Prof. Dr. Reinhold Kosfeld
Fachbereich Wirtschaftswissenschaften
Universität Kassel

Lösungen zu den Aufgaben zur Multivariaten Statistik
Teil 3: Aufgaben zur Varianzanalyse

1. Worauf bezieht sich der Unterschied zwischen einer einfaktoriellen und mehrfaktoriellen Varianzanalyse?

   Lösung:

   Während bei der einfaktoriellen Varianzanalyse der Einfluss einer nominalskalierten unabhängigen Variablen (=Faktor) auf eine metrisch skalierte abhängige Variable untersucht wird, treten in der mehrfaktoriellen Varianzanalyse zusätzliche Faktoren als Einflussgrößen hinzu. Hierbei können neben den Haupteffekten der Faktoren ebenfalls die Wechselwirkungen (Interaktionen) zwischen den Faktoren relevant sein, die in der einfaktoriellen Varianzanalyse nicht auftreten können.

2. Zur Eingliederung von Langzeitarbeitslosen in den Arbeitsmarkt sind zwei Methoden „Fitness for the Job“ eingesetzt worden. Eine Evaluation der Trainingsmethoden nach einem Jahr hat Aufschluss über die Verweildauern in der Arbeitslosigkeit (in Mon.) erbracht:

Trainingsmethode 1	10, 8, 6, 12
Trainingsmethode 2	4, 3, 6, 5, 8, 6

   Ist eine unterschiedliche Wirksamkeit der beiden Trainingsmethoden auf einem Signifikanzniveau von 5% statistisch gesichert?

   Lösung:
   Abh. Variable: Verweildauer in der Arbeitslosigkeit

   Faktor A: Trainingsmethode | H_0: $\alpha_1 = \alpha_2 = 0$

   Faktorstufe: Methode 1 Methode 2

   Streuungszerlegung:

$$\underbrace{\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}\right)^2}_{=Q} = \underbrace{\sum_{j=1}^{p} n_j\left(\bar{y}_j-\bar{y}\right)^2}_{=Q_A} + \underbrace{\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}_j\right)^2}_{=Q_R}$$

   Mittelwerte der Faktorstufen:

   A1: $\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1=4} y_{i1} = \frac{1}{4}(10+8+6+12) = \frac{1}{4}\cdot 36 = 9$

   A2: $\bar{y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2=6} y_{i2} = \frac{1}{6}(4+3+6+5+8+6) = \frac{1}{6}\cdot 32 = 5{,}33$

   Gesamtmittelwert ($N = n_1 + n_2 = 4 + 6 = 10$):

$$\bar{y} = \frac{1}{N}\sum_{j=1}^{p=2}\sum_{i=1}^{N=10} y_{ij} = \frac{1}{10}(10+8+6+12+4+3+6+5+8+6) = \frac{1}{10}\cdot 68 = 6{,}8$$

Abweichungsquadratsummen:

$$Q_A = \sum_{j=1}^{2} n_j(\bar{y}_j - \bar{y})^2 = 4\cdot\underbrace{(9-6{,}8)^2}_{=2{,}2^2} + 6\cdot\underbrace{(5{,}33-6{,}8)^2}_{=(-1{,}47)^2} = 19{,}36 + 12{,}97 = 32{,}33$$

$$Q_R = \sum_{j=1}^{p=2}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$$

$$= (10-9)^2 + (8-9)^2 + (6-9)^2 + (12-9)^2$$

$$+ (4-5{,}33)^2 + (3-5{,}33)^2 + (6-5{,}33)^2 + (5-5{,}33)^2 + (8-5{,}33)^2 + (8-5{,}33)^2 + (6-5{,}33)^2$$

$$= 1^2 + (-1)^2 + (-3)^2 + 3^2$$

$$+ (-1{,}33)^2 + (-2{,}33)^2 + 0{,}67^2 + (-0{,}33)^2 + 2{,}67^2 + 0{,}67^2$$

$$= 1 + 1 + 9 + 9 + 1{,}7689 + 5{,}4289 + 0{,}4489 + 0{,}1089 + 7{,}1289 + 0{,}4489$$

$$= 35{,}3334$$

Mittlere Abweichungsquadratsummen:

$$MQ_A = \frac{Q_A}{p-1} = \frac{32{,}33}{2-1} = 32{,}33$$

$$MQ_R = \frac{Q_R}{(n_1+n_2)-2} = \frac{35{,}3334}{10-2} = 4{,}4167$$

Prüfgröße:

$$F_A = \frac{MQ_A}{MQ_R} = \frac{32{,}33}{4{,}4167} = 7{,}3199$$

Kritischer Wert ($\alpha = 0{,}05$):

$$F_{p-1;N-p;0{,}95} = F_{1;8;0{,}95} = 5{,}32$$

Testentscheidung:

$F_A = 7{,}3199 > F_{1;8;0{,}95} = 5{,}32$ => H_0 ablehnen (unterschiedliche Wirksamkeit der beiden Trainingsmethoden ist gegeben)

3. Begründen Sie die Anzahl der Freiheitsgrade der Abweichungsquadratsummen in der einfaktoriellen Varianzanalyse!

<u>Lösung:</u>

<u>Totale Abweichungsquadratsumme (Q_T)</u>:

Zur der Ermittlung der totalen Abweichungsquadratsumme muss aus den N Beobachtungswerten vorab das Gesamtmittel $\bar{y}$ berechnet werden. Hat man die ersten N-1 Abweichungen y_{ij}-$\bar{y}$ ermittelt, dann steht auch die letzte Abweichung y_{Np}-$\bar{y}$ fest. Man sagt, dass durch die Berechnung des Gesamtmittels $\bar{y}$ aus den Stichprobendaten 1 Freiheitsgrad verloren geht. Q_T besitzt daher N-1 Freiheitsgrade.

Erklärte Abweichungsquadratsumme durch den Faktor A (Q_A):

Das Gesamtmittel $\bar{y}$ könnte genauso gut aus den Mittelwerten $\bar{y}_j$ der p Stufen des Faktors A berechnet werden. Hat man die ersten p-1 Abweichungen $\bar{y}_j - \bar{y}$ bestimmt, liegt wiederum die letzte Abweichung $\bar{y}_p - \bar{y}$ fest, so dass Q_A p-1 Freiheitsgrade besitzt.

Residualabweichungsquadratsumme (Q_R):

Zur Berechnung der Residualabweichungsquadsratsumme Q_R wird auf p Mittelwerte der Stufen des Faktors A, $\bar{y}_p$, zurückgegriffen. Mit der Kenntnis der ersten N-p Abweichungen $y_{ij} - \bar{y}_p$, liegen auch die letzten p Abweichungen fest. Q_R hat damit N-p Freiheitsgrade. (Hinweis: Die Anzahl der Freiheitsgrade von Q_R lässt sich auch residual bestimmen.)

4. Erläutern Sie die unterschiedlichen Effekte im vollständigen Modell einer zweifachen Varianzanalyse!

Lösung:

Vollständiges Modell der zweifachen Varianzanalyse:

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

α_j gibt den Haupteffekt des Faktors A auf die abhängige Variable y wieder, β_k den Haupteffekt des Faktors B. $(\alpha\beta)_{jk}$ ist der Effekt, der aus einer Wechselwirkung der beiden Faktoren A und B resultiert (Interaktionseffekt). ε_{ijk} gibt schließlich den Zufallseinfluss auf y wieder, der durch die aufgrund ihrer geringen Bedeutung nicht berücksichtigten Einflussgrößen („Zufallsfaktoren“) zustande kommt.

5. Was versteht man unter einem balanciertem Design? Welches Schätzproblem tritt bei einem unbalancierten Designs auf?

Lösung:

Von einem balancierten Design (orthogonales Design) spricht man, wenn die sich aus den Kombinationen der Faktorstufen ergebenden Zellen alle gleich besetzt sind.

Bei einem unbalancierten Design (nonorthogonalen Design) lässt sich die Gesamtsumme der Abweichungsquadrate nicht mehr exakt in die auf die Faktoren zurückführbaren Anteile zerlegen. Vielmehr übersteigt oder unterschreitet die Summe der Einzelabweichungsquadrate die Gesamtsumme. Der Grund hierfür ist der Verlust der Orthogonalität des Designs, der zu Abhängigkeiten zwischen den Faktoren führt mit der Folge, dass die Kreuzproduktsummen nicht mehr verschwinden. Für die nonorthogonale Varianzanalyse sind verschiedene Verfahren wie z.B. die Regressionsmethode oder der experimentelle Ansatz entwickelt worden, die auf eine Separierung der Einzeleinflüsse abstellen.

6. Überprüfen Sie die beiden Haupteffekte einer Varianzanalyse der Umsätze (=abhängige Variable) auf Signifikanz! (α=0,05)

		Preisstrategie 1	Preisstrategie 2
Werbe-Strategie	1	20, 18, 22	16, 14, 18
	2	24, 19, 23	17, 16, 18

<u>Lösung:</u>

Abhängige Variable: Umsatz (metrisch skaliert)

Faktor A:	Werbestrategie		Faktor B:	Preisstrategie	
Faktorstufen:	A1	A2	Faktorstufen:	B1	B2

$H_0 : \alpha_1 = \alpha_2 = 0$ und $\beta_1 = \beta_2 = 0$

Streuungszerlegung:

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

Mittelwerte der Stufen des Faktors A:

$$\bar{y}_{1\bullet} = \frac{1}{\underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{q}}\sum_{k=1}^{q=2}\sum_{i=1}^{n=3} y_{i1k} = \frac{1}{6}\left(20+18+22+16+14+18\right) = \frac{108}{6} = 18$$

$$\bar{y}_{2\bullet} = \frac{1^2}{\underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{q}}\sum_{k=1}^{q=2}\sum_{i=1}^{n=3} y_{i2k} = \frac{1}{6}\left(24+19+23+17+16+18\right) = \frac{117}{6} = 19{,}5$$

Mittelwerte der Stufen des Faktors B:

$$\bar{y}_{\bullet 1} = \frac{1}{\underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{p}}\sum_{j=1}^{p=2}\sum_{i=1}^{n=3} y_{ij1} = \frac{1}{6}\left(20+18+22+24+19+23\right) = \frac{126}{6} = 21$$

$$\bar{y}_{\bullet 2} = \frac{1}{\underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{p}}\sum_{j=1}^{p=2}\sum_{i=1}^{n=3} y_{ij2} = \frac{1}{6}\left(16+14+18+17+16+18\right) = \frac{99}{6} = 16{,}5$$

Gesamtmittelwert:

$$\bar{y} = \frac{1}{\underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{p}\cdot\underset{\underset{2}{\uparrow}}{q}} \sum_{j=1}^{p=2}\sum_{k=1}^{q=2}\sum_{i=1}^{n=3} = \frac{1}{12}\left(20+18+22+16+14+18+24+19+23+17+16+18\right) = \frac{225}{12} = 18{,}75$$

Abweichungsquadratsummen:

$$Q_A = \underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{q}\cdot\sum_{j=1}^{p=2}\left(\bar{y}_{j\bullet}-\bar{y}\right)^2 = 6\cdot\left[(18-18{,}75)^2+(19{,}5-18{,}75)^2\right]$$

$$= 6\cdot\left[(-0{,}75)^2+0{,}75^2\right] = 6{,}75$$

$$Q_B = \underset{\underset{3}{\uparrow}}{n}\cdot\underset{\underset{2}{\uparrow}}{p}\cdot\sum_{k=1}^{q=2}\left(\bar{y}_{\bullet k}-\bar{y}\right)^2 = 6\cdot\left[(21-18{,}75)^2+\left(16{,}5-18{,}75)^2\right)\right]$$

$$= 6\cdot\left[2{,}25^2+(-2{,}25)^2\right] = 60{,}75$$

$$Q_R = \sum_{j=1}^{p=2}\sum_{k=1}^{q=2}\sum_{i=1}^{n_{jk}=3}\left(y_{ijk}-\bar{y}_{j\bullet}-\bar{y}_{\bullet k}+\bar{y}\right)^2$$

Komb. A1B1 $= (20 - 18 - 21 + 18{,}75)^2 + (18 - 18 - 21 + 18{,}75)^2 + (22 - 18 - 21 + 18{,}75)^2$

Komb. A1B2 $+ (16 - 18 - 16{,}5 + 18{,}75)^2 + (14 - 18 - 16{,}5 + 18{,}75)^2 + (18 - 18 - 16{,}5 + 18{,}75)^2$

Komb. A2B1 $+ (24 - 19{,}5 - 21 + 18{,}75)^2 + (19 - 19{,}5 - 21 + 18{,}75)^2 + (23 - 19{,}5 - 21 + 18{,}75)^2$

Komb. A2B2 $+ (17 - 19{,}5 - 16{,}5 + 18{,}75)^2 + (16 - 19{,}5 - 16{,}5 + 18{,}75)^2 + (18 - 19{,}5 - 16{,}5 + 18{,}75)^2$

$= (-0{,}25)^2 + (-2{,}25)^2 + 1{,}75^2 + 0{,}25^2 + (-1{,}75)^2 + 2{,}25^2 + 2{,}25^2 + (-2{,}75)^2 + 1{,}25^2 + (-0{,}25)^2 + (-1{,}25)^2 + 0{,}75^2$

$= 0{,}0625 + 5{,}0625 + 3{,}0625 + 0{,}0625 + 3{,}0625 + 5{,}0625 + 5{,}0625 + 7{,}5625 + 1{,}5625 + 0{,}0625 + 1{,}5625 + 0{,}5625$

$= 32{,}75$

Mittlere Abweichungsquadratsummen:

$$MQ_A = \frac{Q_A}{p-1} = \frac{6{,}75}{2-1} = 6{,}75$$

$$MQ_B = \frac{Q_B}{q-1} = \frac{60{,}75}{2-1} = 60{,}75$$

$$MQ_R = \frac{Q_R}{n\cdot p\cdot q-(p+q-1)} = \frac{32{,}75}{3\cdot 2\cdot 2-(2+2-1)} = \frac{32{,}75}{9} = 3{,}6389$$

Prüfgrößen:

$$F_A = \frac{MQ_A}{MQ_R} = \frac{6{,}75}{3{,}6389} = 1{,}855$$

$$F_B = \frac{MQ_B}{MQ_R} = \frac{60{,}75}{3{,}6389} = 16{,}695$$

Kritischer Wert ($\alpha = 0{,}05$)

$F_{p-1;n\cdot p\cdot q-(p+q-1);0{,}95} = F_{1;9;0{,}95} = 5{,}12$

Testentscheidungen:

$F_A = 1{,}855 < F_{1;9;0{,}95} = 5{,}12$ => Faktor A nicht signifikant (Werbestrategie)

$F_B = 16{,}695 > F_{1;9;0{,}95} = 5{,}12$ => Faktor B signifikant (Preisstrategie)

7. Wie ist der Interaktionseffekt in Aufg. 6 zu bewerten? (α=0,05)

<u>Lösung</u>:

Test des Interaktionseffekts in Aufg. 6:

Zellenmittelwerte ($\hat{=}$ Mittelwerte der Kombinationen der Faktorstufen)

$$\bar{y}_{11} = \frac{1}{\underset{\substack{\uparrow\\3}}{n}}(20+18+22) = \frac{60}{3} = 20, \quad \bar{y}_{12} = \frac{1}{\underset{\substack{\uparrow\\3}}{n}}(16+14+18) = \frac{48}{3} = 16,$$

$$\bar{y}_{21} = \frac{1}{\underset{\substack{\uparrow\\3}}{n}}(24+19+23) = \frac{66}{3} = 22, \quad \bar{y}_{22} = \frac{1}{\underset{\substack{\uparrow\\3}}{n}}(17+16+18) = \frac{51}{3} = 17,$$

Abweichungsquadratsummen:

$$\begin{aligned}
Q_{AxB} &= \underset{\substack{\uparrow\\3}}{n}\cdot \sum_{j=1}^{p=2}\sum_{k=1}^{q=2}\left(\bar{y}_{jk} - \bar{y}_{j\bullet} - \bar{y}_{\bullet k} + \bar{y}\right)^2 \\
&= 3\cdot\left[(20-18-21+18{,}75)^2 + (16-18-16{,}5+18{,}75)^2 + (22-19{,}5-21+18{,}75)^2 + (17-19{,}5-16{,}5+18{,}75)^2\right] \\
&= 3\cdot[(-0{,}25)^2 + 0{,}25^2 + 0{,}25^2 + (-0{,}25)^2] = 0{,}75
\end{aligned}$$

$$\begin{aligned}
Q_R &= \sum_{i=1}^{n=3}\sum_{j=1}^{p=2}\sum_{k=1}^{q=2}\left(y_{ijk} - \bar{y}_{jk}\right)^2 \\
&= (20-20)^2 + (18-20)^2 + (22-20)^2 + (16-16)^2 + (14-16)^2 + (18-16)^2 \\
&\quad + (24-22)^2 + (19-22)^2 + (23-22)^2 + (17-17)^2 + (16-17)^2 + (18-17)^2 \\
&= 0+4+4+0+4+4+4+9+1+0+1+1 \\
&= 32
\end{aligned}$$

Mittlere Abweichungssummen:

$$MQ_{AxB} = \frac{Q_{AxB}}{(p-1)(q-1)} = \frac{0{,}75}{(2-1)(2-1)} = 0{,}75$$

$$MQ_R = \frac{Q_R}{n \cdot p \cdot q - p \cdot q} = \frac{32}{3 \cdot 2 \cdot 2 - 2 \cdot 2} = \frac{32}{8} = 4$$

Prüfgröße:

$$F_{AxB} = \frac{MQ_{AxB}}{MQ_R} = \frac{0{,}75}{4} = 0{,}1875$$

Kritischer Wert ($\alpha = 0{,}05$):

$$F_{(p-1)(q-1);n \cdot p \cdot q - p \cdot q;0{,}95} = F_{1;8;0{,}95} = 5{,}32$$

Testentscheidung:

$F_{AxB} = 0{,}1875 < F_{1;\,8;\,0{,}95} = 5{,}32$ => Interaktionseffekt AxB nicht signifikant