

Regressions-Modelle für varianzanalytische Versuchspläne

Beispiel: Faktorieller Versuchsplan mit 2 Faktoren:

Zeilenfaktor "Arbeitsbedingung" mit den Stufen "nüchtern", "hübsch" und "altmodisch".

Spaltenfaktor "Geschlecht" mit den Stufen "weiblich" und "männlich".

Gruppengrößen:

	Frauen	Männer	Gesamt
nüchtern	$N_{nw} = 18$	$N_{nm} = 24$	$N_n = 42$
hübsch	$N_{hw} = 15$	$N_{hm} = 20$	$N_h = 35$
altmodisch	$N_{aw} = 12$	$N_{am} = 16$	$N_a = 28$
Gesamt	$N_w = 45$	$N_m = 60$	$N = 105$

Angenommene "wahre" Mittelwerte:

	Frauen	Männer	Gesamt
nüchtern	$\mu_{nw} = 148$	$\mu_{nm} = 172$	$\mu_n = 161.71$
hübsch	$\mu_{hw} = 168$	$\mu_{hm} = 146$	$\mu_h = 155.43$
altmodisch	$\mu_{aw} = 128$	$\mu_{am} = 126$	$\mu_a = 126.86$
Gesamt	$\mu_w = 149.33$	$\mu_m = 151.07$	$\mu = 150.32$

Varianzanalytische Effekte

Zeileneffekte: Der Effekt der Zeile ("Reihe") r ist definiert als Abweichung des Zeilenmittels μ_r vom Gesamtmittel μ :

$$a_r := \mu_r - \mu, \tag{2}$$

also:

$$a_n = \mu_n - \mu = 161.71 - 150.32 = 11.39$$

$$a_h = \mu_h - \mu = 155.43 - 150.32 = 5.11$$

$$a_a = \mu_a - \mu = 126.86 - 150.32 = -23.46$$

Spalteneffekte: Der Effekt der Spalte s ist definiert als Abweichung des Spaltenmittels μ_s vom Gesamtmittel μ :

$$b_s := \mu_s - \mu, \tag{3}$$

also:

$$b_w = \mu_w - \mu = 149.33 - 150.32 = -0.99$$

$$b_m = \mu_m - \mu = 151.07 - 150.32 = 0.75$$

Wechselwirkungseffekte: Der Wechselwirkungseffekt der Zelle z ist definiert als Abweichung des

Zellenmittels μ_z von dem fiktiven Wert, der allein aufgrund der Haupteffekte zu erwarten wäre. Diesen fiktiven Wert bestimmen wir, indem wir zum Gesamtmittel μ den Effekt der zur Zelle z gehörenden Reihe $r(z)$ und den Effekt der zur Zelle z gehörenden Spalte $s(z)$ addieren:

$$c_z := \mu_z - (\mu + a_{r(z)} + b_{s(z)}), \quad (4)$$

also:

$$\begin{aligned} c_{nw} &= \mu_{nw} - (\mu + a_n + b_w) = 148 - (150.32 + 11.39 + (-0.99)) &= -12.72 \\ c_{nm} &= \mu_{nm} - (\mu + a_n + b_w) = 172 - (150.32 + 11.39 + 0.75) &= 9.54 \\ c_{hw} &= \mu_{hw} - (\mu + a_h + b_w) = 168 - (150.32 + 5.11 + (-0.99)) &= 13.56 \\ c_{hm} &= \mu_{hm} - (\mu + a_h + b_w) = 146 - (150.32 + 5.11 + 0.75) &= -10.30 \\ c_{aw} &= \mu_{aw} - (\mu + a_a + b_w) = 128 - (150.32 + (-23.46) + (-0.99)) &= 2.13 \\ c_{am} &= \mu_{am} - (\mu + a_a + b_w) = 126 - (150.32 + (-23.46) + 0.75) &= -1.61 \end{aligned}$$

Fehler: Der "Fehler", mit dem ein Meßwert X_i behaftet ist, ist seine Abweichung vom Mittel der zugehörigen Zelle $z(i)$:

$$e_i := X_i - \mu_{z(i)}. \quad (5)$$

Ist z.B. $X_i = 142$ der Meßwert einer Frau unter Arbeitsbedingung "nüchtern", dann ist

$$e_i = X_i - \mu_{nw} = 142 - 148 = -6$$

Aus diesen Definitionen ergibt sich die folgende "Modellgleichung":

$$X_i = \mu + a_{r(i)} + b_{s(i)} + c_{z(i)} + e_i, \quad (6)$$

wobei $r(i)$, $s(i)$ und $z(i)$ für die Reihe, Spalte und Zelle der Person i stehen. Für den o.g. Meßwert X_i einer Frau unter Arbeitsbedingung "nüchtern" bedeutet das:

$$X_i = \mu + a_n + b_w + c_{nw} + e_i = 150.32 + 11.39 + (-0.99) + (-12.72) + (-6) = 142.$$

Durch einige Umformungen der Modellgleichung (5), die gleich erläutert werden, kommt man zu einer Gleichung, die formal den Aufbau einer multiplen Regressionsgleichung hat:

$$\begin{aligned} \hat{X}_i &= \beta_0 && \text{(Gesamtmittel)} \\ &+ \beta_n \cdot \delta_i^n + \beta_h \cdot \delta_i^h + \beta_a \cdot \delta_i^a && \text{(Zeileneffekte)} \\ &+ \beta_w \cdot \delta_i^w + \beta_m \cdot \delta_i^m && \text{(Spalteneffekte)} \\ &+ \beta_{nw} \cdot \delta_i^{nw} + \beta_{nm} \cdot \delta_i^{nm} + \beta_{hw} \cdot \delta_i^{hw} + \beta_{hm} \cdot \delta_i^{hm} + \beta_{aw} \cdot \delta_i^{aw} + \beta_{am} \cdot \delta_i^{am} && \text{(Wechselwirkungseffekte)} \end{aligned} \quad (7)$$

Diese neue Gleichung beruht auf den folgenden Veränderungen der Modellgleichung (5):

- Es ist nicht mehr eine Gleichung für den tatsächlichen Meßwert X_i angegeben, sondern für den

"geschätzten Meßwert" \hat{X}_i , und das ist das zugehörige Zellenmittel $\mu_{z(i)}$ (wegen seiner Least-squares-Eigenschaften). Nach der Definition des Fehlers (Gleichung (4)) ist $\mu_{z(i)} = X_i - e_i$; in der Modellgleichung (5) braucht also nur die Addition von e_i zu unterbleiben, um \hat{X}_i statt X_i zu erhalten.

- Für das Gesamtmittel μ ist die Bezeichnung β_0 verwendet.
- Es sind nicht mehr die Buchstaben a , b und c verwendet, um zwischen Zeilen-, Spalten- und Wechselwirkungseffekten zu unterscheiden. Alle Effekte heißen β , und der Index zeigt, ob es sich um einen Zeileneffekt handelt (z.B. β_n statt a_n ; den Index n gibt es nur für eine Zeile) oder um einen Spalteneffekt (z.B. β_w statt b_w) oder um einen Wechselwirkungseffekt (z.B. β_{nw} statt c_{nw}).
- In der Gleichung tauchen alle Effekte auf und nicht nur die am Meßwert X_i beteiligten. Sie werden aber mit "Indikatorvariablen" oder "Dummyvariablen" multipliziert. Z.B. ist δ_i^n eine Variable, die angibt, ob die Person i unter Arbeitsbedingung nüchtern gearbeitet hat. Trifft dies zu, dann ist $\delta_i^n = 1$, anderenfalls ist $\delta_i^n = 0$. (Beachten: Das hochgestellte n ist kein Exponent, sondern ein "Oben-Index".) Entsprechend gibt z.B. δ_i^w an, ob Person i eine Frau ist ($\delta_i^w = 1$) oder nicht ($\delta_i^w = 0$), und δ_i^{nw} gibt an, ob Person i zur Zelle "Frauen unter Arbeitsbedingung nüchtern" gehört ($\delta_i^{nw} = 1$) oder nicht. Damit stehen die am Meßwert X_i nicht beteiligten Effekte zwar in der Gleichung; sie werden aber durch Multiplikation mit 0 "herausgefiltert", und nur die am Meßwert X_i beteiligten Effekte sind mit 1 multipliziert.

Äußerlich hat die so modifizierte Gleichung (6) die Form einer multiplen Regressionsgleichung: X ist das Kriterium, die (durch verschiedene "Oben-Indices" unterschiedenen) δ -Variablen bilden die Prädiktoren, und die verschiedenen Effekte sind die zugehörigen "Regressionsgewichte". Ein zusätzlicher Summand in einer solchen multiplen Regressionsgleichung wird oft als a oder α bezeichnet und an das Ende der Gleichung geschrieben; die Bezeichnung β_0 und die Position am Anfang der Gleichung findet sich aber ebenfalls in vielen Darstellungen der multiplen Regression.

Allerdings ist noch ein Problem zu beheben, bevor Gleichung (6) wirklich als multiple Regressionsgleichung behandelt werden kann: Es bestehen mehrere Abhängigkeiten zwischen den Prädiktoren:

- Da jede Person zu genau einer Zeile gehört, ist $\delta_i^n + \delta_i^h + \delta_i^a = 1$.
- Da jede Person zu genau einer Spalte gehört, ist $\delta_i^w + \delta_i^m = 1$.
- Mit den δ -Werten für die Wechselwirkungen ist es etwas komplizierter.
 - Bei Personen, die unter Arbeitsbedingung "nüchtern" gearbeitet haben, ist genau eine beiden Dummyvariablen δ_i^{nw} und δ_i^{nm} gleich 1, und bei diesen Personen ist auch $\delta_i^n = 1$. Bei Personen, die unter einer der beiden anderen Arbeitsbedingungen gearbeitet haben, sind dagegen die drei Dummyvariablen δ_i^n , δ_i^{nm} und δ_i^{nw} alle null. In beiden Fällen gilt damit $\delta_i^n = \delta_i^{nw} + \delta_i^{nm}$.Ähnliche Überlegungen für die Haupt- und Wechselwirkungseffekte, an denen eine der Arbeitsbedingungen "hübsch" und "altmodisch" beteiligt ist, ergeben $\delta_i^h = \delta_i^{hw} + \delta_i^{hm}$ und $\delta_i^a = \delta_i^{aw} + \delta_i^{am}$.
- Für den Faktor Geschlecht ergibt sich entsprechend: Bei Frauen ist $\delta_i^w = 1$, und genau eine der drei Dummyvariablen δ_i^{nw} , δ_i^{hw} und δ_i^{aw} ist ebenfalls 1. Bei Männern sind diese vier Dummyvariablen dagegen alle null. In beiden Fällen ist

$$\delta_i^w = \delta_i^{nw} + \delta_i^{hw} + \delta_i^{aw}$$

Für die Haupt- und Wechselwirkungseffekte, an denen das Geschlecht "männlich" beteiligt ist, ergibt sich entsprechend:

$$\delta_i^m = \delta_i^{nm} + \delta_i^{hm} + \delta_i^{am}$$

Derartige Abhängigkeiten der Prädiktoren führen zu Problemen mit der multiplen Regression. Man kann diese Probleme auf verschiedene Arten lösen: Man kann bei der Codierung der Gruppenzugehörigkeit mit Dummyvariablen bleiben und einen anderen Effektbegriff zugrundelegen.¹ Man kann aber auch den varianzanalytischen Effektbegriff beibehalten und zu einer anderen Codierung der Gruppenzugehörigkeit übergehen. Weitere Codierungsmethoden für die Gruppenzugehörigkeit wurden vor allem für hypothesengeleitete Mittelwertsvergleiche entwickelt.

Dummycodierung für Referenzgruppen-bezogene Effekte

In manchen Situationen ist ein anderer Effektbegriff als der varianzanalytische sehr sinnvoll. Wenn in einem Experiment mit einer Experimentalgruppe und einer Kontrollgruppe die "wahren" Mittelwerte für die Kontrollgruppe 23 und für die Experimentalgruppe 27 betragen, dann kann man auch einfach sagen: Der Effekt der Zugehörigkeit zur Experimentalgruppe (und damit der Effekt der nur in der Experimentalgruppe realisierten Bedingung) besteht in einer Hebung des Mittelwerts um 4 Punkte *im Vergleich zur Kontrollgruppe*. Damit legen wir aber einen anderen Effektbegriff zugrunde als in der Varianzanalyse: Der Effekt ist nicht mehr als Abweichung eines Gruppenmittels vom Gesamtmittel definiert, sondern als Abweichung von einer "Referenzgruppe" - der Kontrollgruppe. Durch eine Beschreibung mit diesem Referenzgruppen-bezogenen Effektbegriff wird das, was man in einem Experiment untersuchen will, sogar noch klarer ausgedrückt als mit dem varianzanalytischen.

Bei gleicher Größe von Experimentalgruppe und Kontrollgruppe würde sich ein Gesamtmittel von 25 ergeben. Also besteht der varianzanalytische Effekt der Zugehörigkeit zur Experimentalgruppe in einer Anhebung des Mittelwerts um 2 Punkte *im Vergleich zum Gesamtmittel*, und der varianzanalytische Effekt der Zugehörigkeit zur Kontrollgruppe besteht in einer Senkung des Mittelwerts um 2 Punkte *im Vergleich zum Gesamtmittel*.

Beim Vergleich von Experimentalgruppe und Kontrollgruppe ist es unmittelbar naheliegend, die Kontrollgruppe als Referenzgruppe zu wählen und damit den Effekt der Zugehörigkeit zur Experimentalgruppe relativ zur Kontrollgruppe zu bestimmen. In anderen Situationen kann man formal ähnlich verfahren. Wenn etwa bei einem Vergleich zweier Therapieformen A und B die wahren Mittelwerte eines Erfolgsmaßes $\mu_A = 23$ und $\mu_B = 27$ betragen, dann kann man die Gruppe A zur Referenzgruppe erklären und sagen: Der Effekt der Zugehörigkeit zur Gruppe B besteht in einer Anhebung des Mittelwerts um 4 Punkte *im Vergleich zur Gruppe A*. Ebenso gut kann man aber auch sagen: Der Effekt der Zugehörigkeit zur Gruppe A besteht in einer Senkung des Mittelwerts um 4 Punkte *im Vergleich zur Gruppe B*. Dabei ist dann die Gruppe B Referenzgruppe.

¹Daß die Dummycodierung der Gruppenzugehörigkeit implizit mit einer Veränderung des Effektbegriffs verbunden ist, wird in vielen Lehrbüchern nicht erwähnt. Darauf wird in den Schlußbemerkungen des vorliegenden Skripts näher eingegangen.

Verallgemeinert: Auch wenn es nicht durch die Situation vorgegeben ist, welche Gruppe Referenzgruppe ist, kann man den Referenzgruppen-bezogenen Effektbegriff verwenden; es ist dann aber besonders wichtig, die Bedeutung eines Effekts im Auge zu behalten - etwa indem man ihn in einem Satz formuliert, in dem ausdrücklich die Formulierung "im Vergleich zu ..." vorkommt.

In einem mehrfaktoriellen Versuchsplan wird eine Zelle zur Referenzgruppe erklärt. Ihre Wahl ist grundsätzlich beliebig, solange man die durch diese Wahl entstehende Bedeutung der Effekte im Auge behält. Für "Haupteffekte" gilt dann die folgende Regel:

Referenzgruppen-bezogene Haupteffekte in mehrfaktoriellen Versuchsplänen:

Haupteffekte gibt es nur für die Faktorstufen, die in der Referenzgruppe nicht realisiert sind. Sie werden auf der "Referenzstufe" des anderen Faktors (bzw. der anderen Faktoren) bestimmt.

Beispiel: Für die obige Untersuchung mit den Faktoren Arbeitsbedingung und Geschlecht soll im folgenden die Zelle "Männer unter Arbeitsbedingung nüchtern" Referenzgruppe sein. Dann gibt es Haupteffekte im Faktor Arbeitsbedingung nur für die Stufen "hübsch" und "altmodisch", und im Faktor Geschlecht gibt es nur einen Haupteffekt der Stufe "weiblich".

Daß es für die in der Referenzgruppe realisierten Stufen "nüchtern" und "männlich" keine Haupteffekte gibt, ist ähnlich zu verstehen wie beim Vergleich von Experimentalgruppe und Kontrollgruppe, bei dem es nur einen Haupteffekt der Zugehörigkeit zur Experimentalgruppe gibt. Man könnte zwar auch sagen, daß die Effekte der Stufen "nüchtern" und "männlich" immer null sind (und entsprechend der Effekt der Zugehörigkeit zur Kontrollgruppe); aber es wäre irreführend, für eine Größe, deren Wert 0 sich ohne jeden Bezug zu den Mittelwerten automatisch aus der Definition ergibt, dieselbe Bezeichnung "Effekt" zu verwenden wie für die auf Mittelwertsdifferenzen beruhenden Effekte.

Nach der obigen Regel werden die Haupteffekte für die Stufen des Faktors Arbeitsbedingung auf der Referenzstufe des Faktors Geschlecht bestimmt. Das ist die in der Referenzgruppe realisierte Stufe, also "männlich". Als Symbol verwenden wir wieder β mit entsprechendem Index, müssen aber im Auge behalten, daß diese Symbole jetzt andere Effekte bezeichnen als früher. Damit ergibt sich:

$$\beta_h := \mu_{hm} - \mu_{nm} = 146 - 172 = -26$$

$$\beta_a := \mu_{am} - \mu_{nm} = 126 - 172 = -46$$

Entsprechend wird der Haupteffekt des Geschlechts "weiblich" auf der Referenzstufe "nüchtern" des Faktors Arbeitsbedingung bestimmt:

$$\beta_w := \mu_{nw} - \mu_{nm} = 148 - 172 = -24$$

In Worten:

- Der "Haupteffekt" der Arbeitsbedingung "hübsch" besteht darin, daß unter dieser Arbeitsbedingung *bei den Männern* der Mittelwert der Meßwerte um 26 Punkte erniedrigt ist *im Vergleich zur Arbeitsbedingung "nüchtern"*.
- Der "Haupteffekt" der Arbeitsbedingung "altmodisch" besteht darin, daß unter dieser Arbeitsbedingung *bei den Männern* der Mittelwert der Meßwerte um 46 Punkte erniedrigt ist *im*

Vergleich zur Arbeitsbedingung "nüchtern".

- Der "Haupteffekt" des Geschlechts "weiblich" besteht darin, daß bei den Frauen *unter Arbeitsbedingung "nüchtern"* der Mittelwert der Meßwerte um 46 Punkte erniedrigt ist *im Vergleich zu den Männern.*

Diese Formulierungen machen deutlich, wie wichtig es ist, die genaue Bedeutung Referenzgruppen-bezogener "Haupteffekte" im Auge zu behalten. Hätten wir z.B. die Gruppe "Männer unter Arbeitsbedingung hübsch" zur Referenzgruppe erklärt, dann würde der "Haupteffekt" des Geschlechts "weiblich" aufgrund der Mittelwerte unter Arbeitsbedingung "hübsch" bestimmt, und es ergäbe sich $\beta_w := \mu_{hw} - \mu_{hm} = 168 - 146 = 22$. Es hängt also von der Wahl der Referenzstufe im Faktor Arbeitsbedingung ab, ob der Haupteffekt des Geschlechts "weiblich" negativ oder positiv ist! Das liegt daran, daß eine "disordinale Interaktion" vorliegt. Allgemein ist der Begriff "Haupteffekt" beim Referenzgruppen-Ansatz problematisch; denn es handelt sich nur um einen Effekt auf *einer*, meist willkürlich gewählten Stufe des anderen Faktors. (Deshalb ist der Begriff "Haupteffekt" in den obigen Formulierungen der Ergebnisse auch in Anführungszeichen gesetzt.)

Konsequenz: Bei den "Haupteffekten" ist es immer wichtig, die Referenzstufen beider Faktoren im Auge zu behalten: Mit welcher Stufe desselben Faktors wird verglichen, und auf welcher Stufe des anderen Faktors erfolgt der Vergleich?

Für die Wechselwirkungseffekte gilt die folgende Regel:

Referenzgruppen-bezogene Wechselwirkungseffekte in mehrfaktoriellen Versuchsplänen:

Wechselwirkungseffekte gibt es nur für Kombinationen von Faktorstufen, die alle in der Referenzgruppe nicht realisiert sind. Sie geben an, um welchen Betrag der Mittelwert eine Stufenkombination oberhalb (bzw. unterhalb bei negativem Vorzeichen) des Werts liegt, der aufgrund des Mittelwerts der Referenzgruppe und aufgrund der Haupteffekte (sowie ggf. der Wechselwirkungseffekte niedrigerer Ordnung²) zu erwarten wäre.

Bei der Beispieluntersuchung gibt es nach dieser Regel keine Wechselwirkungseffekte für die Stufenkombinationen nw, nm, hm und am, da jeweils bei mindestens einem Faktor dieselbe Stufe wie bei der Referenzgruppe vorliegt.

Nähere Begründung: Für die Stufenkombination nw folgt aus der Definition des Haupteffekts β_w die Gleichung $\mu_{nw} = \mu_{nm} + \beta_w$. Daher kann es keine Abweichung des Mittelwerts μ_{nw} von dem Wert ergeben, der aufgrund des Mittelwerts μ_{nm} der Referenzgruppe und aufgrund des Haupteffekts β_w der Stufe "weiblich" ergeben würde. (Einen Haupteffekt der Stufe "nüchtern" gibt es ohnehin nicht!) Man könnte zwar von einem Wechselwirkungseffekt sprechen, der automatisch null ist; aber dagegen sprechen ähnliche Gründe wie bei den Haupteffekten.

Für die Stufenkombination hw (Frauen unter Arbeitsbedingung "hübsch") wäre aufgrund des Mittelwerts der Referenzgruppe sowie aufgrund der Haupteffekte der Stufen "hübsch" und "weiblich" ein fiktiver Mittelwert von $\mu_{nm} + \beta_h + \beta_w = 172 + (-26) + (-24) = 122$ zu erwarten.

²Zu "Wechselwirkung niedrigerer Ordnung": In einem dreifaktoriellen Versuchsplan gibt es auch Wechselwirkungen 2. Ordnung. Für die Definition der entsprechenden Wechselwirkungseffekte sind bei der Bestimmung eines "fiktiven Mittelwerts" auch die Wechselwirkungen 1. Ordnung zu berücksichtigen.

Der tatsächliche Mittelwert $\mu_{hw} = 168$ liegt um 46 Punkte oberhalb dieses fiktiven Werts, und das ist dann der Wechselwirkungseffekt β_{hw} . Zusammengefaßt:

$$\beta_{hw} := \mu_{hw} - (\mu_{nm} + \beta_h + \beta_w) = 168 - (172 + (-26) + (-24)) = 46.$$

Entsprechend für den Wechselwirkungseffekt der Stufenkombination aw (Frauen unter Arbeitsbedingung "altmodisch"): Der "fiktive Mittelwert" $\mu_{nm} + \beta_a + \beta_w = 172 + (-46) + (-24) = 102$ wird vom tatsächlichen Mittelwert $\mu_{aw} = 128$ abgezogen:

$$\beta_{aw} := \mu_{aw} - (\mu_{nm} + \beta_a + \beta_w) = 128 - (172 + (-46) + (-24)) = 26.$$

Mit den so definierten Effekten kann man nun die folgende Gleichung aufstellen, bei der \hat{X}_i wieder der Mittelwert der Zelle ist, zu der der Meßwert X_i gehört, während β_0 der Mittelwert der Referenzgruppe ist:

$$\hat{X}_i = \beta_0 \quad + \beta_h \cdot \delta_i^h + \beta_a \cdot \delta_i^a \quad + \beta_w \cdot \delta_i^w \quad + \beta_{hw} \cdot \delta_i^{hw} + \beta_{aw} \cdot \delta_i^{aw} \quad (8)$$

(Ref.gr) (ggf. Zeileneffekt) (ggf. Spalteneffekt) (ggf. Wechselwirkungseffekt)

In Worten:

- Die Summe beginnt immer mit dem Mittelwert der Referenzgruppe.
- Wenn die Person i unter Arbeitsbedingung "nüchtern" gearbeitet hat, ist $\delta_i^h = 0$ und $\delta_i^a = 0$; es wird also faktisch keiner der Zeileneffekte addiert. Hat Person i dagegen unter einer anderen Arbeitsbedingung gearbeitet, dann ist eine der beiden Dummyvariablen δ_i^h und δ_i^a gleich 1, und dann wird der entsprechende Zeileneffekt addiert.
- Wenn die Person i ein Mann ist, ist $\delta_i^w = 0$; es wird also faktisch kein Spalteneffekt addiert. Ist die Person i dagegen eine Frau, dann ist $\delta_i^w = 1$, und dann wird der Spalteneffekt β_w addiert.
- Wenn die Person i zu einer der Zellen nw, nm und am gehört, zu denen es keine Wechselwirkung gibt, dann ist $\delta_i^{hw} = 0$ und $\delta_i^{aw} = 0$; es wird also faktisch keiner der Wechselwirkungseffekte addiert. Gehört die Person dagegen zu einer der Zellen hw und aw, zu denen es einen Wechselwirkungseffekt gibt, dann ist eine der beiden Dummyvariablen δ_i^{hw} und δ_i^{aw} gleich 1, und dann wird der entsprechende Wechselwirkungseffekt addiert.

Wieder hat unsere Gleichung die Form einer multiplen Regressionsgleichung; aber diesmal gibt es keine unzulässigen Abhängigkeiten zwischen den Prädiktoren (den δ -Variablen) mehr. Es gibt zwar noch die multiplikativen Abhängigkeiten $\delta_i^{hw} = \delta_i^h \cdot \delta_i^w$ und $\delta_i^{aw} = \delta_i^a \cdot \delta_i^w$, aber multiplikative Abhängigkeiten der Prädiktoren sind in Regressionsgleichungen "erlaubt".

Effektcodierung der Gruppenzugehörigkeit

Beim Referenzgruppen-bezogenen Ansatz wurde die störende Abhängigkeit der Prädiktoren (also der Dummyvariablen in Gleichung (6)) durch einen veränderten Effektbegriff beseitigt, bei dem es für bestimmte Faktorstufen keinen Haupteffekt gibt, und für verschiedene Stufenkombinationen keinen Wechselwirkungseffekt. Bei einer anderen Lösung des Problems behält man den varianzanalytischen Effektbegriff bei und verwendet eine andere Codierung der Gruppenzugehörigkeit, die sog. Effektcodierung. Dabei kommt dann eine Regressionsgleichung

heraus, die sich - ähnlich wie Gleichung (7) für die Dummycodierung - systematisch herleiten ließe. Das ist aber etwas mühsam, und deshalb soll gleich das Ergebnis dargestellt werden, und dann wird gezeigt, nach welchem Prinzip die entsprechenden Codiervariablen gebildet sind.

Ebenso wie bei der Dummycodierung kann man auch bei der Effektcodierung für jeden Faktor eine Stufe festlegen, für die kein Haupteffekt in der Regressionsgleichung auftauchen soll, und dann verschwinden auch alle Wechselwirkungseffekte, an denen mindestens eine der Faktorstufen beteiligt ist, für die kein Haupteffekt vorkommt. Wenn man wieder die Arbeitsbedingung "nüchtern" und das Geschlecht "männlich" aus Gleichung (6) loswerden will, dann entsteht eine Regressionsgleichung folgender Art:

$$\hat{X}_i = \beta_0 + \beta_h \cdot \gamma_i^h + \beta_a \cdot \gamma_i^a + \beta_w \cdot \gamma_i^w + \beta_{hw} \cdot \gamma_i^{hw} + \beta_{aw} \cdot \gamma_i^{aw} \quad (9)$$

(Gesamtmittel) (Zeileneffekte) (Spalteneffekte) (Wechselwirkungseffekte)

Dabei sind \hat{X}_i und die verschiedenen β -Größen wie in Gleichung (6) zu interpretieren. Insbesondere sind also die Haupteffekte β_h , β_a und β_w sowie die Wechselwirkungseffekte β_{hw} und β_{aw} als varianzanalytische Effekte zu verstehen und nicht als Referenzgruppen-bezogene Effekte wie in der Regressionsgleichung (7) für die Dummycodierung.

Die Prädiktoren in Regressionsgleichung (8) sind die Codiervariablen γ_i^h , γ_i^a , γ_i^w , γ_i^{hw} und γ_i^{aw} . Genaue Definitionen dieser Codiervariablen folgen später (Gleichungen (14), (15), (11), (19) und (20)). Zunächst sollen die Werte dieser Codiervariablen, mit denen man faktisch arbeitet, in einer Tabelle zusammengestellt werden. Dann wird gezeigt, wie sie faktisch arbeiten, und daraus ergibt sich dann die passende Definition.

In der folgenden Tabelle ist also zunächst dargestellt, mit welchen Werten der Codiervariablen γ_i^h , γ_i^a , γ_i^w , γ_i^{hw} und γ_i^{aw} faktisch gearbeitet wird, wenn man sich entschieden hat, daß die Haupteffekte der Arbeitsbedingung "nüchtern" und des Geschlechts "männlich" aus der Regressionsgleichung verschwinden sollen. Natürlich hängt dies davon ab, zu welcher Zelle des Versuchsplans die Person i gehört, und deshalb sind diese Zellen in der ersten Spalte der Tabelle einzeln aufgeführt. In den weiteren Spalten folgen dann die Werte für die jeweils im Spaltenkopf genannten Codiervariablen.

Effektcodierung der verschiedenen Zellen

Zelle	γ_i^h	γ_i^a	γ_i^w	γ_i^{hw}	γ_i^{aw}
nw	$-N_h / N_n$	$-N_a / N_n$	1	$-N_h / N_n$	$-N_a / N_n$
nm	$-N_h / N_n$	$-N_a / N_n$	$-N_w / N_m$	$(N_h / N_n) \cdot (N_w / N_m)$	$(N_a / N_n) \cdot (N_w / N_m)$
hw	1	0	1	1	0
hm	1	0	$-N_w / N_m$	$-N_w / N_m$	0
aw	0	1	1	0	1
am	0	1	$-N_w / N_m$	0	$-N_w / N_m$

Das Prinzip, das dieser Tabelle zugrundeliegt, ist bei diesem Beispiel am einfachsten für die Spalteneffekte zu verstehen, also für die Codiervariable γ_i^w .

- Ist die Person i eine Frau (Zellen nw, hw und aw), dann ist $\gamma_i^w = 1$. Die Addition des Produkts $\beta_w \cdot \gamma_i^w$ in Gleichung (8) bewirkt also, daß der Effekt β_w des Geschlechts "weiblich" addiert wird.
- Ist die Person i ein Mann (Zellen nm, hm und am), dann ist $\gamma_i^w = -N_w / N_m$. Die Addition des

Produkts $\beta_w \cdot \gamma_i^w$ in Gleichung (8) bewirkt also nicht unmittelbar, daß der Effekt des Geschlechts "männlich" addiert wird, sondern $\beta_w \cdot (-N_w / N_m)$. Die folgende Überlegung zeigt aber, daß dies faktisch darauf hinausläuft, den Effekt des Geschlechts "männlich" zu addieren.

Um das nachzuvollziehen, sollte man sich (etwa aus der Statistikausbildung) daran erinnern, daß zwischen den Haupteffekten der Stufen desselben Faktors eine Abhängigkeit besteht. Bei gleichen Gruppengrößen gilt bekanntlich die Regel, daß die Summe der Zeileneffekte null ist, und ebenso die Summe der Spalteneffekte. Bei ungleichen Gruppengrößen (wie in unserem Beispiel) ergibt sich eine Summe von null, wenn man die Effekte mit den entsprechenden Gruppengrößen multipliziert. Für den Faktor "Geschlecht" bedeutet das

$$N_w \cdot \beta_w + N_m \cdot \beta_m = 0. \quad (10)$$

Löst man diese Gleichung nach β_m auf, dann erhält man

$$\beta_m = -N_w \cdot \beta_w / N_m = \beta_w \cdot (-N_w / N_m), \quad (11)$$

wobei das zweite Gleichheitszeichen auf einer einfachen algebraischen Umformung beruht, durch die folgende Überlegung erleichtert wird: Ist die Person i ein Mann (gehört sie also zu einer der Zellen nm , hm und am), dann ist (laut Tabelle) $\gamma_i^w = -N_w / N_m$. Die Addition des Produkts $\beta_w \cdot \gamma_i^w$ in Gleichung (8) läuft also bei Männern darauf hinaus, daß $\beta_w \cdot (-N_w / N_m)$ addiert wird, und nach Gleichung (10) bedeutet dies, daß der Effekt des Geschlechts "männlich" addiert wird. Man könnte auch sagen, daß der Wert γ_i^w für Männer so gewählt ist, daß das Produkt $\beta_w \cdot \gamma_i^w$ in Gleichung (8) eine Art "Ersatzdarstellung" des Effekts β_m ergibt, die auf der in Gleichung (9) ausgedrückten Abhängigkeit der Effekte β_w und β_m beruht.

Nachdem geklärt ist, was die in der Tabelle angegebenen Werte der Codiervariablen γ_i^w faktisch bewirken, ist es leicht, diese Codiervariable so zu definieren, daß genau diese Wirkung erzielt wird: Wenn das Produkt $\beta_w \cdot \gamma_i^w$ in Gleichung (8) für Frauen den Effekt β_m ergeben soll und für Männer eine auf Gleichung (10) beruhende Ersatzdarstellung des Effekts β_m , dann müssen genau die in der Tabelle angegebenen Werte der Codiervariablen γ_i^w verwendet werden, und das erreicht man durch die Definition

$$\gamma_i^w := \delta_i^w - \delta_i^m \cdot N_w / N_m. \quad (12)$$

Daß diese Definition in der Tat bei allen Zellen zu den in der Tabelle angegebenen Werten für die Codiervariable γ_i^w führt, läßt sich leicht nachprüfen. Da aber im weiteren Verlauf noch für andere Codiervariablen eine Definitionsgleichung zu erstellen sein wird, bei der bestimmte, bereits bekannte Werte der Codiervariablen in den verschiedenen Zellen herauskommen sollen, soll kurz angegeben werden, wie man eine solche Definitionsgleichung erstellt. Das Prinzip ist noch leichter zu erkennen, wenn man Gleichung (11) folgendermaßen schreibt:

$$\gamma_i^w := \delta_i^w \cdot 1 + \delta_i^m \cdot (-N_w / N_m).$$

In dieser Form ist das Prinzip leicht zu erkennen: Für alle Werte, die bei der zu definierenden Codiervariablen vorkommen sollen (in unserem Fall also für die Werte 1 und $-N_w / N_m$) bildet man Dummyvariablen, die angeben, in welchen Zellen der entsprechende Wert herauskommen

soll. (Hier: Der Wert 1 soll beim Geschlecht "weiblich" herauskommen, und der Wert $-N_w / N_m$ beim Geschlecht "männlich".) Mit diesen Dummyvariablen multipliziert man die Werte, bildet die Summe der so entstandenen Produkte und vereinfacht die Gleichung.

Für Zeileneffekte, also für die Codiervariablen γ_i^h und γ_i^a , sind aufgrund ganz ähnlicher Überlegungen die in der Tabelle angegebenen Werte so gewählt, daß die für die Zeileneffekte stehende Summe $\beta_h \cdot \gamma_i^h + \beta_a \cdot \gamma_i^a$ in Gleichung (8) unter den Arbeitsbedingungen "hübsch" und "altmodisch" den entsprechenden Zeileneffekt β_h bzw. β_a ergibt. (Wollen Sie es überprüfen?) Für den Effekt β_n der Arbeitsbedingung "nüchtern", der (gemäß anfänglicher Vorgabe) aus der Regressionsgleichung verschwinden soll, wird wieder eine Ersatzdarstellung gewählt, die auf der Abhängigkeit der Haupteffekte beruht. Das Grundprinzip, daß die zu einem Faktor gehörenden Haupteffekte nach Multiplikation mit den Gruppengrößen zusammen null ergeben, läßt sich hier in der Gleichung

$$N_n \cdot \beta_n + N_h \cdot \beta_h + N_a \cdot \beta_a = 0, \quad (13)$$

ausdrücken. Löst man diese Gleichung nach β_n auf, dann ergibt sich

$$\beta_n = -(N_h \cdot \beta_h + N_a \cdot \beta_a) / N_n = \beta_h \cdot (-N_h / N_n) + \beta_a \cdot (-N_a / N_n), \quad (14)$$

wobei das zweite Gleichheitszeichen wieder auf einer einfachen algebraischen Umformung beruht. Legt man nun unter Arbeitsbedingung "nüchtern" die in der Tabelle der Codiervariablenwerte angegebenen Werte von $\gamma_i^h = (-N_h / N_n)$ und $\gamma_i^a = (-N_a / N_n)$ zugrunde, dann läßt sich aus Gleichung (13) entnehmen, daß unter dieser Arbeitsbedingung die Summe $\beta_h \cdot \gamma_i^h + \beta_a \cdot \gamma_i^a$ in Gleichung (8) eine Ersatzdarstellung für den Effekt β_n ergibt.

Zusammengefaßt: Die Wahl der in der entsprechenden Tabelle angegebenen Werte für die Codiervariablen γ_i^h und γ_i^a bewirkt, daß die Summe $\beta_h \cdot \gamma_i^h + \beta_a \cdot \gamma_i^a$ in Gleichung (8) für die Arbeitsbedingungen "hübsch" und "altmodisch" den entsprechenden Zeileneffekt β_h bzw. β_a ergibt, und für Arbeitsbedingung "nüchtern" ergibt diese Summe eine auf Gleichung (13) beruhende Ersatzdarstellung des Effekts β_n . All das wird erreicht, indem man die beiden Codiervariablen γ_i^h und γ_i^a durch die Gleichungen

$$\gamma_i^h := \delta_i^h - \delta_i^n \cdot N_h / N_n, \quad (15)$$

und

$$\gamma_i^a := \delta_i^a - \delta_i^n \cdot N_a / N_n. \quad (16)$$

definiert.

Auch bei den Codiervariablen γ_i^{hw} und γ_i^{aw} für die Wechselwirkungseffekte sind in der Tabelle der Codiervariablenwerte ganz ähnliche Prinzipien zugrundegelegt wie bei den Haupteffekten. Für die beiden Zellen hw und aw, deren Wechselwirkungseffekte β_{hw} und β_{aw} in der Regressionsgleichung (8) vorkommen, sind die Werte der beiden Codiervariablen so gewählt, daß der entsprechende Wechselwirkungseffekt mit 1 multipliziert wird und der andere mit 0. Für alle übrigen Zellen erfolgt eine Ersatzdarstellung, die wieder auf Abhängigkeitsregeln beruht.

Allerdings gibt es für Wechselwirkungseffekte mehrere derartige Abhängigkeiten, die sich folgendermaßen formulieren lassen:

- Die Wechselwirkungseffekte, an denen dieselbe Stufe des Spaltenfaktors beteiligt ist, ergeben nach Multiplikation mit der entsprechenden Gruppengröße eine Summe 0. Je nach dem, ob man das Geschlecht "weiblich" oder das Geschlecht "männlich" konstant hält, ergibt sich also

$$N_{nw} \cdot \beta_{nw} + N_{hw} \cdot \beta_{hw} + N_{aw} \cdot \beta_{aw} = 0 \quad (17.w)$$

bzw.

$$N_{nm} \cdot \beta_{nm} + N_{hm} \cdot \beta_{hm} + N_{am} \cdot \beta_{am} = 0. \quad (16.m)$$

- Die Wechselwirkungseffekte, an denen dieselbe Stufe des Zeilenfaktors beteiligt ist, ergeben ebenfalls nach Multiplikation mit der entsprechenden Gruppengröße eine Summe 0. Für die drei Stufen des Faktors "Arbeitsbedingung" ergibt sich also

$$N_{nw} \cdot \beta_{nw} + N_{nm} \cdot \beta_{nm} = 0 \quad (18.n)$$

bzw.

$$N_{hw} \cdot \beta_{hw} + N_{hm} \cdot \beta_{hm} = 0 \quad (17.h)$$

bzw.

$$N_{aw} \cdot \beta_{aw} + N_{am} \cdot \beta_{am} = 0 \quad (17.a)$$

Aus diesen Gleichungen lassen sich nun die folgenden Ersatzdarstellungen für die Wechsel-

wirkungseffekte herleiten, die in der Regressionsgleichung (8) nicht vorkommen:³

$$\beta_{nw} = \beta_{hw} \cdot (-N_h / N_n) + \beta_{aw} \cdot (-N_a / N_n), \quad (19.nw)$$

$$\beta_{nm} = \beta_{hw} \cdot (N_h / N_n) \cdot (N_w / N_m) + \beta_{aw} \cdot (N_a / N_n) \cdot (N_w / N_m), \quad (18.nm)$$

$$\beta_{hm} = \beta_{hw} \cdot (-N_w / N_m), \quad (18.hm)$$

und

$$\beta_{am} = \beta_{aw} \cdot (-N_w / N_m). \quad (18.am)$$

Aufgrund dieser Ersatzdarstellungen läßt sich leicht nachprüfen, daß die Werte der Codiervariablen γ_i^{hw} und γ_i^{aw} in der Codiervariablen-tabelle so gewählt sind, daß für jede Zelle, deren Wechselwirkungseffekt in der Regressionsgleichung (8) nicht vorkommt, die in dieser Regressionsgleichung vorkommende Summe $\beta_{hw} \cdot \gamma_i^{hw} + \beta_{aw} \cdot \gamma_i^{aw}$ eine Ersatzdarstellung des Wechselwirkungseffekts für die jeweilige Zelle ergibt.

Für eine formelmäßige Definition kann man nun in der Codiervariablen-tabelle feststellen, daß bei allen Zellen der Wert der beiden Codiervariablen γ_i^{hw} und γ_i^{aw} gleich dem Produkt der Codiervariablen für den Zeileneffekt und den Spalteneffekt ist, also

$$\gamma_i^{hw} := \gamma_i^h \cdot \gamma_i^w \quad (20)$$

und

$$\gamma_i^{aw} := \gamma_i^a \cdot \gamma_i^w. \quad (21)$$

Ein allgemeiner Beweis, daß diese Definitionen nicht nur im speziellen Fall unserer Untersuchung sinnvoll sind, würde aber zu weit führen.⁴

³Für die Gleichungen (18.nw), (18.hm) und (18.am) löst man zunächst die Gleichungen (16.w), (17.h) und (17.a) nach β_{nw} bzw. β_{hm} bzw. β_{am} auf. Dann kann man die Quotienten von Gruppengrößen (wie z.B. N_{hw} / N_{nw}), die sich bei dieser Auflösung ergeben, vereinfachen: Die Proportionalitäten

$$N_{nw} : N_{hw} : N_{aw} = N_{nm} : N_{hm} : N_{am} = N_n : N_h : N_a$$

und

$$N_{nw} : N_{nm} = N_{hw} : N_{hm} = N_{aw} : N_{am} = N_w : N_m$$

gehen aus den Gruppengrößen der Beispieluntersuchung hervor. (Entsprechendes gilt bei allen Versuchsplänen mit "orthogonalen" Gruppengrößen.)

Für Gleichung (18.nm) ergibt sich zunächst $\beta_{nm} = \beta_{nw} \cdot (-N_w / N_m)$ aus Gleichung (17.n), und dann ersetzt man β_{nw} nach Gleichung (18.nw).

⁴Ansatz des Beweises: Zunächst konstruiert man Definitionen der Codiervariablen γ_i^{hw} und γ_i^{aw} nach den im Anschluß an Gleichung (11) dargestellten Prinzipien und zeigt dann, daß die in den Gleichungen (19) und (20) vorgenommenen Definitionen zum gleichen Ergebnis führen, (Fortsetzung...)

Kontrastcodierung der Gruppenzugehörigkeit

Bisher sind wir davon ausgegangen, daß die "wahren Mittelwerte" für Zeilen, Spalten und Zellen bekannt sind. In der Realität ist das natürlich anders. Nehmen wir also an, es handle sich bei den bisher zugrundegelegten Mittelwerten nur um Durchschnitte aus einer Untersuchung. Ob dann der recht deutliche Effekt des Faktors Arbeitsbedingung signifikant ist, kann allein aufgrund der Mittelwerte nicht entschieden werden, da das auch von der Varianz innerhalb der Gruppen abhängt. Aber selbst wenn dieser Effekt signifikant ist, kann in der Varianzanalyse lediglich die Nullhypothese verworfen werden, daß *alle* "wahren" Zeilenmittelwerte gleich sind. Konkret ist es auch bei einem signifikanten Haupteffekt des Faktors Arbeitsbedingung durchaus denkbar, daß die wahren Mittelwerte unter den Arbeitsbedingungen "nüchtern" und "hübsch" identisch sind und daß lediglich ein Unterschied zum wahren Mittelwert unter Arbeitsbedingung "altmodisch" besteht.

Hat man vor der Untersuchung spezifische Hypothesen zu solchen Fragen, dann kann man die Gruppenzugehörigkeit auch in einer Weise codieren, durch die Signifikanztests zu derartigen Fragen ermöglicht werden. In der Beispieluntersuchung können entsprechende Hypothesen durch eine Umbenennung der Arbeitsbedingungen vorbereitet werden. Wenn man die beiden ersten Arbeitsbedingungen als "funktional gestaltet" bzw. "dekorativ gestaltet" und die letzte als "ungestaltet" bezeichnet, dann legt das folgende Fragestellungen nahe:

- Besteht ein Unterschied zwischen gestalteten und ungestalteten Arbeitsbedingungen?
- Besteht ein Unterschied zwischen den zwei Arten der Gestaltung?

Für die erste Frage werden also die beiden Gruppen mit gestalteter Arbeitsbedingung zusammengefaßt, und der Mittelwert dieser Gruppe wird mit dem Mittelwert für ungestaltete Arbeitsbedingung verglichen. Bei der zweiten Frage bleibt dagegen die ungestaltete Arbeitsbedingung unberücksichtigt, und es werden die Mittelwerte der beiden gestalteten Arbeitsbedingungen verglichen. Man sagt auch, daß in beiden Fällen "Kontraste" von Mittelwerten gebildet werden, und deshalb wird die Codierung der Gruppenzugehörigkeit, mit der solche Unterschiede untersucht werden, auch als "Kontrastcodierung" bezeichnet.

Im folgenden werden zunächst Effekte definiert, die die beiden obigen Fragen beantworten. Dann wird gezeigt, welche Ersatzdarstellungen möglich sind, und dann werden entsprechende Codiervariablen für die Effekte definiert, die in die Regressionsgleichung eingehen sollen.

Für den Vergleich von gestalteten und ungestalteten Arbeitsbedingungen sind zunächst die beiden Mittelwerte zu definieren, die verglichen ("kontrastiert") werden sollen. Dazu faßt man die beiden gestalteten Arbeitsbedingungen in einer Gruppe g (für "gestaltete Arbeitsbedingung") zusammen und bildet einen "wahren" Mittelwert μ_g für diese Gruppe. Dabei ist die unterschiedliche Größe der beiden zusammengefaßten Gruppen zu berücksichtigen:

$$\mu_g = (N_f \cdot \mu_f + N_d \cdot \mu_d) / N_g = (42 \cdot 161.71 + 35 \cdot 155.43) / 77 = 158.86, \quad (22)$$

⁴(...Fortsetzung)

wenn man die Codiervariablen γ_i^h , γ_i^w , γ_i^a und γ_i^w aufgrund ihrer Definitionsgleichungen ersetzt.

Allerdings sind die Definitionen in den Gleichungen (19) und (20) nur bei orthogonalen Versuchsplänen sinnvoll (vgl. Fußnote 3 zu den Proportionalitäten, die die Orthogonalität ausmachen). Bei nicht orthogonalen Versuchsplänen müßte bei der Herleitung der Ersatzdarstellungen in Fußnote 3 auf die Vereinfachung der Gruppengrößen verzichtet werden.

wobei die Indices f und d für "funktionale" und "dekorative" Gestaltung stehen. (Die Zahlenwerte gelten natürlich nur, wenn die bisher zugrundegelegten Mittelwerte die "wahren" sind, was jetzt wieder angenommen werden soll.) Der Mittelwert μ_u für **ungestaltete** Arbeitsbedingung ist natürlich gleich dem alten Mittelwert μ_a ; also $\mu_u = 126.86$. (Hier liegt ja keine Zusammenfassung von Gruppen vor, sondern lediglich eine Umbenennung.) Aufgrund dieser Mittelwerte können wir Effekte als Abweichung der Mittelwerte μ_g und μ_u vom Gesamtmittel μ definieren:

$$\beta_g := \mu_g - \mu = 158.86 - 150.32 = 8.54, \quad (23)$$

und

$$\beta_u := \mu_u - \mu = 126.86 - 150.32 = -23.46. \quad (24)$$

Um die beiden "gestalteten" Arbeitsbedingungen miteinander zu vergleichen, verfahren wir bei der Definition von Effekten so, als gäbe es die Gruppe mit "ungestalteter" Arbeitsbedingung nicht. Entsprechend übernimmt das Mittel μ_g (gestaltete Arbeitsbedingungen) die bisherige Rolle des Gesamtmittels μ , und die Effekte β_f und β_d sind jetzt als Abweichung des entsprechenden Mittelwerts vom Mittelwert μ_g definiert, also

$$\beta_f := \mu_f - \mu_g \quad (25)$$

bei funktionaler Gestaltung, und

$$\beta_d := \mu_d - \mu_g \quad (26)$$

bei dekoraktiver Gestaltung des Arbeitsplatzes.

Aus diesen Definitionen ergeben sich die Abhängigkeiten⁵

$$N_g \cdot \beta_g + N_u \cdot \beta_u = 0 \quad (27)$$

und

$$N_f \cdot \beta_f + N_d \cdot \beta_d = 0, \quad (28)$$

aus denen sich die Ersatzdarstellungen

⁵Herleitung von Gleichung (26):

$$\begin{aligned} N_g \cdot \beta_g + N_u \cdot \beta_u &= N_g \cdot (\mu_g - \mu) + N_u \cdot (\mu_u - \mu) \\ &= (N_g \cdot \mu_g + N_u \cdot \mu_u) - N \cdot \mu \\ &= 0 \end{aligned}$$

Die erste Zeile ergibt sich aus den Definitionen von β_g und β_u (Gleichungen (22) und (23)). Die zweite Zeile ergibt sich aus einer algebraischen Umformung und der trivialen Gleichung $N = N_g + N_u$. Für das letzte Gleichheitszeichen verwendet man die Gleichung $\mu = (N_g \cdot \mu_g + N_u \cdot \mu_u) / N$, die sich aus dem allgemeinen Gesetz über den Mittelwert zusammengefaßter Gruppen ergibt.

Gleichung (27) läßt sich ganz ähnlich beweisen.

$$b_u = \beta_g \cdot (-N_g / N_u), \quad (29)$$

und

$$\beta_d = \beta_f \cdot (-N_f / N_d) \quad (30)$$

für die Effekte β_u und β_d ergeben.⁶

Im folgenden werden Schritt für Schritt Codiervariablen für die Kontrastcodierung definiert. Das Verständnis dieser Variablen wird erleichtert, wenn man sich eine ähnliche Tabelle anlegt, wie sie für die Effektcodierung nach der Regressionsgleichung (8) angegeben war, und bei jeder Definition einer neuen Codiervariablen für alle Zellen die Werte einträgt, die sich aus der Definition ergeben. Dazu der Hinweis, daß wieder 5 Codiervariablen gebildet werden, die aber größtenteils anders bezeichnet sind als bei der Effektcodierung.

Wenn man mit den in den Gleichungen (28) und (29) angegebenen Ersatzdarstellungen für die Effekte β_d und β_u arbeitet, dann werden die Effekte β_g und β_f in einer Regressionsgleichung auftauchen - multipliziert mit Codiervariablen γ_i^g und γ_i^f , deren Werte nach folgendem Prinzip zu bestimmen sind: Für jede Person i soll die Aufnahme der Summe $\beta_g \cdot \gamma_i^g + \beta_f \cdot \gamma_i^f$ in die Regressionsgleichung bewirken, daß die Differenz $\mu_{r(i)} - \mu$ (also die Abweichung des entsprechenden Zeilenmittels $\mu_{r(i)}$ vom Gesamtmittel μ) addiert wird. Daraus ergeben sich die folgenden Überlegungen zu den erforderlichen Werten der Codiervariablen γ_i^g und γ_i^f :

- Für Personen mit "funktional" gestalteter Arbeitsbedingung ist $\mu_f - \mu$ zu addieren, und das ist gleich $\beta_g + \beta_f$ (Gleichungen (22) und (24) addieren!). Also muß für alle Personen mit "funktional" gestalteter Arbeitsbedingung $\gamma_i^g = 1$ und $\gamma_i^f = 1$ gewählt werden.
- Für Personen mit "dekorativ" gestalteter Arbeitsbedingung ist $\mu_d - \mu$ zu addieren, und das ist gleich $\beta_g + \beta_d$ (Gleichungen (22) und (25) addieren!). Also muß für alle Personen mit "dekorativ" gestalteter Arbeitsbedingung $\gamma_i^g = 1$ und $\gamma_i^f = -N_f / N_d$ gewählt werden (Ersatzdarstellung von β_d nach Gleichung (29)).
- Für Personen mit "ungestalteter" Arbeitsbedingung ist $\mu_u - \mu$ zu addieren, also β_u (Gleichung (23)). Unter Rückgriff auf die Ersatzdarstellung von β_u (Gleichung (28)) erreichen wir dies durch die Codiervariablenwerte $\gamma_i^g = -N_g / N_u$ und $\gamma_i^f = 0$ für aller Personen mit "ungestalteter" Arbeitsbedingung.

Nach dem im Anschluß an Gleichung (11) dargestellten Verfahren zur Konstruktion einer Definitionsgleichung für Codiervariablen ergeben sich aus diesen Vorgaben die Definitionen

$$\gamma_i^g := \delta_i^g - \delta_i^u \cdot N_g / N_u \quad (31)$$

und

⁶Ebenso könnte man aus den Gleichungen (26) und (27) Ersatzdarstellungen für die Effekte β_g und β_f herleiten und eine Regressionsgleichung herleiten, in der die Effekte β_u und β_d mit Codiervariablen γ_i^u und γ_i^d auftreten. Bei der Kontrastcodierung gibt es nur selten Gründe für die Entscheidung, welche Effekte man in der Regressionsgleichung auftauchen läßt und für welche man eine Ersatzdarstellung wählt.

$$\gamma_i^f := \delta_i^f - \delta_i^d \cdot N_f / N_d. \quad (32)$$

Zusammengefaßt: Bei der Kontrastcodierung für Haupteffekte kommen dieselben Methoden zur Anwendung wie bei der varianzanalytischen Effektcodierung für einen zweistufigen Faktor. Zur Bestimmung von Effekten werden die zu kontrastierenden Mittelwerte mit einem "Gesamtmittel" verglichen, das (unter Berücksichtigung der Gruppengrößen) aus den zu vergleichenden Mittelwerten gebildet wird. Wegen der Abhängigkeit der so definierten Effekte taucht jedoch nur einer von beiden unmittelbar (also mit einem γ_i -Wert multipliziert) in der Regressionsgleichung auf. Die Codiervariablen (also die γ_i -Werte) sind dann so konstruiert, daß für den anderen Effekt eine "Ersatzdarstellung" gewählt wird. Für Gruppen, die in einen Kontrast nicht eingehen (z.B. die Gruppe u beim Kontrast von μ_f und μ_d) wird die entsprechende Codiervariable auf 0 gesetzt.

Die Übereinstimmung der Kontrastcodierung von Haupteffekten mit der Effektcodierung für einen zweistufigen Faktor bedeutet auch, daß für den zweistufigen Faktor Geschlecht die Definitionen des Effekts β_w und der Codiervariablen γ_i^w unmittelbar aus der varianzanalytischen Effektcodierung übernommen werden können.

Wie bei den anderen Codierungsverfahren kommen auch bei der Kontrastcodierung Wechselwirkungseffekte als β -Gewichte in der Regressionsgleichung vor. Dabei gilt das folgende Prinzip: Jeder Kontrast im Zeilenfaktor kann mit jedem Kontrast im Spaltenfaktor in Wechselwirkung treten. Zu jeder solchen Wechselwirkung zweier Kontraste gibt es jeweils vier Wechselwirkungseffekte, von denen aber nur einer unmittelbar in die Regressionsgleichung eingeht, während die anderen in Form einer "Ersatzdarstellung" berücksichtigt werden.

Für unsere Beispieluntersuchung: Im Zeilenfaktor (Arbeitsbedingung) gibt es die Kontraste "gestaltet vs. ungestaltet" und "funktional vs. dekorativ", und im Spaltenfaktor (Geschlecht) lediglich den Kontrast "weiblich vs. männlich". Also kann jeder der beiden Kontraste im Zeilenfaktor mit dem einzigen Kontrast im Spaltenfaktor in Wechselwirkung treten.

Für die Wechselwirkung der Kontraste "gestaltet vs. ungestaltet" und "weiblich vs. männlich" lassen sich die folgenden Wechselwirkungseffekte bilden:

$$\beta_{gw} = \mu_{gw} - (\mu + \beta_g + \beta_w) = 157.09 - (150.32 + 8.54 + (-0.99)) = -0.78, \quad (33.gw)$$

$$\beta_{gm} = \mu_{gm} - (\mu + \beta_g + \beta_m) = 160.18 - (150.32 + 8.54 + 0.75) = 0.57, \quad (32.gm)$$

$$\beta_{uw} = \mu_{uw} - (\mu + \beta_u + \beta_w) = 128 - (150.32 + (-23.46) + (-0.99)) = 2.13, \quad (32.uw)$$

und

$$\beta_{um} = \mu_{um} - (\mu + \beta_u + \beta_m) = 126 - (150.32 + (-23.46) + 0.75) = -1.61, \quad (32.um)$$

wobei der Ausdruck in Klammern jeweils der "fiktive" Mittelwert ist, der aufgrund des Gesamtmittels und der Haupteffekte zu erwarten wäre. Die Mittelwerte μ_{gw} und μ_{gm} sind natürlich unter Berücksichtigung der Gruppengrößen bestimmt:

$$\mu_{gw} = (N_{fw} \cdot \mu_{fw} + N_{dw} \cdot \mu_{dw}) / N_{gw} = (18 \cdot 148 + 15 \cdot 168) / 33 = 157.09 \quad (34)$$

und

$$\mu_{gm} = (N_{fm} \cdot \mu_{fm} + N_{dm} \cdot \mu_{dm}) / N_{gm} = (24 \cdot 172 + 20 \cdot 146) / 44 = 160.18. \quad (35)$$

In Worten ausgedrückt, besagt z.B. der Effekt β_{gw} : Frauen unter gestalteten Arbeitsbedingungen erzielen durchschnittlich 0.78 Punkte weniger, als es aufgrund des Gesamtmittels μ sowie der Haupteffekte β_g und β_w zu erwarten wäre.

Zwischen diesen Wechselwirkungseffekten bestehen wieder Abhängigkeiten, die sich in einer Kettengleichung als

$$\begin{aligned} N_{gw} \cdot \beta_{gw} + N_{gm} \cdot \beta_{gm} &= N_{uw} \cdot \beta_{uw} + N_{um} \cdot \beta_{um} = N_{gw} \cdot \beta_{gw} + N_{uw} \cdot \beta_{uw} = N_{gm} \cdot \beta_{gm} + N_{um} \\ &\cdot \beta_{um} \\ &= 0 \end{aligned} \quad (36)$$

zusammenfassen lassen.⁷ Daraus läßt sich herleiten: Definieren wir eine Codiervariable γ_i^{gw} als

$$\gamma_i^{gw} := \gamma_i^g \cdot \gamma_i^w, \quad (37)$$

dann ergibt das Produkt $\beta_{gw} \cdot \gamma_i^{gw}$ für jede Zelle eine Darstellung bzw. Ersatzdarstellung des passenden Wechselwirkungseffekts β_{gw} , β_{gm} , β_{uw} oder β_{um} .⁸

Für die Wechselwirkung der Kontraste "funktional vs. dekorativ" und "weiblich vs. männlich" ist zu beachten, daß in den Kontrast "funktional vs. dekorativ" nur die Personen mit gestalteter Arbeitsbedingung eingehen. Daher übernimmt das Mittel μ_g dieser Personen die Funktion eines Gesamtmittels.⁹ Dementsprechend lassen sich die folgenden Wechselwirkungseffekte bilden:

$$\beta_{fw} = \mu_{fw} - (\mu_g + \beta_f + \beta_w), \quad (38.fw)$$

$$\beta_{fm} = \mu_{fm} - (\mu_g + \beta_f + \beta_m), \quad (37.fm)$$

$$\beta_{dw} = \mu_{dw} - (\mu_g + \beta_d + \beta_w), \quad (37.dw)$$

und

⁷Aufbauprinzip von Gleichung (35): Zuerst die beiden Wechselwirkungseffekte mit g, dann die mit u, dann die mit w, dann die mit m. Nach Multiplikation mit den jeweiligen Gruppengrößen ergibt sich jedesmal eine Summe 0.

⁸Übungsaufgabe: Erstellen Sie aufgrund der Definition in Gleichung (36) eine Wertetabelle für die Codiervariable γ_i^{gw} für alle Zellen und überprüfen Sie (unter Verwendung der in Gleichung (35) zusammengefaßten Abhängigkeiten), daß das Produkt $\beta_{gw} \cdot \gamma_i^{gw}$ für jede Zelle den passenden Wechselwirkungseffekt ergibt - entweder direkt oder als Ersatzdarstellung.

⁹Allgemein: Bei der Wechselwirkung zweier Kontraste wird die Funktion eines Gesamtmittels vom Mittelwert derjenigen Personen übernommen, die in beide Kontraste eingehen.

$$\beta_{dm} = \mu_{dm} - (\mu_g + \beta_d + \beta_m). \quad (37.dm)$$

Aufgrund der Abhängigkeiten

$$\begin{aligned} N_{fw} \cdot \beta_{fw} + N_{fm} \cdot \beta_{fm} &= N_{dw} \cdot \beta_{dw} + N_{dm} \cdot \beta_{dm} = N_{fw} \cdot \beta_{fw} + N_{dw} \cdot \beta_{dw} = N_{fm} \cdot \beta_{fm} + N_{dm} \cdot \beta_{dm} \\ &= 0 \end{aligned} \quad (39)$$

läßt sich zeigen: Definieren wir eine Codiervariable γ_i^{fw} als

$$\gamma_i^{fw} := \gamma_i^f \cdot \gamma_i^w, \quad (40)$$

dann ergibt das Produkt $\beta_{fw} \cdot \gamma_i^{fw}$ für jede Zelle eine Darstellung bzw. Ersatzdarstellung des passenden Wechselwirkungseffekts β_{fw} , β_{fm} , β_{dw} oder β_{dm} .

Damit sind alle Effekte und Codiervariablen definiert, die wir für eine Regressionsgleichung brauchen. Diese lautet jetzt:

$$\hat{X}_i = \beta_0 + \beta_g \cdot \gamma_i^g + \beta_f \cdot \gamma_i^f + \beta_w \cdot \gamma_i^w + \beta_{gw} \cdot \gamma_i^{gw} + \beta_{fw} \cdot \gamma_i^{fw}. \quad (41)$$

Was unterscheidet diese Regressionsgleichung von denjenigen für die Dummycodierung (Gleichung (7)) und für die varianzanalytische Effektcodierung (Gleichung (8))? Die Definition der Effekte geht davon aus, daß es sinnvoll ist, die drei Arbeitsbedingungen zunächst in "gestaltete" und "ungestaltete" zu unterteilen, und dann die "gestalteten" noch einmal in "funktional" und "dekorativ" gestaltete. Entsprechend dieser Einteilung enthält dann die Regressionsgleichung (40) insgesamt 5 Effekte (die β -Werte), die jeweils einem bestimmten Kontrast von Arbeitsbedingungen bzw. dem Kontrast der Geschlechter "männlich" und "weiblich" bzw. der Wechselwirkung eines Arbeitsbedingungskontrasts mit dem Geschlecht entsprechen.

Schlußbemerkungen

In der vorangehenden Darstellung wurde gezeigt, welche verschiedenen Möglichkeiten es gibt, in varianzanalytischen Versuchsplänen die Gruppenzugehörigkeit durch Codiervariablen zu erfassen. Diese Codiervariablen dienen dann als Prädiktoren in einer Regressionsgleichung, deren β -Coeffizienten Haupteffekte und Wechselwirkungseffekte sind. Diese Effekte sind aber je nach Codiermethode unterschiedlich definiert.

In vielen Lehrbüchern wird zwar (mehr oder weniger "kochbuchmäßig") mit Gleichungen oder verbalen Regeln angegeben, wie man Werte für Codiervariablen bestimmt und wie man daraus eine Regressionsgleichung zusammenstellt; es wird aber versäumt, auf die unterschiedliche Bedeutung der β -Größen bei Dummy-, Effekt- und Kontrastcodierung hinzuweisen.

In dieser Situation mag man sich fragen, woher dann die Definitionen der mit den verschiedenen Codiermethoden verbundenen unterschiedlichen Effektbegriffe kommen. Das soll am Beispiel der Dummycodierung erläutert werden. Aus den Regeln für diese Codiermethode geht hervor, daß es genau eine Zelle gibt, bei der alle in der Regressionsgleichung auftretenden Dummyvariablen null sind. Daraus und aus den Regeln für die Konstruktion der übrigen Dummyvariablen läßt sich dann mathematisch zurückrechnen, daß die Größen β_0 , β_h , β_a , β_w , β_{hw}

und β_{aw} die im Abschnitt zur Dummycodierung näher erläuterte Bedeutung als Referenzgruppen-Mittel und Referenzgruppen-bezogene Effekte haben.¹⁰ Für die Effektcodierung und die Kontrastcodierung läßt sich die im vorangehenden Text dargestellte Bedeutung der β -Gewichte ähnlich aus den entsprechenden Codierregeln rekonstruieren.

Insgesamt erscheint es fragwürdig, lediglich Regeln für die Codierung der Gruppenzugehörigkeit darzustellen, so daß mühsam rekonstruiert muß, was die β -Gewichte in der Regressionsgleichung bedeuten, die der Computer in Sekundenbruchteilen ausgibt, sobald man eine Codierung der Gruppenzugehörigkeit angegeben hat. Für das Verständnis hilfreicher ist es wohl, wie in der vorangehenden Darstellung von möglichen Bedeutungen des Effektbegriffs - also der β -Gewichte - auszugehen. Dann kann man für eine anliegende Fragestellung die am ehesten passende Bedeutung auswählen und in Abhängigkeit von dieser erwünschten Bedeutung die passenden Codiervariablen definieren.

Ausdrücklich ist noch einmal darauf hinzuweisen, daß die β -Gewichte aufgrund der "wahren" Mittelwerte definiert sind. In empirischen Untersuchungen sind diese natürlich in aller Regel unbekannt. Schätzwerte lassen sich aber bilden, indem man anstelle der wahren Mittelwerte die entsprechenden Stichprobendurchschnitte in die Definitionsformeln einsetzt. Mit Signifikanztests, die hier nicht dargestellt werden sollen, kann dann geprüft werden, welche β -Gewichte (also welche Effekte) signifikant von 0 verschieden sind.

Sehr problematisch ist es jedoch, erst aufgrund der Daten zu bestimmen, welche Codierung der Gruppenzugehörigkeit man wählen könnte, um zu signifikanten Ergebnissen zu kommen. Die

¹⁰In der Darstellung von Regeln für die Dummycodierung fehlt sogar meistens die Bezeichnung "Referenzgruppe". Diese Bezeichnung ist eine verbale Umschreibung der Funktion dieser Gruppe, die sich aus der "rückgerechneten" Bedeutung der Effekte ergibt.

Dieser Rückrechnung liegen die folgenden Überlegungen zugrunde. Aufgrund der Least-squares-Eigenschaften von Mittelwerten muß der Schätzwert \hat{X}_i für alle Personen das Zellenmittel sein. Für die Mitglieder derjenigen Gruppe, bei der alle Dummyvariablen in Regressionsgleichung (7) null sind, ergibt sich dann unmittelbar, daß β_0 der "wahre" Mittelwert dieser Gruppe ist. Die Bezeichnung dieser Gruppe als Referenzgruppe und die entsprechende Interpretation der Effekte ergibt sich ebenfalls, indem man sich ansieht, welche Werte der Codiervariablen sich aus den Codierregeln für verschiedene andere Zellen ergeben und was daraus für die β -Gewichte folgt. Für unsere Beispieluntersuchung folgt z.B. aus den Codierregeln: Wenn für die Zelle nm alle Codiervariablen null sind, dann ist für die Zelle hm die Codiervariable für den Effekt β_h gleich 1, und für alle anderen Codiervariablen ergibt sich bei dieser Zelle ein Wert von 0. Da für diese Zelle $\hat{X}_i = \mu_{hm}$ gelten muß, ergibt sich aus Regressionsgleichung (7) zunächst $\mu_{hm} = \beta_0 + \beta_h$, und damit $\beta_h = \mu_{hm} - \mu_{nm}$ aufgrund der bereits gefundenen Bedeutung von β_0 . Für die Zellen am und nw gilt entsprechend: Sind bei Zelle nm alle Codiervariablen null, dann ist nach den Regeln für die Dummycodierung für Zelle am nur die Codiervariable für β_a eins, und für Zelle nw nur die Codiervariable für β_w , und alle anderen Codiervariablen sind für die Zellen am und nm 0. Also ist $\mu_{am} = \beta_0 + \beta_a$ und $\mu_{nw} = \beta_0 + \beta_w$, und damit $\beta_a = \mu_{am} - \mu_{nm}$ und $\beta_w = \mu_{nw} - \mu_{nm}$. Mit ähnlichen Überlegungen für weitere Gruppen kommt man dann zu entsprechenden Feststellungen über die weiteren β -Gewichte in Regressionsgleichung (7). Die Bezeichnung "Referenzgruppe" für die Gruppe, bei der alle Codiervariablen 0 sind, faßt dann rückblickend ihre Bedeutung für die Definition der Haupteffekte zusammen.

Leichtigkeit, mit der man das mit entsprechenden Computerprogrammen ausprobieren kann, ist zwar sehr verführerisch; aber die Wahrscheinlichkeit, daß bei irgendeiner nachträglich ausgesuchten Codierung ein β -Gewicht, das in Wirklichkeit 0 sein müßte, trotzdem signifikant von 0 verschieden wird, ist größer als das "nominelle" α -Niveau des entsprechenden Signifikanztests.

Glücklicherweise gibt es aber eine Lösung dieses Problems: Wenn man erst aufgrund der Daten ("post hoc") bestimmt, welche Codierung der Gruppenzugehörigkeit verwendet werden soll, dann kann man dies durch strengere Anforderungen bei der Signifikanzprüfung ausgleichen. Aber das ist eine andere Geschichte und soll ein andermal erzählt werden.