesamt	1	76.00	8	5.477
	2	67.00	8	5.782
	3	73.00	8	4.986
	Insgesamt	72.00	24	6.440

Wir können nun die **Interaktionseffekte AB_{jk} berechnen**. Für die Faktorkombination A_1 und B_1 erhält man unter Verwendung von Gleichung (3.35):

$$AB_{11} = \bar{y}_{11} - (\bar{y}_{1\bullet} - \bar{y}) - (\bar{y}_{\bullet 1} - \bar{y}) - \bar{y}$$

$$= 73 - (68 - 72) - (76 - 72) - 72 = 73 - (-4) - 4 - 72 = 1.$$

Ein Plus von 1 erhält man ebenfalls unter Verwendung der Formel (3.36):

$$AB_{11} = \bar{y}_{11} - \bar{y}_{1\bullet} - \bar{y}_{\bullet 1} + \bar{y}$$

$$= 73 - 68 - 76 + 72 = 1.$$

In der Kombination der Faktorstufen A_1 und B_1 erbringen die Arbeiter also eine um 1 höhere Produktionsleistung als nach Korrektur der Durchschnittsleistung um die getrennten Einflüsse der Faktoren A und B zu erwarten ist.

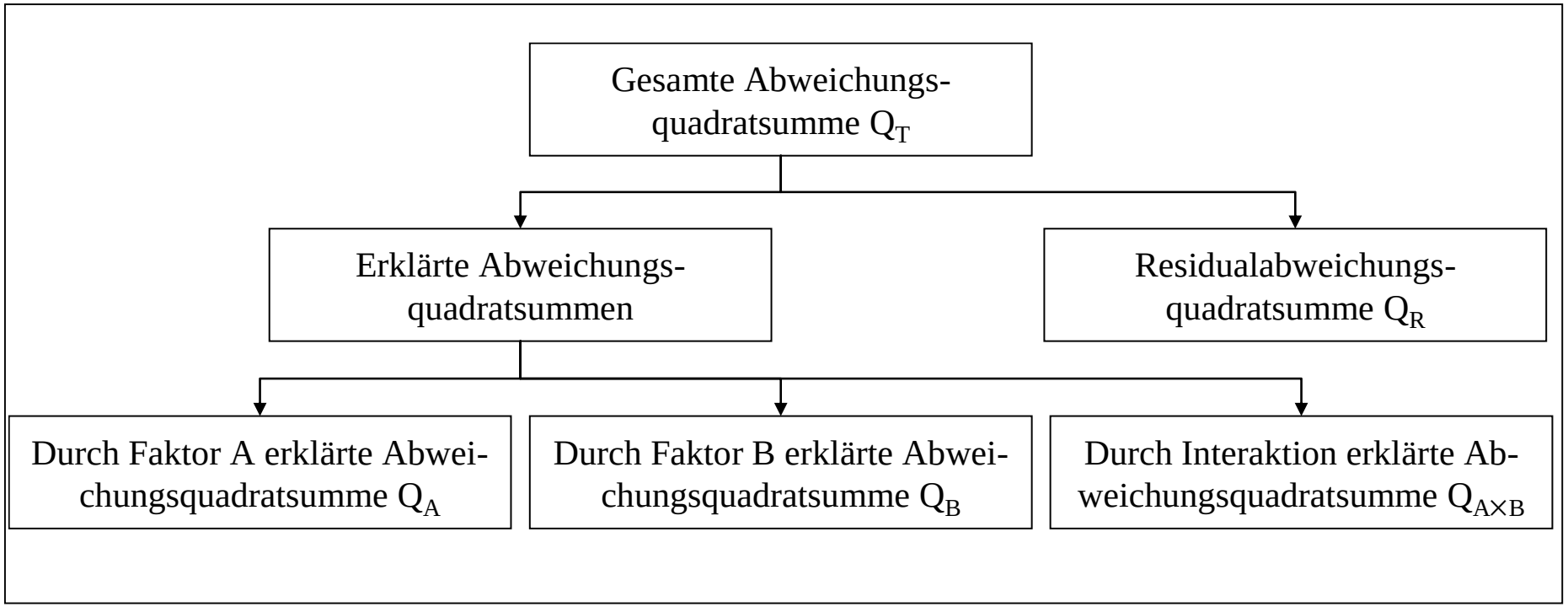
Führt man die Berechnung ebenfalls für alle anderen Kombinationen der Faktorstufen durch, dann erhält man die aus den Stichprobendaten ermittelten Interaktionseffekte AB_{jk}, die in einer Kreuztabelle zusammengestellt werden:

Faktor A (Prod.methode)	Faktor B (Arbeitsgruppe)		
	1	2	3
1	$AB_{11} = 1$	$AB_{12} = 1$	$AB_{13} = -2$
2	$AB_{21} = 4$	$AB_{22} = -5$	$AB_{23} = 1$
3	$AB_{31} = -2$	$AB_{32} = 3$	$AB_{33} = -1$
4	$AB_{41} = -3$	$AB_{42} = 1$	$AB_{43} = 2$

Die aus Stichprobendaten berechneten Interaktionseffekte werden aufgrund des Stichprobenfehlers stets mehr oder weniger stark von 0 abweichen. Die Frage, ob die Interaktionseffekte substanziell oder bloße Zufallseffekte sind, lässt sich in der Varianzanalyse ganz analog zur Beurteilung der getrennten Faktoreffekte behandeln. Wir vergleichen die durch die Interaktion AxB erklärte Streuung unter Berücksichtigung der Freiheitsgrade mit der Reststreuung. Sofern erstere letztere signifikant übersteigt, stufen wir den Interaktionseffekt als substanziell ein. Andernfalls liegt er im Bereich des Stichprobenfehlers und würde dann keinen inhaltlichen Erklärungsbeitrag leisten. ♦

Streuungszerlegung, mittlere Abweichungs quadratsummen und Prüfgröße

Bei der zweifaktoriellen Varianzanalyse mit Wechselwirkungen lässt sich die totale Abweichungsquadratsumme Q_T in vier Komponenten aufspalten:



Aufspaltung der totalen (gesamten) Abweichungsquadratsumme:

$$(3.38)\quad Q_T = \sum_{j=1}^{p}\sum_{k=1}^{q}\sum_{i=1}^{n}(y_{ijk}-\bar{y})^2 = \underbrace{n\cdot q\sum_{j=1}^{p}\left(\bar{y}_{j\bullet}-\bar{y}\right)^2}_{Q_A} + \underbrace{n\cdot p\sum_{k=1}^{q}\left(\bar{y}_{\bullet k}-\bar{y}\right)^2}_{Q_B}$$

$$+\underbrace{n\sum_{j=1}^{p}\sum_{k=1}^{q}\left(\bar{y}_{jk}-\bar{y}_{j\bullet}-\bar{y}_{\bullet k}+\bar{y}\right)^2}_{Q_{A\times B}} + \underbrace{\sum_{j=1}^{p}\sum_{k=1}^{q}\sum_{i=1}^{n}\left(y_{ijk}-\bar{y}_{jk}\right)^2}_{Q_R}$$

Neu hinzugekommen ist hier die auf die Interaktion AxB zurückzuführende Abweichungsquadratsumme

$$(3.39)\quad Q_{A\times B} = n\cdot\sum_{j=1}^{p}\sum_{k=1}^{q}\underbrace{\left(\bar{y}_{jk}-\bar{y}_{j\bullet}-\bar{y}_{\bullet k}+\bar{y}\right)}_{=AB_{jk}}{}^2$$

Die Restabweichungsquadratsumme Q_R ergibt sich jetzt durch Summation der quadrierten Abweichungen der Beobachtungen y_{ijk} von den Zellenmittelwerten $\bar{y}_{jk}$:

$$(3.40)\quad Q_R = \sum_{j=1}^{p}\sum_{k=1}^{q}\sum_{i=1}^{n}\left(y_{ijk}-\bar{y}_{jk}\right)^2$$

Q_{AxB} besitzt (p-1)(q-1) Freiheitsgrade, so dass die mittlere Abweichungsquadrat-summe der Interaktion AxB

$$(3.41)\ MQ_{A\times B} = \frac{Q_{A\times B}}{(p-1(q-1)}$$

lautet. Die beiden **Prüfgrößen** der getrennten Effekte der Faktoren A und B,

$$(3.42)\ F_A = \frac{Q_A/(p-1)}{Q_R/(N-p\cdot q)} = \frac{MQ_A}{MQ_R}$$

und

$$(3.43)\ F_B = \frac{Q_B/(q-1)}{Q_R/(N-p\cdot q)} = \frac{MQ_B}{MQ_R}$$

werden um die Prüfgröße für den Interaktionseffekt ergänzt:

$$(3.44)\ F_{A\times B} = \frac{Q_{A\times B}/[(p-1)(q-1)]}{Q_R/(N-p\cdot q)} = \frac{MQ_{A\times B}}{MQ_R} \quad .$$

Die Quellen der Streuung sind wiederum der besseren Übersicht halber in einer ANOVA-Tabelle zusammengefasst:

Quelle	Abweichungsquadratsumme	Freiheits-grade	Mittlere Abweichungs-quadratsumme
insgesamt	$Q_T = \sum_{j=1}^{p} \sum_{k=1}^{q} \sum_{i=1}^{n} \left(y_{ijk} - \bar{y}\right)^2$	N – 1	$MQ_T = \frac{Q_T}{N-1}$
Faktor A	$Q_A = n \cdot q \sum_{j=1}^{p} \left(\bar{y}_{j\bullet} - \bar{y}\right)^2$	p – 1	$MQ_A = \frac{Q_A}{p-1}$
Faktor B	$Q_B = n \cdot p \sum_{k=1}^{q} \left(\bar{y}_{\bullet k} - \bar{y}\right)^2$	q – 1	$MQ_B = \frac{Q_B}{q-1}$
Interaktion	$Q_{A\times B} = n \sum_{j=1}^{p} \sum_{k=1}^{q} \left(\bar{y}_{jk} - \bar{y}_{j\bullet} - \bar{y}_{\bullet k} + \bar{y}\right)^2$	(p – 1)(q – 1)	$MQ_{A\times B} = \frac{Q_{A\times B}}{(p-1)(q-1)}$
Rest (nicht erklärt)	$Q_R = \sum_{j=1}^{p} \sum_{k=1}^{q} \sum_{i=1}^{n} \left(y_{ijk} - \bar{y}_{jk}\right)^2$	N – p·q	$MQ_R = \frac{Q_R}{N - p \cdot q}$

Beispiel 3.10: In der zweifaktoriellen Varianzanalyse ohne Wechselwirkung hatten wir für die abhängige Variable Produktionsleistung je Schicht eine totale Abweichungsquadratsumme

$$Q_T = 954$$

berechnet, die bei der Hinzunahme der Wechselwirkung AxB unverändert bleibt. Wir können ebenso die Abweichungsquadratsummen der getrennten Faktoreffekte von A und B übernehmen, da sie sich ebenfalls nicht durch die Interaktion AxB verändern:

$$Q_A = 276 \quad \text{und} \quad Q_B = 336.$$

[Verständnisfrage: Warum werden Q_T, Q_A und Q_B durch die Einbeziehung der Interaktion AxB nicht tangiert?]

Zu berechnen bleiben daher noch die Abweichungsquadratsumme, die auf die Wechselwirkung AxB zurückzuführen ist, Q_{AxB}, und die Residualabweichungsquadratsumme Q_R. Für Q_{AxB} erhält man

$$\begin{aligned}
Q_{A\times B} &= n\sum_{j=1}^{4}\sum_{k=1}^{3}\left(\bar{y}_{jk} - \bar{y}_{j\bullet} - \bar{y}_{\bullet k} + \bar{y}\right)^2 \\
&= 2\cdot\left[(73-68-76+72)^2 + (64-68-67+72)^2 + \ldots + (73-70-73+72)^2\right] \\
&= 2\cdot[1^2 + 1^2 + (-2)^2 + 4^2 + (-5)^2 + 1^2 + (-2)^2 + 3^2 + (-1)^2 + (-3)^2 + 1^2 + 2^2] \\
&= 2\cdot 76 = 152
\end{aligned}$$

Die nicht erklärte Abweichungsquadratsumme (Restabweichungsquadratsumme) Q_R erhält man als Differenz zwischen der totalen Abweichungsquadratsumme Q_T und den durch die beiden Faktoren A und B erklärten Abweichungsquadratsummen:

$$Q_R = Q_T - Q_A - Q_B - Q_{A\times B} = 954 - 276 - 336 - 152 = 954 - 764 = 190$$

Eine Kontrollrechnung unter Anwendung der Rechenformel (3.40) zeigt die Richtigkeit dieses Ergebnisses:

i	j (Faktor-stufe A)	k (Faktor-stufe B)	y_{ijk}	$\bar{y}_{jk}$	$y_{ijk} - \bar{y}_{jk}$	$(y_{ijk} - \bar{y}_{jk})^2$
1	1	1	69	73	-4	16
2	1	1	77	73	4	16
1	1	2	61	64	-3	9
2	1	2	67	64	3	9
1	1	3	65	67	-2	4
2	1	3	69	67	2	4
1	2	1	79	81	-2	4
2	2	1	83	81	2	4

1	2	2	62	63	-1	1
2	2	2	64	63	1	1
1	2	3	72	75	-3	9
2	2	3	78	75	3	9
1	3	1	80	79	1	1
2	3	1	78	79	-1	1
1	3	2	74	75	-1	1
2	3	2	76	75	1	1
1	3	3	80	77	3	9
2	3	3	74	77	-3	9
1	4	1	75	71	4	16
2	4	1	67	71	-4	16
1	4	2	70	66	4	16
2	4	2	62	66	-4	16
1	4	3	70	73	-3	9
2	4	3	76	73	3	9
Summe		-	-	-	0	$Q_R = 190$

Die mittleren Abweichungsquadratsummen MQ_A und MQ_B bleiben gegenüber der zweifaktoriellen Varianzanalyse ohne Wechselwirkungen unverändert:

$$MQ_A = 92 \quad \text{und } MQ_B = 168.$$

Als mittlere Abweichungsquadratsumme der Wechselwirkung AxB erhält man

$$MQ_{A\times B} = \frac{Q_{A\times B}}{(p-1)\cdot(q-1)} = \frac{152}{(4-1)\cdot(3-1)} = \frac{152}{6} = 25{,}333$$

und als mittlere Residualabweichungsquadratsumme

$$MQ_R = \frac{Q_R}{N - p\cdot q} = \frac{190}{24 - 4\cdot 3} = \frac{190}{12} = 15{,}833$$

Hiermit ergeben sich die Prüfgrößen

$$F_A = \frac{MQ_A}{MQ_R} = \frac{92}{15{,}833} = 5{,}811$$

$$F_B = \frac{MQ_B}{MQ_R} = \frac{168}{15{,}833} = 10{,}611$$

$$F_{A\times B} = \frac{MQ_{A\times B}}{MQ_R} = \frac{25{,}333}{15{,}833} = 1{,}600$$

der F-Tests.

Wir testen die Signifikanz der getrennten Faktoreffekte und des Interaktionseffekts bei einem Signifkanzniveau (Irrtumswahrscheinlichkeit) α von 5%. Die kritischen Werte hierzu entnehmen wir wiederum einer F-Verteilungstabelle:

	Quantile der F-Verteilung						
		v_1					
F(x)	v_2	1	2	3	4	5	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0,95	12	4,75	3,89	3,49	3,26	3,11	3,00
0,99		9,33	6,93	5,95	5,41	5,06	4,82
0,95	15	4,54	3,68	3,29	3,06	2,90	2,79
0,99		8,68	6,36	5,42	4,89	4,56	4,32
0,95	20	4,35	3,49	3,10	2,87	2,71	2,60
0,99		8,10	5,85	4,94	4,43	4,10	3,87
0,95	25	4,24	3,39	2,99	2,76	2,60	2,49
0,99		7,77	5,57	4,68	4,18	3,86	3,63
0,95	30	4,17	3,32	2,92	2,69	2,53	2,42
0,99		7,56	5,39	4,51	4,02	3,70	3,47
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Kritischer Wert für Faktor A:

$$F_{p-1;N-p\cdot q;1-\alpha} = F_{3;12;0,95} = 3,49$$

Kritischer Wert für Faktor B:

$$F_{q-1;N-p\cdot q;1-\alpha} = F_{2;12;0,95} = 3,89$$

Kritischer Wert für Interaktion AxB:

$$F_{(p-1)(q-1);N-p\cdot q;1-\alpha} = F_{6;12;0,95} = 3,00$$

Damit ergeben sich die folgenden **Testentscheidungen**:

$F_A = 5,811 > F_{3;12;0,95} = 3,49 \quad \Rightarrow$ Nullhypothese (3.19) ablehnen

$F_B = 10,611 > F_{2;12;0,95} = 3,89 \quad \Rightarrow$ Nullhypothese (3.20) ablehnen

$F_{AxB} = 1,600 < F_{6;12;0,95} = 3,00 \quad \Rightarrow$ Nullhypothese (3.33) annehmen

Wie bereits in der zweifaktoriellen Varianzanalyse ohne Wechselwirkung wird hier die Signifikanz der getrennten Effekte der Faktoren A und B bestätigt. Dagegen wird ein substanzieller Einfluss der Interaktion zwischen der Produktionsmethode und der Arbeitsgruppe nicht bestätigt. Die durch die Interaktion AxB erklärte Varianz geht nicht bedeutsam über die Zufallsstreuung hinaus. Es kann daher nicht davon ausgegangen werden, dass sich systematische Produktivitätssteigerung durch ganz bestimmte Kombinationen der Produktionsmethode mit den Arbeitsgruppen erzielen lassen.

Zweifaktorielle Varianzanalyse mit Interaktion mit SPSS

Mit SPSS erhält man eine (erweiterte) ANOVA-Tabelle für die zweifaktorielle Varianzanalyse über die Menüpunkte

Analysieren
  Allgemeines lineares Modell
    Univariat …

Nach Zuweisung der Variablen Produktionsergebnis zu Feld „Abhängige Variable“ und der Variablen Produktionsmethode und Arbeitsgruppe zum Feld „Feste Faktoren“. Das vollständige Modell (gesättigtes Modell) ist voreingestellt, so dass in dem Fenster „Modell“ keine Anpassung vorgenommen werden muss.

Tests der Zwischensubjekteffekte

Abhängige Variable: Produktionsergebnis je Arbeitsschicht

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	764.000[a]	11	69.455	4.387	.009
Konstanter Term	124416.000	1	124416.00	7857.853	.000
Prodmeth	**276.000**	**3**	**92.000**	**5.811**	**.011**
Arbeiter	**336.000**	**2**	**168.000**	**10.611**	**.002**
Prodmeth * Arbeiter	**152.000**	**6**	**25.333**	**1.600**	**.230**
Fehler	**190.000**	**12**	**15.833**		
Gesamt	125370.000	24			
Korrigierte Gesamtvariation	**954.000**	**23**			

a. R-Quadrat = .801 (korrigiertes R-Quadrat = .618)

Die zum Vergleich der F-Werte identische Testentscheidung (α = 0,05) erhalten wir unter Verwendung der in der Spalte „Signifikanz“ ausgewiesenen p-Werte:

für Faktor A: $p = 0{,}011 < \alpha = 0{,}05 \Rightarrow$ Nullhypothese (3.19) ablehnen

für Faktor B: $p = 0{,}002 < \alpha = 0{,}05 \Rightarrow$ Nullhypothese (3.20) ablehnen

für Interaktion AxB: $p=0{,}230 > \alpha=0{,}05 \Rightarrow$ Nullhypothese (3.33) annehmen ♦

3.2.2 Nonorthogonale Varianzanalyse

Bei ungleichen Zellenhäufigkeiten (=Häufigkeiten der Kombinationen der Faktorstufen A_j mit B_k) njk sind die **Faktoren nicht mehr orthogonal** („korrelierte“ Faktoren). Man spricht dann von einem **unbalancierten Design**. Wenn die Faktoreffekte nicht mehr orthogonal sind (miteinander „korrelieren“), dann hat das zur Folge, dass sich die separat berechneten Abweichungsquadratsummen nicht mehr zur totalen Abweichungsquadratsumme Q_T aufaddieren. Bei der **zweifaktoriellen Varianzanalyse mit Wechselwirkung** gilt bei ungleichen Zellenhäufigkeiten n_{jk} also nicht mehr

$$Q_A + Q_B + Q_{AxB} + Q_R = Q_T \qquad \text{(bei orthogonaler Varianzanalyse),}$$

sondern

$$Q_A + Q_B + Q_{AxB} + Q_R \neq Q_T \qquad \text{(bei nonorthogonaler Varianzanalyse).}$$

Der Grund hierfür ist, dass die berechneten Effekte nicht mehr unabhängig voneinander sind. In dem Fall

$$Q_A + Q_B + Q_{AxB} > Q_T - Q_R$$

würde der Gesamteinfluss von A, B und AxB überschätzt werden, während er in dem Fall

$$Q_A + Q_B + Q_{AxB} < Q_T - Q_R$$

zu gering eingeschätzt, d.h. unterschätzt würde. Als Folge hiervon wären die aus der orthogonalen Varianzanalyse unverändert übernommenen Signifikanztests nicht mehr korrekt.

Während bei experimentellen Daten häufig eine „Orthogonalität“, d.h. eine gleiche Zellenbesetzung hergestellt werden kann, ist dies im Allgemeinen bei Beobachtungsdaten nicht möglich. Das Problem ungleicher Zellenbesetzung ist damit als ein für die Wirtschafts- und Sozialwissenschaften praktisch relevanter Fall anzusehen.

Beispiel 3.11: Nonorthogonale Varianzanalyse

Wir ändern unser obiges Beispiel geringfügig ab, indem wir in die Faktorkombination A_1 mit B_2 (Zelle j=1; k=1) die beiden zusätzlichen Beobachtungswerte y_{311}=70 und y_{411}=74 aufnehmen. Damit liegen nun N=26 Beobachtungen vor, die in der folgenden Kreuztabelle zusammengestellt worden sind:

Faktor A (Produkt.-methode)	Faktor B (Arbeitsgruppe)		
	1	2	3
1	$y_{111} = 69$, $y_{211} = 77$, $y_{311} = 70$, $y_{411} = 74$	$y_{112} = 61$, $y_{212} = 67$	$y_{113} = 65$, $y_{213} = 69$
2	$y_{121} = 79$, $y_{221} = 83$	$y_{122} = 62$, $y_{222} = 64$	$y_{123} = 72$, $y_{223} = 78$
3	$y_{131} = 80$, $y_{231} = 78$	$y_{132} = 74$, $y_{232} = 76$	$y_{133} = 80$, $y_{233} = 74$
4	$y_{141} = 75$, $y_{241} = 67$	$y_{142} = 70$, $y_{242} = 62$	$y_{143} = 70$, $y_{243} = 76$

Jetzt sind die Zellen nicht mehr gleich stark besetzt. Wir wollen die Abweichungsquadratsummen bei ungleicher Zellenbesetzung mit den bekannten Formeln ermitteln, wozu wir allerdings Verallgemeinerungen vornehmen müssen:

- Die Anzahl der Beobachtungswerte der Kombination $A_j x B_k$, die bisher für jede Zelle gleich n war, wird durch die Anzahl n_{jk} der y-Werte in der jeweiligen Zelle j;k ersetzt.

- Die Anzahl der y-Werte in jeder Stufe des Faktors A, die bisher in jeder der p Stufen des Faktors A gleich q·n war, wird durch die stufenweise unterschiedliche Anzahl $n_{j\bullet}$ ersetzt.

- Die Anzahl der y-Werte in jeder Stufe des Faktors B, die bisher in jeder der q Stufen des Faktors B gleich p·n war, wird durch die stufenweise unterschiedliche Anzahl $n_{\bullet k}$ ersetzt.

Der Mittelwert der Faktorkombination $A_1 x B_1$ (Zelle j=1; k=1) ergibt sich dann mit n_{11}=4 durch

$$\bar{y}_{11} = \frac{1}{n_{11}} \cdot \sum_{i=1}^{n_{11}} y_{i11} = \frac{1}{4} \cdot (69 + 77 + 70 + 74) = \frac{1}{4} \cdot 290 = 72{,}5$$

Da in allen anderen Zellen dieselben y-Werte wie vorher liegen, bleiben ihre Zellenmittelwerte unverändert.

In der 1. Stufe des Faktors A erhält man mit n_{11}=4 und $n_{1\bullet}$=8 den Mittelwert

$$\bar{y}_{1\bullet} = \frac{1}{n_{1\bullet}} \cdot \sum_{k=1}^{q} \sum_{i=1}^{n_{1k}} y_{i1k} = \frac{1}{8} \cdot (69 + 77 + 70 + 74 + 61 + 67 + 65 + 69) = 69$$

und in der 1. Stufe des Faktors B

$$\bar{y}_{\bullet 1} = \frac{1}{n_{\bullet 1}} \cdot \sum_{j=1}^{p} \sum_{i=1}^{n_{j1}} y_{ij1}$$

$$= \frac{1}{10} \cdot (69 + 77 + 70 + 74 + 79 + 83 + 80 + 78 + 75 + 67) = 75{,}2$$

Alle anderen Mittelwerte der Faktorstufen bleiben unverändert. Der Übersichtlichkeit halber sind die Zellenmittelwerte sowie die Zeilen- und Spaltenmittelwerte in einer Kreuztabelle zusammengestellt.

Faktor A (Produktions methode)	Faktor B (Arbeitsgruppe)			
	1	2	3	$\bar{y}_{j\bullet}$
1	$\bar{y}_{11} = 72{,}5$	$\bar{y}_{12} = 64$	$\bar{y}_{13} = 67$	$\bar{y}_{1\bullet} = 69$
2	$\bar{y}_{21} = 81$	$\bar{y}_{22} = 63$	$\bar{y}_{23} = 75$	$\bar{y}_{2\bullet} = 73$
3	$\bar{y}_{31} = 79$	$\bar{y}_{32} = 75$	$\bar{y}_{33} = 77$	$\bar{y}_{3\bullet} = 77$
4	$\bar{y}_{41} = 71$	$\bar{y}_{42} = 66$	$\bar{y}_{43} = 73$	$\bar{y}_{4\bullet} = 70$
$\bar{y}_{\bullet k}$	$\bar{y}_{\bullet 1} = 75{,}2$	$\bar{y}_{\bullet 2} = 67$	$\bar{y}_{\bullet 3} = 73$	$\bar{y} = 72$

Für den Gesamtmittelwert $\bar{y}$ erhalten wir mit p=4 und q=3 den Wert

$$\bar{y} = \frac{1}{N}\sum_{j=1}^{p}\sum_{k=1}^{q}\sum_{i=1}^{n_{jk}} y_{ijk} = \frac{1}{26}\cdot(69+77+\ldots+70+76) = 72,$$

der sich durch die Aufnahme der beiden zusätzlichen y-Werte in die Zelle j=1; k=1 nicht verändert hat.

Berechnung der Abweichungsquadratsummen

Totale Abweichungsquadratsumme:

$$Q_T = \sum_{j=1}^{p}\sum_{k=1}^{q}\sum_{i=1}^{n_{jk}}(y_{ijk}-\bar{y})^2 = (69-72)^2+(77-72)^2+\ldots+(70-72)^2+(76-72)^2 = 962$$

Abweichungsquadratsumme des Faktors A:

$$Q_A = \sum_{j=1}^{p} n_{j\bullet}(\bar{y}_{j\bullet}-\bar{y})^2$$

$$= 8\cdot(69-72)^2+6\cdot(73-72)^2+6\cdot(77-72)^2+6\cdot(70-72)^2 = 72+6+150+24 = 252$$

Abweichungsquadratsumme des Faktors B:

$$Q_B = \sum_{k=1}^{q} n_{\bullet k}\cdot(\bar{y}_{\bullet k}-\bar{y})^2$$

$$= 10\cdot(75{,}2-72)^2+8\cdot(67-72)^2+8\cdot(73-72)^2 = 102{,}4+200+8 = 310{,}4$$

Fakt. A	Fakt. B				Erwartete Mittelwerte ohne Interaktion	Interaktions-effekt	Beitrag zur Wechselwirkung
j	k	$\bar{y}_{jk}$	$\bar{y}_{j\bullet}$	$\bar{y}_{\bullet k}$	$\bar{y}_{jk}^{erw} = (\bar{y}_{j\bullet} + \bar{y}_{\bullet k} - \bar{y})$	$AB_{jk} = \bar{y}_{jk} - \bar{y}_{jk}^{erw}$	$n_{jk}(\bar{y}_{jk} - \bar{y}_{jk}^{erw})^2$
1	1	72,5	69	75,2	72,2	0,3	0,36
1	2	64	69	67	64	0	0
1	3	67	69	73	70	-3	18
2	1	81	73	75,2	76,2	4,8	46,08
2	2	63	73	67	68	-5	50
2	3	75	73	73	74	1	2
3	1	79	77	75,2	80,2	-1,2	2,88
3	2	75	77	67	72	3	18
3	3	77	77	73	78	-1	2
4	1	71	70	75,2	73,2	-2,2	9,68
4	2	66	70	67	65	1	2
4	3	73	70	73	71	2	8
Summe							$Q_{A\times B} = 159$

Erwarteter Mittelwert ohne Interaktion: $\bar{y}_{jk}^{erw} = \bar{y} + (\bar{y}_{j\bullet} - \bar{y}) + (\bar{y}_{\bullet k} - \bar{y}) = \bar{y}_{j\bullet} + \bar{y}_{\bullet k} - \bar{y}$

z.B. Faktorkombination A_1B_1: $n_{11} = 4$

$$\bar{y}_{11}^{erw} = \bar{y} + (\bar{y}_{1\bullet} - \bar{y}) + (\bar{y}_{\bullet 1} - \bar{y}) = 72 + (69 - 72) + (75{,}2 - 72) = 72 + (-3) + 3{,}2 = 72{,}2$$

$$AB_{11} = \bar{y}_{11} - \bar{y}_{11}^{erw} = 72{,}5 - 72{,}2 = 0{,}3$$

Beitrag zur Wechselwirkung: $n_{11} \cdot (\bar{y}_{11} - \bar{y}_{11}^{erw})^2 = 4 \cdot 0{,}3^2 = 0{,}36$

Die Restabweichungsquadratsumme berechnen wir ebenfalls in einer Arbeitstabelle:

i	j (Fakt. A)	k (Fakt. B)	y_{ijk}	$\bar{y}_{jk}$	$y_{ijk} - \bar{y}_{jk}$	$(y_{ijk} - \bar{y}_{jk})^2$
1	1	1	69	72,5	-3,5	12,25
2	1	1	77	72,5	4,5	20,25
3	1	1	70	72,5	-2,5	6,25
4	1	1	74	72,5	1,5	2,25
1	1	2	61	64	-3	9
2	1	2	67	64	3	9
1	1	3	65	67	-2	4
2	1	3	69	67	2	4
1	2	1	79	81	-2	4

2	2	1	83	81	2	4
1	2	2	62	63	-1	1
2	2	2	64	63	1	1
1	2	3	72	75	-3	9
2	2	3	78	75	3	9
1	3	1	80	79	1	1
2	3	1	78	79	-1	1
1	3	2	74	75	-1	1
2	3	2	76	75	1	1
1	3	3	80	77	3	9
2	3	3	74	77	-3	9
1	4	1	75	71	4	16
2	4	1	67	71	-4	16
1	4	2	70	66	4	16
2	4	2	62	66	-4	16
1	4	3	70	73	-3	9
2	4	3	76	73	3	9
Summe		-	-	-	0	$Q_R = 199$

Während die totale Abweichungsquadratsumme Q_T einen Wert von 962 annimmt, summieren sich die übrigen Abweichungsquadrate zu 920,4:

$$(Q_T = 962) \neq$$
$$(Q_A = 252) + (Q_B = 310{,}4) + (Q_{A\times B} = 159) + (Q_R = 199) = 920{,}4$$

Wegen

$$(Q_A = 252) + (Q_B = 310{,}4) + (Q_{A\times B} = 159) = 721{,}4$$
$$< (Q_T = 962) - (Q_R = 199) = 763$$

wird der Gesamteinfluss der Faktoren A und B zusammen mit der Interaktion AxB unterschätzt.

Der Grund hierfür liegt darin, dass bei ungleich besetzten Zellen offenbar die „gemischten“ Abweichungsquadratsummen (siehe hierzu Exkurs: Streuungszerlegung bei der einfaktoriellen Varianzanalyse) nicht gleich 0 sind und wie im orthogonalen Fall vernachlässigt werden können. Diese „gemischten“ Abweichungsquadratsummen stellen aber ebenfalls Effekte der beiden Faktoren A und B sowie ihrer Interaktion dar. Aus diesem Grund wird tatsächlich nicht die geringere Abweichungsquadratsumme von

$$(Q_A = 252) + (Q_B = 310{,}4) + (Q_{A\times B} = 159) = 721{,}4$$

sondern eine Abweichungsquadratsumme von

$$(Q_T = 962) - (Q_R = 199) = 763$$

durch die A, B und AxB erklärt. Die Differenz von 41,6 (=763 – 721,4) ist daher insgesamt dem systematischen Einfluss der Faktoren A und B sowie der Interaktion AxB zuzuordnen. Dass die Differenz von 41,6 eine systematische und keine Residualkomponente (Restkomponente) ist, steht also fest. Als schwierig erweist sich jedoch die Aufteilung der Differenz

$$[\,(Q_T = 962) - (Q_R = 199) = 763\,] -$$
$$[\,(Q_A = 252) + (Q_B = 310{,}4) + (Q_{A\times B} = 159) = 721{,}4\,] = 41{,}6$$

auf die Einzeleinflüsse A, B und AxB. ♦

Korrekturmethoden der nonorthogonalen Varianzanalyse

Um die Gleichheit der Summe der Abweichungsquadrate von A, B und AxB mit der Differenz Q_T-Q_R herzustellen, gibt es in der **nonorthogonalen Varianzanalyse** unterschiedliche **Korrekturmethoden**:

- Hierarchische Methode

Die Haupteffekte, also die Faktoren A und B, werden sukzessive berechnet. Der Haupteffekt A enthält den mit B gemeinsamen Effekt, d.h. die Faktorstufen A_1, A_2, …, A_p werden jeweils zusätzlich nach den Stufen B_1, B_2, …, B_q des Faktors B untergliedert.

- Experimentelle Methode

Die Haupteffekte werden gegenseitig korrigiert. Bei der Berechnung eines Haupteffekts werden vor- und nachstehende Faktoren berücksichtigt.

- Regressionsmethode

Alle Effekte werden simultan auf ihren Einfluss untersucht. Entsprechend wird eine simultane Korrektur gemeinsamer Einflüsse vorgenommen.

Die Korrekturmethoden der nonorthogonalen Varianzanalyse sind äußerst aufwendig, so dass sie praktisch nur mit einem Statistikprogramm effizient durchgeführt werden können. SPSS enthält die dazu erforderlichen Optionen in Form einer Wahl von 4 Typen. Wir bevorzugen in der Anwendung den für Beobachtungsdaten vorteilhaften Typ III, der die „Regressionsmethode“ beinhaltet und die Standardeinstellung in SPSS ist.

Korrekturmethoden in SPSS:

Typ I: Hierarchischer Ansatz

Typ II: Experimenteller Ansatz

Typ III: Regressionsmethode (Voreinstellung)

Beispiel 3.12: Für den um zwei Beobachtungswerte der abhängigen Variablen Produktionsergebnis je Schicht erweiterten Datensatz soll der nonorthogonale Fall der zweidimensionalen Varianzanalyse mit SPSS illustriert werden. Die Wahl des Verfahrens verläuft wie im Fall der orthogonalen Varianzanalyse. Wir verwenden hierbei zunächst die **Regressionsmethode** (Voreinstellung: Typ III). In der von SPSS ausgegebenen (erweiterten) ANOVA-Tabelle sind die interessierenden Quellen wieder durch Fettdruck und eine größere Schrift hervorgehoben.

Tests der Zwischensubjekteffekte

Abhängige Variable: Produktionsergebnis je Arbeitsschicht

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	**763.000[a]**	**11**	**69.364**	**4.880**	**.003**
Konstanter Term	129675.174	1	129675.17	9122.877	.000
Prodmeth	**303.571**	**3**	**101.190**	**7.119**	**.004**
Arbeiter	**344.818**	**2**	**172.409**	**12.129**	**.001**
Prodmeth * Arbeiter	**151.000**	**6**	**25.167**	**1.771**	**.177**
Fehler	**199.000**	**14**	**14.214**		
Gesamt	135746.000	26			
Korrigierte Gesamtvariation	**962.000**	**25**			

a. R-Quadrat = .793 (korrigiertes R-Quadrat = .631)

Hier wird die Quelle „Korrigiertes Modell“ verständlich, da sie die korrigierte Abweichungsquadratsumme der Effekte A, B und AxB von 763 wiedergibt, die genau der Differenz Q_T - Q_R = 962 – 199 = 763 entspricht. Die Quadratsummen der Effekte A, B und AxB summieren sich dagegen nicht zu 763: $Q_A + Q_B + Q_{AxB}$ = 303,571 + 344,818 +151 = 799,389.

Die ANOVA-Tabelle der **experimentellen Methode** (Typ II) hat folgende Gestalt:

Tests der Zwischensubjekteffekte

Abhängige Variable: Produktionsergebnis je Arbeitsschicht

Quelle	Quadratsumme vom Typ II	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	**763.000[a]**	**11**	**69.364**	**4.880**	**.003**
Konstanter Term	134784.000	1	134784.00	9482.291	.000
Prodmeth	**301.600**	**3**	**100.533**	**7.073**	**.004**
Arbeiter	**360.000**	**2**	**180.000**	**12.663**	**.001**
Prodmeth * Arbeiter	**151.000**	**6**	**25.167**	**1.771**	**.177**
Fehler	**199.000**	**14**	**14.214**		
Gesamt	135746.000	26			
Korrigierte Gesamtvariation	**962.000**	**25**			

a. R-Quadrat = .793 (korrigiertes R-Quadrat = .631)

Auch mit der **experimentellen Methode** stimmt die korrigierte Abweichungsquadratsumme der Effekte A, B und AxB (Quelle: „Korrigiertes Modell“) mit der Differenz Q_T - Q_R = 962 – 199 = 763 überein. Die Quadratsummen der Effekte A, B und AxB summieren sich auch hier nicht zu 763: $Q_A + Q_B + Q_{AxB}$ = 301,6 + 360 +151 = 812,6.

Ergänzend geben wir noch die ANOVA-Tabelle der **hierarchischen Methode** an:

Tests der Zwischensubjekteffekte

Abhängige Variable: Produktionsergebnis je Arbeitsschicht

Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	**763.000[a]**	**11**	**69.364**	**4.880**	**.003**
Konstanter Term	134784.000	1	134784.00	9482.291	.000
Prodmeth	**252.000**	**3**	**84.000**	**5.910**	**.008**
Arbeiter	**360.000**	**2**	**180.000**	**12.663**	**.001**
Prodmeth * Arbeiter	**151.000**	**6**	**25.167**	**1.771**	**.177**
Fehler	**199.000**	**14**	**14.214**		
Gesamt	135746.000	26			
Korrigierte Gesamtvariation	**962.000**	**25**			

a. R-Quadrat = .793 (korrigiertes R-Quadrat = .631)

Allein mit der hierarchischen Methode wird eine Übereinstimmung der Abweichungsquadratsummen der Effekte A, B und AxB mit der Differenz $Q_T – Q_R = 962 – 199 = 763$ hergestellt: $Q_A + Q_B + Q_{AxB} = 252 + 360 +151 = 763$. Allerdings dürfte die Annahme, dass ein Faktor A in hierarchischer Beziehung zu einem anderen Faktor B steht, in Anwendungen mit Beobachtungsdaten eher die Ausnahme als die Regel sein.

♦