Das bedeutet nun, daß die Korrelationen von drei Variablen nicht beliebige Kombinationen von Zahlen im Bereich von -1 bis $+1$ bilden können. Ist beispielsweise $r_{xz} = 0.70$ und $r_{yz} = 0.30$, dann kann man diese Zahlen in die Formel einsetzen und erhält

$$-0.47 \leq r_{xy} \leq 0.89 .$$

Daß solche Grenzen für die Korrelation r_{xy} bestehen müssen, ist auch intuitiv nachvollziehbar. Bestünde eine hohe negative Korrelation r_{xy}, dann könnten nicht beide Variablen positiv mit z korrelieren. (Extremfall: Bei $r_{xy} = -1$ könnte y als lineare Umpolung von x betrachtet werden, und dann wäre $r_{xz} = -r_{yz}$.) Bestünde andererseits eine sehr hohe Korrelation r_{xy} von nahezu 1, dann müßten die Korrelationen beider Variablen mit z nahezu gleich sein. (Extremfall: Bei $r_{xy} = 1$ wäre y eine Lineartransformation von x und damit $r_{xz} = r_{yz}$.) Die obige Ungleichung erlaubt es, die Grenzen für die Korrelation r_{xy}, die sich aus solchen Überlegungen ergeben, genau anzugeben.

>>

b) Multiple Regression und Korrelation

Wir wollen annehmen, der Schulerfolg eines Schülers solle vorhergesagt werden, und wir hätten seine Testwerte aus drei verschiedenen Tests zur Verfügung: Aus einem Test des räumlichen Vorstellungsvermögens (X_i), einem Test der verbalen Intelligenz (Y_i) und einem Test des rechnerischen Denkens (Z_i). Wir könnten uns nun überlegen, daß eine Schätzgleichung der folgenden Art sinnvoll wäre:

$$\hat{S}_i = 0.5 \cdot X_i + 1.5 \cdot Y_i + 1.2 \cdot Z_i - 70$$

Dabei steht $\hat{S}_i$ für "geschätzter Schulerfolg". Geschätzt werden könnte z.B. die Gesamtpunktzahl S_i, die der Schüler in den verschiedenen Jahresabschlußtests des ersten Jahres erreicht. Die Zahlen sind nicht beliebig gewählt; die ersten drei haben die Funktion von Gewichtszahlen.[153] Das räumliche Vorstellungsvermögen (X_i) spielt in der Schule eine nur geringe Rolle und erhält eine niedrige Gewichtszahl; die verbale Intelligenz (Y_i) spielt eine große Rolle und erhält daher eine hohe Gewichtszahl; die hohe Gewichtszahl für das Rechnen (Z_i) könnte durch den Schultyp (z.B. Handelsschule oder naturwissenschaftliches Gymnasium) bedingt sein. Die Subtraktion der Konstanten 70 hat mit den Durchschnitten zu tun: Alle Tests haben ein arithmetisches Mittel von l00; die Gesamtpunktzahl hat ein arithmetisches Mittel von 250. Die subtrahierte Konstante 70 ist so gewählt, daß für einen Schüler, der in allen 3 Tests den durchschnittlichen Meßwert 100 erzielt, auch eine durchschnittliche Schulleistung ($\hat{S}_i = 250$) herauskommt.

Eine "gewichtete Summe", bei der verschiedene Werte mit einer "Gewichtszahl" multipliziert

---

[153] vgl. dazu auch S. @

werden und die Summe dieser Produkte gebildet wird, nennt man auch eine Linearkombination dieser Werte. Die Gewichtszahlen können dabei auch ein negatives Vorzeichen haben. (Eine solche negative Gewichtszahl könnte z.B. auftreten, wenn einer der Tests so gepolt wäre, daß eine hohe Punktzahl eine niedrige Leistung bedeutet.) Ähnlich, wie es bereits für den Begriff der Lineartransformation dargestellt wurde (vgl. S. 59), gibt es auch verschiedene Definitionen der Linearkombination. Sie unterscheiden sich darin, daß nach der einen Definition zusätzlich zu der gewichteten Summierung auch noch eine Konstante addiert oder subtrahiert werden kann (wie z.B. bei uns die 70), während andere Definitionen das nicht zulassen. Die letzteren Definitionen sprechen also nur dann von einer Linearkombination, wenn die "Verrechnungsformel" nur aus einer gewichteten Summation ohne Addition oder Subtraktion einer Konstanten besteht. Wir wollen uns hier der ersten Definition anschließen und beispielsweise die rechte Seite der obigen Gleichung für die Schätzung des Schulerfolgs ebenfalls als Linearkombination der Testwerte bezeichnen.

Wir haben die Gewichtszahlen in unserer Schätzgleichung nach Plausibilitätsgesichtspunkten festgesetzt. Wir hätten natürlich auch andere Gewichtszahlen wählen können, und wären wieder zu einer Linearkombination gekommen. Fast jede dieser Schätzgleichungen würde zu verschieden hohen Korrelationen zwischen Schätzwerten und tatsächlichen s_i Werten, also zu verschiedenen Vorhersagegenauigkeiten führen.

Wir könnten nun sagen: Wir wählen diejenige Linearkombination als Schätzgleichung, bei der die Korrelation zwischen Schätzwerten ($\hat{s}_i$) und tatsächlichen s_i-Werten am höchsten ist. Man könnte aber auch sagen: Wir wählen diejenige Linearkombination, bei der die Schätzfehlervarianz (also das arithmetische Mittel der quadrierten Differenzen s_i - $\hat{s}_i$) am niedrigsten ist. Niedrige Schätzfehlervarianz bedeutet dann ja auch niedrigen Standardschätzfehler und hohen Schätzungseffekt. Es läßt sich mathematisch beweisen, daß diejenige Linearkombination, die zur niedrigsten Schätzfehlervarianz führt, auch die höchstmögliche Korrelation eines als Linearkombination errechneten Schätzwerts $\hat{s}_i$ mit den tatsächlichen Werten hat.[154] Wir kommen daher nicht in einen Konflikt; wir würden uns vielmehr für diejenige Linearkombination (also diejenigen Gewichtszahlen) entscheiden, bei der die Schätzfehlervarianz am geringsten ist.

Diese Linearkombination, die die kleinste Schätzfehlervarianz mit sich bringt, nennt man die multiple lineare Regressionsgleichung oder auch einfach die multiple Regressionsgleichung.

[154] Für Spezialisten: Dagegen gilt nicht das Gegenteil: Haben wir die Schätzfehlervarianz durch geeignete Wahl der Gewichtszahlen zu einem Minimum gemacht, dann könnten wir die Schätzwerte einer Lineartransformation unterziehen. Die Korrelation zwischen Schätzwerten und tatsächlichen s_i-Werten würde sich dadurch nicht ändern. Warum? (vgl. S. @) Die Schätzfehlervarianz würde dagegen steigen. Dieselben Schätzwerte könnte man aber auch direkt als Linearkombination erhalten. Konsequenz: Haben wir die maximal mögliche Korrelation von Schätzwerten und tatsächlichen Werten, dann haben wir noch nicht automatisch die niedrigste Schätzfehlervarianz.

Offenbar besteht eine Parallele zu der in Abschnitt @3a abgeleiteten linearen Regressionsgleichung. Auch für sie galt ja, daß jede andere lineare Schätzgleichung zu einer höheren Schätzfehlervarianz führt.

Um die Überlegungen zu verallgemeinern, übertragen wir zwei Begriffe, die bereits aus der einfachen Regression bekannt sind, auf die Schätzung oder Vorhersage einer Variable aus mehreren anderen: Die zu schätzende Variable nennen wir das *Kriterium*; die Variablen, aufgrund deren geschätzt (vorhergesagt) wird, nennen wir die *Prädiktoren*.

Was war in unserem Beispiel Kriterium, und was waren die Prädiktoren?

Kriterium war s_i (die Gesamtzahl der Punkte aus den Jahresabschlußtests). Prädiktoren waren die drei Testwerte x_i, y_i und z_i. Mit diesen Begriffen können wir die wichtigsten Begriffe und Tatsachen zu multipler Korrelation und Regression folgendermaßen zusammenfassen:

Die lineare multiple Regressionsgleichung ist diejenige Linearkombination der Prädiktoren, bei der die Schätzfehlervarianz (also das arithmetische Mittel der quadrierten Differenzen von geschätztem und tatsächlichem Kriteriumswert) am geringsten ist. Diese spezielle Linearkombination hat auch die höchste Korrelation von Schätzwert und tatsächlichem Kriteriumswert, die sich mit einer Linearkombination dieser Prädiktoren erzielen läßt. Man nennt diese höchstmögliche Korrelation auch die multiple Korrelation des Kriteriums mit den Prädiktoren.

Man könnte bei der multiplen Regression natürlich auch von einer multiplen Determination sprechen: Das Kriterium wird als statistisch (nicht unbedingt kausal) determiniert durch die Prädiktoren aufgefaßt. Das Quadrat der multiplen Korrelation kann wieder als Determinationskoeffizient aufgefaßt werden: Es gibt an, welcher relative Anteil der Varianz (also der Unterschiede) der Kriteriumswerte als determiniert durch die Unterschiede in den Prädiktoren aufgefaßt werden kann. Auch die Schätzfehlervarianz, den Standardschätzfehler und den Schätzungseffekt, die wir bei Verwendung der multiplen Korrelation in Kauf nehmen, können wir angeben, indem wir die multiple Korrelation an Stelle der Produkt-Moment-Korrelation einsetzen.

Natürlich gibt es auch Formeln für die Berechnung der multiplen Korrelation und der Gewichtszahlen in der multiplen Regression. Dies Formeln werden aber relativ komplex, wenn mehr als zwei Prädiktoren vorliegen.[155] Daher berechnet man die Gewichtszahlen und die multiple Korrelation in aller Regel mit dem Computer.

[155] Recht einfache Formeln für die Gewichtszahlen in der multiplen Regression kann man allerdings mit Hilfe der sogenannten Matrix-Algebra aufstellen. Dabei handelt es sich um Verfahren, bei denen nicht nur einzelne Zahlen, sondern ganze Tabellen von Zahlen - sog. Matrizen - zueinander addiert oder miteinander multipliziert werden. Eine Behandlung der multiplen Regression und Korrelation mit diesem Ansatz ist Gegenstand von Lehrveranstaltungen im Hauptstudium zum Thema "Multivariate Verfahren".

<<

Relativ überschaubare Formeln für Gewichtszahlen und multiple Korrelation entstehen dagegen, wenn man nur zwei Prädiktoren hat. Aus diesen Formeln kann man bereits einige Grundprinzipien herleiten, die dann auch auf Situationen mit mehr als zwei Prädiktoren übertragbar sind.

Ausgangspunkt der folgenden Überlegungen soll eine Feststellung sein, die sich aus der Abbildung @ zum Schätzungseffekt (S. ?) ergab: Mit einem einzelnen Prädiktor erhalten wir bei einer Vorhersagevalidität von 0.70 nur einen Schätzungseffekt von knapp 29%, und für einen Schätzungseffekt von 50% wäre eine Vorhersagevalidität von 0.87 erforderlich. In der Praxis gilt aber schon die Konstruktion eines Tests mit einer Vorhersagevalidität um 0.70 als großer Erfolg. Es soll nun gezeigt werden, daß und wie sich deutlich bessere Schätzungseffekte erzielen lassen, wenn man mehrere Prädiktoren mit der Technik der multiplen Regression kombiniert. Es wird sich zeigen, daß es dabei auf einige Prinzipien der Zusammenstellung eines guten Satzes von Prädiktoren ankommt, und diese Prinzipien lassen sich aus Formeln für die Gewichtszahlen und für die multiple Regression herleiten.

Überschaubar werden diese Formeln, wenn wir - ähnlich wie beim ersten Zugang zur linearen Regression mit nur einem Prädiktor in Abschnitt @A.III.s.a.(iv) - davon ausgehen, daß Prädiktoren und Kriterium als z-Werte vorliegen. Außerdem ist es sinnvoll, zur Kennzeichnung der verschiedenen Prädiktoren nicht Buchstaben (wie x und y) zu verwenden, sondern die Zahlen 1 und 2 (für "1. Prädiktor" und "2. Prädiktor"), während das Kriterium als c bezeichnet wird. Dann sind r_{1c} und r_{2c} die Korrelationen der beiden Prädiktoren mit dem Kriterium, und die Korrelation der Prädiktoren untereinander ist r_{12} (was natürlich als "r eins zwei" und nicht als "r zwölf" zu lesen ist). Die z-Werte von Person i in den beiden Prädiktoren und im Kriterium werden als z_{i1}, z_{i2} und z_{ic} bezeichnet, und $\hat{z}_{ic}$ ist der Schätzwert für z_{ic}, der sich aus x_{i1} und x_{i2} ergibt. Schließlich werden die Gewichtszahlen der beiden Prädiktoren als b_1 und b_2 bezeichnet, und die multiple Korrelation als R. Eine additive Konstante a brauchen wir nicht einzuführen; denn genau wie bei der Regression mit nur einem Prädiktor wäre die günstigste Konstante a ohnehin 0, wenn Prädiktoren und Kriterium als z-Werte vorliegen.

Damit läßt sich das Problem folgendermaßen formulieren: Gesucht sind Gewichtszahlen b_1 und b_2 für die Schätzung

$$\hat{z}_{ic} = b_1 \cdot z_{i1} + b_2 \cdot z_{i2},$$

und zwar diejenigen Gewichtszahlen, bei denen die Summe

$$\sum_i (z_{ic} - \hat{z}_{ic})^2$$

möglichst klein wird. Die nach diesem Kriterium (also dem Prinzip der kleinsten Quadrate) günstigsten Gewichtszahlen sind

$$b_1 = \frac{r_{1c} - r_{12} \cdot r_{2c}}{1 - r_{12}^2}$$

und

$$b_2 = \frac{r_{2c} - r_{12} \cdot r_{1c}}{1 - r_{12}^2},$$

und für die multiple Korrelation gilt

$$R = \sqrt{b_1 \cdot r_{1c} + b_2 \cdot r_{2c}} = \sqrt{\frac{{r_{1c}}^2 + {r_{2c}}^2 - 2 \cdot r_{1c} \cdot r_{2c} \cdot r_{12}}{1 - r_{12}^2}}.$$

Dabei gilt das zweite Gleichheitszeichen auch dann noch, wenn nicht mehr - wie für die b-Gewichte vorausgesetzt - Kriterium und Prädiktoren als z-Werte gegeben sind.[156]

Bevor aus diesen Formeln Schlußfolgerungen gezogen werden, ist kurz auf einen Spezialfall einzugehen: Ist $r_{12} = 1$ oder $r_{12} = -1$, dann gilt für die Nenner in den Formeln für die Gewichtszahlen b_1 und b_2 die Gleichung $1 - {r_{12}}^2 = 0$. Der Spezialfall zweier perfekt korrelierter Prädiktoren ist aber ohnehin uninteressant[157], und daher sei ein für allemal vereinbart, daß in den

---

[156] Wer versiert in Differentialrechnung ist, kann die Formeln für die Gewichtszahlen b_1 und b_2 folgendermaßen herleiten: Die Summe der Abweichungsquadrate, die durch geeignete Wahl von b_1 und b_2 möglichst klein gemacht werden soll, läßt sich durch die Gleichung

$$\sum_i (z_{ic} - \hat{z}_{ic})^2 = \sum_i (z_{ic} - b_1 \cdot z_{i1} - b_2 \cdot z_{i2})^2$$

$$= \sum_i ({z_{ic}}^2 + {b_1}^2 \cdot {z_{i1}}^2 + {b_2}^2 \cdot {z_{i2}}^2 - 2 \cdot b_1 \cdot z_{ic} \cdot z_{i1} - 2 \cdot b_2 \cdot z_{ic} \cdot z_{i2} - 2 \cdot b_1 \cdot b_2 \cdot z_{ic} \cdot z_{i1})$$

$$= N \cdot (1 + {b_1}^2 + {b_2}^2 - 2 \cdot b_1 \cdot r_{1c} - 2 \cdot b_2 \cdot r_{2c} - 2 \cdot b_1 \cdot b_2 \cdot r_{12}).$$

angeben. Zum Beweis des zweiten Gleichheitszeichen verwendet man die Gleichung

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2ac,$$

deren allgemeine Gültigkeit für beliebige Zahlen a, b und c sich leicht herleiten läßt. Das letzte Gleichheitszeichen in der obigen Gleichung beruht auf folgenden Überlegungen:

- Da das arithmetische Mittel von quadrierten z-Werten in einer Stichprobe 1 ist, ist
  $\sum_i {z_{ic}}^2 = \sum_i {z_{i1}}^2 = \sum_i {z_{i2}}^2 = N.$
- Da die Produkt-Moment-Korrelation zweier Variablen das arithmetische Mittel der Produkte der entsprechenden z-Werte ist, gelten die Gleichungen
  $\sum_i z_{ic} \cdot z_{i1} = N \cdot r_{1c}$,
  $\sum_i z_{ic} \cdot z_{i2} = N \cdot r_{2c}$
  und
  $\sum_i z_{i1} \cdot z_{i2} = N \cdot r_{12}$.

Um die in der letzten Zeile der langen Summe dargestellte Summe der Abweichungsquadrate zu minimieren, kann man die Ableitungen dieser letzten Zeile nach b_1 und nach b_2 bilden und diese Ableitungen 0 setzen, um diejenigen Werte von b_1 und b_2 zu erhalten, bei denen die Summe der Abweichungsquadrate minimal wird.

Für eine Herleitung der Formel für die multiple Korrelation kann man die in Abschnitt A.III.5 dargestellten Formeln für die Varianz und die Kovarianz von Summen verwenden.

[157] Genauer: Bei $r_{12} = 1$ oder $r_{12} = -1$ sind in den Gleichungen für b_1 und b_2 auch die Zähler 0. Ist nämlich $r_{12} = 1$, dann ergibt sich der zweite Prädiktor durch eine Lineartransformation des ersten ohne Umpolung, und damit ist $r_{2c} = r_{1c}$. Bei $r_{12} = -1$ kann dagegen der zweite Prädiktor

folgenden Überlegungen bei allen Bezugnahmen auf die obigen Formeln für die Gewichtszahlen b_1 und b_2 und für die multiple Korrelation R stichschweigend vorausgesetzt wird, daß r_{12} weder 1 noch -1 ist.

Was sich aus diesen allgemeinen Formeln ergibt, soll nun für verschiedene Fälle geklärt werden. Am einfachsten ist der Fall zweier unkorrelierter Prädiktoren, also $r_{12} = 0$. Dann ergibt sich aus den obigen Formeln für die Gewichtszahlen $b_1 = r_{1c}$ und $b_2 = r_{2c}$, und die Konsequenz für die multiple Korrelation wird am deutlichsten, wenn man deren Quadrat, also den entsprechenden Determinationskoeffizienten angibt. Für diesen ergibt sich $R^2 = r_{1c}{}^2 + r_{2c}{}^2$. Nun kann man die Quadrate $r_{1c}{}^2$ und $r_{2c}{}^2$ ebenfalls als Determinationskoeffizienten auffassen, und dann lassen sich die Ergebnisse in einem Satz formulieren, der übrigens auch bei mehr als zwei Prädiktoren gilt:

---

als Ergebnis einer linearen Umpolung des ersten betrachtet werden; also ist $r_{2c} = -r_{1c}$. In beiden Fällen ergibt sich also für b_1 und b_2 der Bruch 0 / 0, dessen Wert nicht definiert ist. Tatsächlich gibt es in dieser Situation keine Linearkombination der Prädiktoren, die im Sinne des Prinzips der kleinsten Quadrate besser als alle übrigen ist. Beweis: Bei $r_{12} = 1$ ist ja $z_{i1} = z_{i2}$ für jede Person i. Für gegebene Gewichtszahlen b_1 und b_2 könnten wir nun mit einer beliebigen Zahl x durch die Gleichungen $b'1 := b_1 + x$ und $b'_2 := b_2 - x$ neue Gewichtszahlen definieren, mit denen dann für jede Person i die Gleichung

$$\begin{aligned} b'_1 \cdot z_{i1} + b'_2 \cdot z_{i2} &= b_1 \cdot z_{i1} + x \cdot z_{i1} + b_2 \cdot z_{i2} - x \cdot z_{i2} = b_1 \cdot z_{i1} + b_2 \cdot z_{i2} + x \cdot (z_{i1} - z_{i2)} \\ &= b_1 \cdot z_{i1} + b_2 \cdot z_{i2} \end{aligned}$$

gelten würde. Das heißt aber: Bei $r_{12} = 1$ gibt es zu jedem Paar von Gewichtszahlen b_1 und b_2 unendlich viele Paare von Gewichtszahlen b'_1 und b'_2, die bei Verwendung in einer linearen Regressionsgleichung zum gleichen Schätzwert $\hat{z}_{ic}$ führen würden wie die Gewichtszahlen b_1 und b_2. Dann ist aber auch die Summe der Abweichungsquadrate $\sum_i (z_i - \hat{z}_i)^2$ bei beiden Paaren von Gewichtszahlen gleich, und damit kann es keine Gewichtszahlen b_1 und b_2 geben, die im Sinne des Prinzips der kleinsten Quadrate besser als *alle* übrigen Paare von Gewichtszahlen sind.

Bei $r_{12} = -1$ gilt entsprechend $z_{i2} = -z_{i1}$ für jede Person i, und damit läßt sich die obige Überlegung mit Gewichtszahlen $b'_1 := b_1 + x$ und $b'_2 := b_2 + x$ wiederholen.

Übrigens gilt bei mehr als zwei Prädiktoren etwas ganz Ähnliches, wenn einer der Prädiktoren sich als Linearkombination der übrigen darstellen läßt. Gilt z.B. bei drei Prädiktoren für jede Person i die Gleichung $z_{i1} = 0.6 \cdot z_{i2} + 0.8 \cdot z_{i2}$, dann lassen sich zu gegebenen Gewichtszahlen b_1, b_2 und b_3 gleichwertige andere Gewichtszahlen mit den Gleichungen $b'_1 := b_1 + x$ sowie $b'_2 = b_2 - 0.6 \cdot x$ und $b'_3 := b_3 - 0.8 \cdot x$ definieren, wobei in allen drei Gleichungen dieselbe Zahl x zu verwenden ist. (Übungsaufgabe: Zeigen Sie, daß in dieser Situation beide Sätze von Gewichtszahlen für jede Person i zum gleichen Schätzwert $\hat{z}_{ic}$ führen würden.)

Faktisch geht man mit solchen Situationen so um, daß man von den Prädiktoren, die sich als Linearkombination der übrigen darstellen lassen, einen streicht, und das wiederholt man so lange, bis keiner der verbleibenden Prädiktoren sich als Linearkombination der übrigen darstellen läßt. Für die verbleibenden Prädiktoren läßt man dann vom Computer eine optimale lineare Regressionsgleichung bestimmen, und dann kann man jede mit dieser Lösung gleichwertige Linearkombination der ursprünglichen Prädiktoren als lineare Regressionsgleichung verwenden.

Sind Kriterium und Prädiktoren als z-Werte gegeben und sind die Prädiktoren unkorreliert, dann ist die Gewichtszahl jedes Prädiktors in der multiplen Regression seine Korrelation mit dem Kriterium, und die durch die multiple Regression determinierte Varianz des Kriteriums ist die Summe der durch die Prädiktoren einzeln determinierten Varianzanteile.

Was das konkret bedeutet, soll an einem Zahlenbeispiel verdeutlicht werden. Bei Korrelationen von $r_{1c} = 0.60$, $r_{2c} = 0.50$ und $r_{12} = 0.00$ addieren sich die durch die beiden Prädiktoren einzeln determinierten Varianzanteile des Kriteriums von ${r_{1c}}^2 = 0.60^2 = 0.36$ und ${r_{2c}}^2 = 0.50^2 = 0.25$ zu einem durch die multiple Regression determinierten Varianzanteil von $R^2 = 0.36 + 0.25 = 0.61$, und das entspricht einer multiplen Korrelation von $R = \sqrt{0.61} \approx 0.78$. Durch die Kombination der beiden Prädiktoren in einer multiplen Regressionsgleichung erreichen wir also eine Güte der Vorhersage, die mit einem einzelnen Prädiktor kaum zu erreichen ist.

Das Beispiel verdeutlicht aber noch mehr. Auch wenn sich bei der Behandlung des Schätzungseffekts gezeigt hat, daß dieser bei einer Validität von 0.70 noch recht gering ist, ist es trotzdem wichtig, mit einzelnen Prädiktoren wenigstens in die Näher dieses Werts zu kommen; denn dann kann die Validität durch Einbeziehung weiterer Prädiktoren in den Bereich kommen, in dem der Schätzungseffekt steil ansteigt (vgl. Abbildung @26, 26, S. @297, 297).

Nun kann man sich fragen, ob denn nicht zu erwarten ist, daß zwei Prädiktoren, die beide mit einem Kriterium korrelieren, auch untereinander korrelieren. Das trifft aber vor allem dann zu, wenn die Korrelationen der Prädiktoren mit dem Kriterium auf demselben Merkmal beruhen, wenn wir also z.B. Schulleistung mit zwei Intelligenztests vorhersagen wollen. Dann korrelieren natürlich auch diese Prädiktoren untereinander. Unkorrelierte Prädiktoren können wir dagegen erhalten, wenn die Prädiktoren unterschiedliche Persönlichkeitsmerkmale erfassen, die in das Kriterium eingehen. Bei der Vorhersage des Kriteriums der Schulleistung wäre z.B. außer einem Intelligenztest als erstem Prädiktor an einen zweiten Prädiktor zu denken, mit dem die Bereitschaft zum Arbeitseinsatz erfaßt wird.

Welche Rolle die Korrelation der Prädiktoren untereinander spielt, zeigt die folgende Abbildung. Wie im obigen Beispiel wird von Prädiktoren ausgegangen, deren Korrelationen mit dem Kriterium $r_{1c} = 0.60$ bzw. $r_{2c} = 0.50$ sind. Für solche Prädiktoren gibt die Kurve an, wie hoch die multiple Korrelation (Ordinatenachse) bei unterschiedlichen Werten der (auf der Abszissenachse angegebenen) Korrelation r_{12} zwischen den Prädiktoren ist.

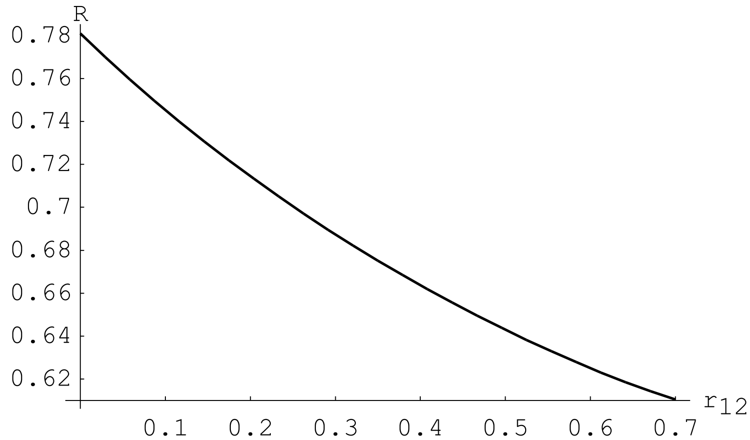

Abb. 27 Die Abhängigkeit der multiplen Korrelation von der Korrelation r_{12} zweier Prädiktoren mit $r_{1c} = 0.60$ und $r_{2c} = 0.50$

Wie man sieht, fällt die multiple Korrelation ab, wenn die Korrelation der Prädiktoren zunimmt, und bei einem Wert von $r_{12} = 0.70$ ist sie kaum noch größer als r_{1c}. Das läßt sich damit erklären, daß der zweite Prädiktor bei einer so hohen Korrelation r_{12} kaum noch etwas zur Vorhersage des Kriteriums beiträgt, was nicht auch schon im ersten enthalten war. Eine weitere Interpretation dieses Abfalls der multiple Korrelation bei zunehmender Korrelation r_{12} wird sich später ergeben.

Daraus ergibt sich eine praktische Konsequenz für die Zusammenstellung von Prädiktoren zur Vorhersage eines Kriteriums, die sich auch auf die Vorhersage mit mehr als zwei Prädiktoren verallgemeinern läßt:

Bei der Vorhersage eines Kriteriums, das auf verschiedenen Personmerkmalen ("Eigenschaften") beruht, ist es günstig, wenn die einzelnen Prädiktoren mit dem Kriterium möglichst hoch, untereinander dagegen möglichst niedrig korrelieren. Dann erfassen sie nämlich unterschiedliche Aspekte des Kriteriums, und die Wahl der Gewichtszahlen in der multiplen Regression bewirkt, daß diese einzelnen Aspekte optimal gewichtet werden.

Zusätzlich zu dieser optimalen Gewichtung verschiedener Personmerkmale, die in ein Kriterium eingehen und die mit unterschiedlichen Prädiktoren erfaßt werden, kann die multiple Regression aber noch auf eine völlig andere Art und Weise zur Verbesserung einer Vorhersage beitragen. Für ein Demonstrationsbeispiel wollen wir annehmen, daß vorhergesagt werden soll, wie oft Sportler (z.B. Fußballspieler) in der nächsten Saison vom Schiedsrichter wegen Foulspiel verwarnt werden. Dazu wird ein Kriterium gebildet, das als "Zahl der Verwarnungen pro Zeit auf dem Spielfeld" berechnet wird.[158] Als Prädiktor wäre z.B. an einen Fragebogen der Aggressivität zu

[158] Wer also als Ersatzspieler in der ganzen Saison nur 120 Minuten eingesetzt wird und sich dabei drei Verwarnungen zuzieht, bekommt einen höheren Kriteriumswert als jemand, der in 1500 Spielminuten 9 Verwarnungen erhält. Die entsprechenden Kriteriumswerte wären nämlich 3 / 120 = 0.025 bzw. 9 / 1500 = 0.006.

denken. Allerdings gibt es ein generelles Problem, wenn man Merkmale wie Aggressivität mit Fragebögen erfaßt. Viele Personen tendieren dazu, sich selbst in Fragebögen so darzustellen, daß sie mit gesellschaftlichen Normen besser übereinstimmen, als es der Realität entspräche. Solche Menschen würden sich in unserer Gesellschaft, in der Aggressivität negativ bewertet wird, als weniger aggressiv darstellen,a ls sie in Wirklichkeit sind. Nun läßt sich diese "Tendenz zur sozialen Erwünschtheit" mit bestimmten "Fang-Fragen" recht gut erfassen,[159] die man zweckmäßigerweise zwischen die Fragen zur Aggressivität im gleichen Fragebogen einstreut. Bei der Auswertung dieses Misch-Fragebogens werden dann aber die Antworten zu den eigentlichen Aggressivitätsfragen und diejenigen zur sozialen Erwünschtheit getrennt ausgezählt. Diese beiden Meßwerte können dann als Prädiktoren verwendet werden.

Nach Transformation in z-Werte sollen also für jeden der Sportler die folgenden Meßwerte zur Verfügung stehen:

z_{i1}: Aggressivitäts-Fragen
z_{i2}: "Fangfragen" zur sozialen Erwünschtheit
z_{ic}: Kriteriumswert (Verwarnungen pro Einsatzzeit).

Nehmen wir nun an, daß sich für diese Meßwerte Korrelationen von $r_{1c} = 0.56$, $r_{2c} = 0.00$ und $r_{12} = -0.60$ ergeben. Nach den allgemeinen Formelen ergibt sich dann für die Gewichtszahlen

$$b_1 = \frac{0.56 - (-0.60) \cdot 0.00}{1 - (-0.60)^2} = 0.875$$

und

$$b_2 = \frac{0.00 - (-0.60) \cdot 0.56}{1 - (-0.60)^2} = 0.525,$$

und mit der Schätzgleichung

$$\hat{z}_{ic} = 0.875 \cdot z_{i1} + 0.525 \cdot z_{i2}$$

läßt sich eine multiple Korrelation von

$$R = \sqrt{\frac{0.56^2 + 0^2 - 2 \cdot 0.56 \cdot 0 \cdot (-0.60)}{1 - (-0.60)^2}} = 0.70$$

erzielen. Was an der multiplen Regressionsgleichung überrascht - zumindest auf den ersten Blick: Obwohl der zweite Prädiktor (also das Maß der Tendenz zu sozial erwünschten Antworten) selbst überhaupt nicht mit dem Kriterium korreliert ($r_{2c} = 0.00$), geht er doch mit beträchtlichem Gewicht in die Schätzgleichung für dieses Kriterium ein und führt damit auch zu einer nicht unbeträchtlichen Verbesserung der Vorhersage. (Allein aufgrund des ersten Prädiktors hätten wir eine "Vorhersage-validität" von $r_{1c} = 0.56$; mit beiden Prädiktoren zusammen - also mit der multiplen Regressions-gleichung - ergibt sich dagegen eine multiple Korrelation von $R = 0.70$.

---

[159] Eine solche "Fang-Frage" könnte z.B. lauten:
Manchmal reizt mich ein gemeiner Witz zum Lachen. . . . . ja - nein
Menschen mit ausgeprägter Tendenz zur sozialen Erwünschtheit würden mit "nein" antworten; wer dagegen ehrlich ist, müßte wohl "ja" antworten.

In diesem Fall ist das auch unmittelbar als sinnvoll nachvollziehbar: Die negative Korrelation der beiden Prädiktoren ($r_{12} = -0.60$) erlaubt den Schluß: Je größer die Tendenz zur sozialen Erwünschtheit, um so mehr wird die tatsächliche Aggressivität über der im Fragebogen zugegebenen liegen. Man könnte auch sagen, daß die Einbeziehung des Maßes der Tendenz zu sozial erwünschten Antworten die für die Vorhersage ungünstige "Beimengung" dieser Tendenz in den Meßwerten des Aggressivitäts-Fragebogens *"unterdrückt"*. Diese Interpretation wird mit der Bezeichnung eines solchen Prädiktors als *"Suppressor"* ausgedrückt. (Englisch: to suppress = unterdrücken).

In einer multiplen Regressionsgleichung ist ein Prädiktor ein Suppressor, wenn sein Beitrag zur Vorhersage des Kriteriums darauf beruht, daß er Anteile eines anderen Prädiktors unterdrückt, die nichts mit dem Kriterium zu tun haben. Insbesondere ist dies der Fall, wenn ein Prädiktor nicht (oder nur geringfügig) mit dem Kriterium korreliert, dafür aber mit einem anderen Prädiktor, der mit dem Kriterium korreliert.

Im obigen Beispiel war es unmittelbar nachvollziehbar, wie die "Fangfragen" zur Erfassung der Tendenz zu sozial erwünschten Antworten dazu beitragen, die Vorhersage des Kriteriums zu verbessern. Dadurch ist dieses Beispiel auch besonders geeignet, eine Gefahr der Interpretation von Gewichtszahlen in multiplen Regressionsgleichungen zu demonstrieren. Wenn man die Suppressor-Funktion des zweiten Prädiktors nicht beachtet, könnte man auf den Gedanken kommen, die Regressionsgleichung $\hat{z}_{ic} = 0.875 \cdot z_{i1} + 0.525 \cdot z_{i2}$ dahingehend zu interpretieren, daß außer der Aggressivität (erster Prädiktor) auch die mit dem zweiten Prädiktor erfaßte Tendenz zur sozialen Erwünschtheit zum unfairen Verhalten von Sportlern beiträgt. Ist also unfaires Verhalten von Sportlern sozial erwünscht? Daß dies eine Fehlinterpretation ist, wird man kaum übersehen, wenn man den zweiten Prädiktor von vorne herein eingeführt hat, um die bekannte Verfälschung von Fragebögen zu sozial unerwünschten Eigenschaften wie der Aggressivität zu korrigieren.

Ähnliche Fehlschlüsse kommen aber in der Forschungspraxis leicht vor, wenn man sich fragt, wie sich ein komplexes Merkmal aus verschiedenen anderen zusammensetzt. Ein (fiktives) Beispiel mag dies veranschaulichen. Bei manchen Lehrern sind Schüler unbeliebt, die einer im Unterricht vorgetragenen Auffassung kritisch widersprechen, und vor allem in der Pubertät wird solche Kritik ja auch oft ziemlich aggressiv vorgetragen. Ein Forscher im Bereich der Pädagogischen Psychologie vertritt aber die Hypothese, daß genau diese in der Pubertät überhöhte Kritikbereitschaft einen positiven Beitrag zum Erreichen der Lernziele des Unterrichts darstellt. Nach dieser Hypothese zeigen die kritischen Äußerungen - trotz ihrer manchmal unpassenden Aggressivität -, daß ein Schüler sich aktiv mit dem Unterrichtsstoff auseinandersetzt, und das führt zu besseren Lernergebnissen als das "brave" Zuhören anderer Schüler. Allerdings werden - so behauptet die Hypothese - von den Schulnoten vor allem die Lernergebnisse derjenigen Schüler belohnt, die brav zuhören und bei ihren Wortmeldungen vor allem versuchen, das zu erraten, worauf der Lehrer hinaus will. Die Lernergebnisse solcher Schüler bestehen natürlich in einer möglichst guten Reproduktion der Auffassungen des Lehrers, und das wird vom Lehrer gut benotet. Anders ist es dagegen bei Klausuren, bei denen die Schüler mehrere Fragen jeweils in wenigen Sätzen beantworten sollen, wenn solche Klausuren nicht vom Lehrer, sondern von neutralen Fachexperten ausgewertet werden, die auch auf selbständige Urteilsbildung achten. Dann werden zusätzlich zu den vom Lehrer vor allem beachteten Kenntnissen auch diejenigen Lernergebnisse erfaßt, sich aus der eher kritischen Auseinandersetzung mit der Auffassung des Lehrers ergeben.

Um die Hypothese zu überprüfen, wird ein geeigneter Unterrichtsgegenstand der Sozialkunde ausgewählt, zu dem es an mehreren Gymnasien in der Oberstufe Leistungskurse gibt. Während des Unterrichts in solchen Kursen wird von entsprechend trainierten Beobachtern für jeden Schüler die Zahl seiner kritischen Äußerungen zu Auffassungen des Lehrers registriert. Außerdem stehen die Abschlußnoten zur Verfügung (Punktzahlen von 0 bis 15, wobei höhere Punktzahlen bessere Noten sind). Schließlich wird im Rahmen der Untersuchung eine Klausur der o.g. Art (mit gleichen Fragen in allen Schulen) durchgeführt und von neutralen Experten benotet, und zwar auf derselben Skala mit 0 bis 15 Punkten. Die Klausuren werden aber erst geschrieben, nachdem der Lehrer sich auf seine Note festgelegt hatte.

Zur statistischen Auswertung dieser Daten stellt der Forscher nun die folgende Überlegung an: Wenn - entsprechend seiner Hypothese - zu den von Klausurnoten erfaßten komplexeren Lernergebnissen nicht nur die von den Lehrernoten berücksichtigen Kenntnisse beitragen sondern auch die kritische Auseinandersetzung, dann müßte sich dies auch in einer multiplen Regressionsgleichung darstellen lassen. Dabei bilden die Klausurnoten das Kriterium. Die Lehrernoten werden als erster Prädiktor verwendet, und die von den Beobachtern erfaßte Zahl kritischer Äußerungen bildet den zweiten Prädiktor. Wenn die Hypothese zutrifft, müßten in dieser multiplen Regressionsgleichung beide Prädiktoren eine positive Gewichtszahl haben. Also wird die multiple Regressionsgleichung für die (in z-Werte transformierten) Daten berechnet, und es ergibt sich

$$\hat{z}_{ic} = 0.60 \cdot z_{i1} + 0.25 \cdot z_{i2},$$

was der Forscher als Bestätigung seiner Hypothese interpretiert.

Ist dieser Schluß zwingend? Bevor gezeigt wird, daß dies nicht der Fall ist, soll nachvollzogen werden, was der Forscher bei seiner Interpretation wohl im Auge hatte. Wenn ein vorherzusagendes Kriterium auf mehreren unkorrelierten Personenmerkmalen beruht, werden ja die Gewichtszahlen der multiplen Regressionsgleichung so bestimmt, daß sich eine (im Sinne des Prinzips der kleinsten Qudrate) optimale Annäherung des Kriteriums durch eine Linearkombination der Prädiktoren erreicht wird. Das interpretiert der Forscher nun so, daß die Gewichtszahlen angeben, mit welchem Gewicht (in einem kausalen Sinn) die von den beiden Prädiktoren erfaßten Merkmale - also die Reproduktion der vom Lehrer vermittelten Kenntnisse und Auffassungen sowie die Bereitschaft zur kritischen Auseinandersetzung - zu den Lernergebnissen beitragen, die von den Klausurnoten (dem Kriterium) erfaßt werden. Dementsprechend würde die obige Schätzgleichung sich bei Korrelationen von $r_{1c} = 0.60$, $r_{2c} = 0.25$ und $r_{12} = 0.00$ aus den allgemeinen Formeln für die Gewichtszahlen b_1 und b_2 ergeben.

Trotzdem ist eine solche Interpretation der Gewichtszahlen als "kausale Gewichte" aber sehr fragwürdig, und dabei ist zusätzlich zu den allgemeinen Problemen der kausalen Interpretation statistischer Beziehungen auch an Suppressor-Effekte zu denken. Die Gewichtszahlen von $b_1 = 0.60$ und $b_2 = 0.25$ in der obigen Regressionsgleichung würde man z.B. auch bei Korrelationen von $r_{1c} = 0.50$, $r_{2c} = 0.01$ und $r_{12} = -0.40$ erhalten (wollen Sie es nachrechnen?), und diese Korrelationen würden eine ganz andere Interpretation der positiven Gewichtszahl b_2 nahelegen, auf die es dem Forscher ja vor allem ankam. Für den zweiten Prädiktor liegt die klassische Situation eines Suppressors vor: So gut wie keine Korrelation mit dem Kriterium[160], aber beträchtliche

---

[160] Mit einer Korrelation von genau $r_{2c} = 0.00$ ergeben sich Gewichtszahlen von $b_1 = 0.60$ und $b_2 = 0.25$ z.B. mit den Korrelationen $r_{1c} = 119 / 240$ und $r_{12} = 5 / 12$. Aber schon in der Definition des Suppressors wurde zugelassen, daß seine Korrelation mit dem Kriterium nicht genau

Korrelation mit einem anderen Prädiktor, der mit dem Kriterium korreliert. Der "Sinn" dieses Suppressors läßt sich auch gut nachvollziehen. Die Bereitschaft zu kritischen Äußerungen (zweiter Prädiktor) trägt vielleicht selbst nicht nennenswert zu den von den Klausurnoten erfaßten Lernergebnissen bei - weder positiv noch negativ (r_{2c} =0.01); aber die negative Korrelation der Prädiktoren ($r_{12} = -0.41$) beruht darauf, daß Lehrer die oft aggressive Kritik von Jugendlichen als "Aufsässigkeit" betrachten und mit schlechten Noten bestrafen. Wenn dies zutrifft, kann man den zweiten Prädiktor ganz ähnlich verwenden, wie die Maße der Tendenz zu sozialer Erwünschtheit beim Aggressionsfragebogen: Um eine "Beimischung" eines Prädiktors zu unterdrücken, die nichts mit dem Kriterium zu tun hat. Bei den Schulnoten läßt sich das folgendermaßen nachvollziehen: Wenn Lehrer kritische Äußerungen mit schlechten Noten bestrafen, dann ist diese Verfälschung um so größer, je mehr kritische Äußerungen von einem Schüler kommen. Um die (von den Klausurnoten besser erfaßten) tatsächlichen Lernergebnisse zu schätzen, sind die vom Lehrer vergebenen Punktzahlen also um so mehr "aufzubessern", je mehr kritische Äußerungen eines Schülers registriert wurden. Genau das geschieht aber, wenn der zweite Prädiktor mit positivem Gewicht in die Regressionsgleichung aufgenommen wird.

Die Überlegungen lassen sich folgendermaßen verallgemeinern:

Die Bestimmung der Gewichtszahlen in der multiplen Regression erfolgt mit der Zielsetzung, das Kriterium möglichst gut vorherzusagen. Zu diesem Ziel können als Suppressoren auch Prädiktoren beitragen, die selbst keinen eigenständigen, z.B. kausalen Bezug zum Kriterium haben, sondern eine unerwünschte "Beimischung" unterdrücken, die ein anderer Prädiktor enthält. Damit sind die Gewichtszahlen in einer multiplen Regressionsgleichung aber ungeeignet, etwas über das Gewicht eines kausalen Beitrags der Prädiktoren zum Kriterium auszusagen.

Nun könnte man versucht sein, die Gefahr solcher Fehlinterpretationen mit dem Hinweis zu verharmlosen, daß es doch leicht ist, Suppressoren daran zu erkennen, daß sie aufgrund ihrer Korrelation mit einem anderen Prädiktor in die Regressionsgleichung eingehen, obwohl ihre Korrelation mit dem Kriterium null ist. Das trifft aber nur für Prädiktoren zu, deren Beitrag zur Vorhersage *ausschließlich* auf einer Funktion als Suppressor beruht. Es gibt aber auch Korrelationsmuster, bei denen sich ein eigenständiger Beitrag zur Vorhersage und eine Funktion als Suppressor überlagern. Die nachfolgende Demonstration einiger derartiger Überlagerungsmöglichkeiten ist auf den ersten Blick etwas langwierig; sie vertieft aber die obige Warnung vor voreiligen Interpretationen der Gewichtszahlen in einer multiplen Regressionsgleichung. Trotzdem: Wem die Demonstration zu lang wird, kann ja bis zum Ende des "Spezialisten"-Abschnitts (S. 334) springen.

Für ein erstes Beispiel der Überlagerung von eigenständigen Beiträgen zur Vorhersage und Suppressor-Effekten ist darauf hinzuweisen, daß z.B. auch Korrelationen von $r_{1c} = 0.55$, $r_{2c} = 0.13$ und $r_{12} = -0.20$ zu der oben betrachteten Regressionsgleichung mit $b_1 = 0.60$ und $b_2 = 0.25$ führen würden. Diese Korrelationen könnten etwa dann entstehen, wenn zwar - wie es die zu untersuchende Hypothese behauptet - die in der Pubertät erhöhte Kritikbereitschaft mancher Schüler zur Bildung

---

0.00, sondern "geringfügig" ist, und das ist bei $r_{2c} = 0.01$ sicher der Fall. Ansonsten gibt es natürlich Grenzfälle, und im weiteren Verlauf wird sich sogar zeigen, daß auch bei nennenswert von 0 abweichenden Korrelationen r_{2c} noch eine Suppressor-Funktion vorliegen kann.

eines eigenständigen Urteils beiträgt (positive Korrelation r_{2c}), wenn aber *außerdem* der Lehrer kritische Äußerungen mit schlechten Noten bestraft. Bei diesen Korrelationen ergäbe sich eine multiple Korrelation von $R = 0.60$. Was die Schätzgleichung und die multiple Korrelation mit Suppressoren zu tun hat, kann man sich am ehesten vergegenwärtigen, wenn man vergleicht, was bei $r_{12} = 0$ herausgekommen wäre. Für den Fall unkorrelierter Prädiktoren wurde bereits festgestellt, daß die Korrelationen der Prädiktoren mit dem Kriterium auch ihre Gewichtszahlen in der multiplen Regressionsgleichung sind und daß die durch diese Regression determinierte Varianz des Kriteriums (also das Quadrat der multiplen Korrelation) die Summe der einzelnen Determinationskoeffizienten ist. Für die Schätzgleichung $\hat{z}_{ic} = 0.55 \cdot z_{i1} + 0.13 \cdot z_{i2}$, die demnach bei $r_{1c} = 0.55$, $r_{2c} = 0.13$ und $r_{12} = 0$ zu verwenden wäre, ergäbe sich also ein Quadrat der multiplen Korrelation von $R^2 = 0.55^2 + 0.13^2 \approx 0.32$ und damit eine multiple Korrelation von $R \approx \sqrt{0.32} = 0.57$. Die höheren Gewichtszahlen b_1 und b_2 und die höhere multiple Korrelation bei einer negativen Korrelation r_{12} lassen sich folgendermaßen interpretieren: Zunächst hat jeder Prädiktor einen Anteil mit dem Kriterium gemeinsam, der - genau wie bei $r_{12} = 0$ - zur Vorhersage des Kriteriums beiträgt. Darüber hinaus haben die beiden Prädiktoren aber miteinander einen Anteil gemeinsam, der nichts mit dem Kriterium zu tun hat: In die Note des Lehrers (den ersten Prädiktor) geht zusätzlich zur Leistung auch die negative Bewertung von "Aufsässigkeit" ein, und die Zahl kritischer Äußerungen (zweiter Prädiktor) ist vielleicht nicht *nur* ein Indikator für Engagement, sondern *auch* für die in der Pubertät verbreitete überzogene Kritikbereitschaft. Dieser gemeinsame Anteil geht aber in beide Prädiktoren mit unterschiedlichem Vorzeichen ein: Die Note des Lehrers wird durch diese Kritikbereitschaft negativ beeinflußt, die beobachtete Zahl kritischer Äußerungen dagegen positiv. In dieser Situation haben dann beide Prädiktoren eine Doppelfunktion. Einerseits trägt jeder "aus eigener Kraft" (also aufgrund seiner positiven Korrelation mit dem Kriterium) zur Vorhersage bei, und außerdem wirkt jeder auch noch als Suppressor für die Kriteriums-fremden Anteile des anderen. Man könnte auch sagen, daß die negativen Einflüsse der Kritikbereitschaft auf den ersten Prädiktor und die positiven Einflüsse auf den zweiten Prädiktor sich gegenseitig aufheben.

Allerdings ist dies nur eine von mehreren möglichen Interpretationen des Zusammenhangs der drei Variablen. Denkbar wäre es bei den angenommenen Korrelationen z.B. auch, daß der Lehrer nicht die Kriteriums-fremden Anteile des zweiten Prädiktors ("überzogene Kritikbereitschaft") negativ bewertet, sondern genau diejenigen, die positiv zu den vom Kriterium erfaßten Leistungen beitragen. Auch dann hätten aber beide Prädiktoren neben ihrem eigenständigen Beitrag zum Kriterium noch eine Suppressor-Funktion. Damit ergibt sich ein weiteres allgemeines Prinzip, das sich am einfachsten für den Fall formulieren läßt, daß die Prädiktoren positiv mit dem Kriterium korrelieren:[161]

Korrelieren zwei Prädiktoren positiv mit einem Kriterium, untereinander dagegen negativ, dann erhöhen sich die Gewichte in der multiplen Regressionsgleichung und auch die multiple Korrelation durch zusätzliche Suppressoreffekte.

Das geht auch unmittelbar aus den Formeln für die Gewichtszahlen b_1 und b_2 hervor. Dort wird im Zähler das Produkt $r_{12} \cdot r_{1c}$ bzw. $r_{12} \cdot r_{2c}$ subtrahiert. Bei positiver Korrelation der beiden

[161] Bei negativen Korrelationen zwischen Prädiktoren und Kriterium kann durch geeignete Umpolung geprüft werden, ob eine Situation entsteht, die das allgemeine Prinzip erfüllt.

Prädiktoren führte dies indirekt zu der in Abbildung 27, 27 (S. 324, 324) dargestellten Senkung der multiplen Korrelation. Ist dagegen die Korrelation der Prädiktoren negativ, dann sind (bei positiven Korrelationen r_{1c} und r_{2c}) die subtrahierten Produkte $r_{12} \cdot r_{1c}$ und $r_{12} \cdot r_{2c}$ negativ, wodurch sich der Zähler nicht mehr vermindert, sondern sogar erhöht, und die Division dieses Zählers durch den Nenner $1 - r_{12}^2$ (der ja bei $r_{12} \neq 0$ kleiner als 1 ist) führt zu einer weiteren Erhöhung der Gewichtszahlen b_1 und b_2. Aus der allgemeinen Gleichung für die multiple Korrelation folgt aber, daß diese (bei positiven Korrelationen r_{1c} und r_{2c}) zunimmt, wenn die Gewichtszahlen b_1 und b_2 größer werden als diese Korrelationen.

Das sieht man auch deutlich, wenn man in Abbildung 27, 27 den Bereich der auf der Abszissenachse angegebenen Korrelationen r_{12} vergrößert wie in der folgenden Abbildung, die zusätzlich noch eine Überraschung beinhaltet.

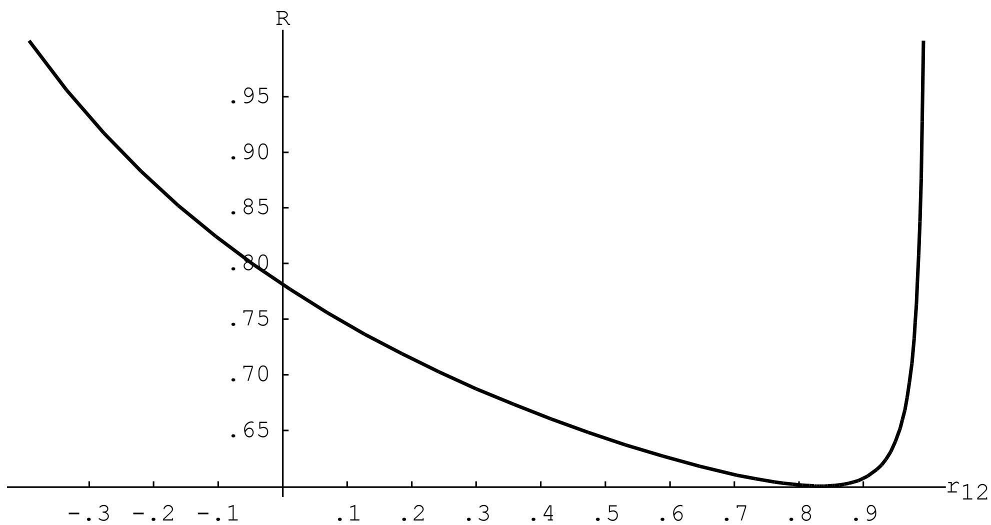

Abb. 28 Die Abhängigkeit der multiplen Korrelation von der Korrelation r_{12} zweier Prädiktoren mit $r_{1c} = 0.60$ und $r_{2c} = 0.40$: Überlagerung mit Suppressor-Effekten

Bevor diese Überraschung - die Zunahme der multiplen Korrelation im rechten Teil der Abbildung - erläutert wird, ist zunächst festzustellen, daß die multiple Korrelation - wie erwartet - weiter zunimmt, wenn die Korrelation r_{12} der Prädiktoren negativ wird.[162]

Wie kommt es aber, daß die multiple Korrelation bei sehr hohen Werten der Korrelation r_{12}

[162] Die Abbildung legt sogar die Frage nahe, ob bei einer noch niedrigeren Korrelation der Prädiktoren (z.B. $r_{12} = -0.50$) nicht eine multiple Korrelation zu erwarten wäre, die größer als 1 ist, was offenbar sinnlos wäre. Eine solche Korrelation r_{12} kann aber bei $r_{1c} = 0.60$ und $r_{2c} = 0.50$ nicht vorkommen. Im Abschnitt zur Partialkorrelation wurde festgestellt, daß es bei Vorgabe von Korrelationen r_{xz} und r_{yz} Grenzen für die Korrelation r_{xy} gibt. Um die entsprechende Formel (S. 316) auf die hiesige Situation anzuwenden, kann man die Prädiktoren als x und y bezeichnen und das Kriterium als z. Dann ergibt sich, daß für die Korrelation der Prädiktoren die Ungleichung $-0.393 \leq r_{12} \leq 0.993$ gelten muß.

wieder ansteigt? Bei der Interpretation von Abbildung 27, 27 war der Abfall der multiplen Korrelation bei zunehmender Korrelation r_{12} vor allem damit erklärt worden, daß der zweite Prädiktor bei einer hohen Korrelation r_{12} kaum noch etwas zur Vorhersage des Kriteriums beiträgt, was nicht auch schon im ersten enthalten war. Unter Rückgriff auf das Suppressor-Konzept kann aber noch ein weiterer Grund genannt werden. Eine hohe Korrelation r_{12} bedeutet, daß die beiden Prädiktoren über das hinaus, was sie zur Vorhersage des Kriteriums beizutragen haben, noch etwas anderes gemeinsam haben. Ist diese Situation sehr stark ausgeprägt (ist also r_{12} sehr hoch), dann kann der zweite Prädiktor, der wegen seiner niedrigeren Korrelation mit dem Kriterium ohnehin schlechter zur Vorhersage des Kriteriums geeignet ist als der erste ($r_{2c} = 0.50$ im Vergleich zu $r_{1c} = 0.60$), auch als Suppressor eingesetzt werden. Tatsächlich ergibt sich z.B. bei Korrelationen von $r_{1c} = 0.60$, $r_{2c} = 0.50$ und $r_{12} = 0.90$ aus der allgemeinen Formel für die Gewichtszahl b_2 ein Wert von $b_2 = -0.21$, während die Gewichtszahl für den ersten Prädiktor $b_1 = 0.79$ beträgt. In diesem Fall geht also der zweite Prädiktor mit einem negativen Gewicht in die multiple Regressionsgleichung ein, obwohl er für sich allein betrachtet aufgrund seiner Korrelation mit dem Kriterium mit positivem Gewicht eingehen könnte. Aber in dieser Konstellation, in der der erste Prädiktor das noch besser kann, ist es günstiger, den zweiten Prädiktor als Suppressor einzusetzen und mit negativem Gewicht zu verrechnen.

Aus der Formel für die Gewichtszahl b_2 kann man auch ersehen, bei welchem Wert der Korrelation r_{12} diese Gewichtszahl von positiven in negative Werte umschlägt. Da der Nenner $1 - r_{12}^2$ immer positiv ist, ergibt sich das Vorzeichen von b_2 aus dem Zähler $r_{2c} - r_{12} \cdot r_{1c}$. Bei positiven Korrelationen r_{1c} und r_{2c} mit $r_{1c} > r_{2c}$ ist dieser Zähler positiv, solange r_{12} kleiner als der Quotient r_{2c} / r_{1c} ist. Ist dagegen $r_{12} > r_{2c} / r_{1c}$, dann ist der Zähler negativ und damit auch die Gewichtszahl b_2. Bei den bisher angenommenen Korrelationen von $r_{1c} = 0.60$ und $r_{2c} = 0.50$ ist beispielsweise $r_{2c} / r_{1c} = 0.50 / 0.60 \approx 0.83$. Genau an dieser Stelle liegt aber in Abbildung 28 das Minimum der multiplen Korrelation. Bei $r_{12} < 0.83$ überwiegt noch der eigene Beitrag des zweiten Prädiktors zur Vorhersage ($b_2 > 0$), während bei größerem r_{12} die Suppressor-Funktion überwiegt ($b_2 < 0$).

Mit diesen Überlegungen läßt sich schließlich auch erklären, warum die multiple Korrelation in der Umgebung des "Umschlagspunkts" kaum größer ist als r_{1c}. Genau in diesem Umschlagspunkt - also für Korrelationen von $r_{1c} = 0.60$, $r_{2c} = 0.50$ und $r_{12} = 0.50 / 0.60$ - erhält man mit den entsprechenden allgemeinen Formeln Gewichtszahlen von $b_1 = 0.60$ und $b_2 = 0.00$ sowie eine multiple Korrelation von $R = 0.60$. Hier ist also der zweite Prädiktor aufgrund der starken Beimischung des Kriteriums-fremden gemeinsamen Anteils der beiden Prädiktoren zu schlecht, um noch einen eigenständigen Beitrag zur Vorhersage zu leisten; andererseits ist aber diese Beimischung Prädiktor noch nicht ausgeprägt genug, um den zweiten als Suppressor zu verwenden. Also ist $b_2 = 0.00$, und dann ist natürlich die optimale Gewichtszahl b_1 des ersten Prädiktors und die Vorhersage-Validität der multiplen Regressionsgleichung (also die multiple Korrelation) genau so groß, als wenn es den zweiten Prädiktor nicht gäbe. Eine ähnliche Situation besteht aber auch in der unmittelbaren Umgebung des Umschlagspunktes. Der Übergang zwischen eigenständigem Beitrag zur Vorhersage (also positivem b_2) und Suppressor-Funktion (negativem b_2) erfolgt kontinuierlich[163], und daher heben sich in der unmittelbaren Umgebung des Umschlagspunkts beide mögliche Beiträge des zweiten Prädiktors nahezu auf.

---

[163] Mathematisch ausgedrückt: Bei gegebenem $r_{1c \text{ und } r2c}$ ist b_2 eine stetige Funktion von r_{12}.

Da Korrelationen über 0.80 zwischen psychologischen Variablen eine Seltenheit sind, sollte ergänzend darauf hingewiesen werden, daß der Umschlagspunkt zwischen eigenständigem Beitrag zur Vorhersage und Suppressor-Funktion wesentlich näher bei 0 liegen kann, wenn die Korrelationen der beiden Prädiktoren mit dem Kriterium sich stärker unterscheiden. Das kann mit einem weiteren Muster von Korrelationen demonstriert werden, das ebenso wie mehrere bereits behandelte Korrelationen zu der multiplen Regressionsgleichung $\hat{z}_{ic} = 0.60 \cdot z_{i1} + 0.25 \cdot z_{i2}$ für die Schätzung von Klausurnoten aus Lehrernoten und Anzahl kritischer Äußerungen im Unterricht führen würde. Die Gewichtszahlen von $b_1 = 0.60$ und $b_2 = 0.25$ würden sich aus den entsprechenden allgemeinen Formeln auch bei Korrelationen von $r_{1c} = 0.40$, $r_{2c} = -0.23$ und $r_{12} = -0.80$ ergeben. Mögliche Interpretation: Anders, als es die Hypothese vermutet, würde bei diesen Korrelationen die Bereitschaft zu kritischen Äußerungen nicht positiv, sondern eher negativ zu den von den Klausurnoten erfaßten Lernergebnissen beitragen ($r_{2c} = -0.23$), und allein auf dieser Grundlage wäre eine negative Gewichtszahl b_2 zu erwarten. Die obigen Überlegungen zum Umschlag zwischen eigenständigem Vorhersagebeitrag und Suppressor-Funktion lassen sich aber folgendermaßen übertragen Bei Korrelationen von $r_{1c} = 0.40$ und $r_{2c} = -0.23$ liegt der "Umschlagspunkt" zwischen eigenständigem Beitrag zur Vorhersage und Suppressor-Funktion bei $r_{12} = -0.23 / 0.40 = -0.575$. Obwohl also der zweite Prädiktor für sich genommen mit negativem Vorzeichen in die Vorhersage des Kriteriums eingehen würde, ergibt sich zusammen mit dem ersten Prädiktor bei Korrelationen von $r_{1c} = 0.40$ und $r_{2c} = -0.23$ eine negative Gewichtszahl b_2 nur für $r_{12} > -0.575$. Ist dagegen die negative Korrelation der Prädiktoren untereinander stärker ausgeprägt (ist also $r_{12} < 0.575$), dann ist es günstiger, auf den eher geringfügigen eigenständigen Vorhersagebeitrag des zweiten Prädiktors zu verzichten und diesen statt dessen mit einer positiven Gewichtszahl als Suppressor einzusetzen.

Zu den Interpretationen der verschiedenen Korrelationsmuster, die alle zu derselben multiplen Regressionsgleichung $\hat{z}_{ic} = 0.60 \cdot z_{i1} + 0.25 \cdot z_{i2}$ für die Schätzung von Klausurnoten aus Lehrernoten und Anzahl kritischer Äußerungen im Unterricht führten, ist eine abschließende Bemerkung angebracht. Diese Interpretationen gingen alle stillschweigend davon aus, daß die Korrelationen r_{1c} und r_{2c} einen Einfluß der von den beiden Prädiktoren erfaßten Persönlichkeitsmerkmale auf diejenigen Lernergebnisse widerspiegeln, die von den Klausurnoten erfaßt werden. Natürlich gibt es - wie bei allen Korrelationen - noch andere denkbare Kausalrichtungen. Was gezeigt werden sollte, läßt sich folgendermaßen zusammenfassen: Selbst wenn diese Kausalrichtung vorliegt, wie es ja von der zu untersuchenden Hypothese behauptet wurde, dann sind die Gewichtszahlen der Prädiktoren in einer multiplen Regressionsgleichung nicht geeignet, den kausalen Beitrag der entsprechenden Persönlichkeitsmerkmale zu den vom Kriterium erfaßten Lernergebnisse quantitativ anzugeben. Dies gilt nicht nur für den zweiten Prädiktor, der in verschiedenen Korrelationsmustern als Suppressor diente (entweder als reiner Suppressor oder auch mit zusätzlichem eigenem Vorhersagebeitrag). Auch die Gewichtszahl des ersten Prädiktors wird von der Anwesenheit eines solchen Suppressors beeinflußt.[164]

---

[164] Genauer: Sind Prädiktoren und Kriterium als z-Werte gegeben und ist der zweite Prädiktor ein "reiner" Suppressor (also $r_{2c} = 0$ und $r_{12} \neq 0$), dann ergibt sich für den Zähler in der Formel für b_1 die Gleichung $r_{1c} - r_{12} \cdot r_{2c} = r_{1c}$. Dieser Zähler ist also das Regressionsgewicht, das auch bei Vorhersage des Kriteriums allein aus dem ersten Prädiktor zu verwenden wäre. Dieser Wert wird aber durch den Nenner $1 - r_{12}^2$ dividiert und damit vergrößert, da die Differenz $1 - r_{12}^2$

Trotz der Komplexität dieser Verhältnisse könnte man aus den Beidpielen den Eindruck bekommen, daß das Zusammenspiel von eigenständigem Vorhersagebeitrag und Suppressor-Funktion eines Prädiktors mit etwas Nachdenken zu durchschauen ist. Dazu ist aber zu bedenken, daß alle bisherigen Beispiele sich auf die noch vergleichsweise überschaubaren Formeln für Gewichtszahlen und multiple Korrelation bei zwei Prädiktoren bezogen. Bereits für drei Prädiktoren würden aber allgemeine Formeln der obigen Art so unübersichtlich, daß sie kaum noch überschaubar wären.[165] Daher soll nur für einen ganz speziellen Fall dargestellt werden, was bei mehr als zwei Prädiktoren noch hinzukommt. Dieser spezielle Fall ist dadurch gekennzeichnet, daß drei Prädiktoren vorhanden sind, bei denen die Korrelationen r_{2c}, r_{3c} und r_{13} alle null sind. Sind außerdem Kriterium und Prädiktoren als z-Werte gegeben, dann ergeben sich die folgenden Formeln für die Gewichtszahlen der drei Prädiktoren:

$$b_1 = \frac{r_{1c} \cdot (1 - r_{23}^2)}{1 - r_{12}^2 - r_{23}^2},$$

$$b_2 = -\frac{r_{12} \cdot r_{1c}}{1 - r_{12}^2 - r_{23}^2}$$

und

$$b_3 = \frac{r_{12} \cdot r_{23} \cdot r_{1c}}{1 - r_{12}^2 - r_{23}^2}.$$

Für Korrelationen von $r_{1c} = 0.36$, $r_{12} = 0.40$ und $r_{23} = 0.60$ ergibt sich damit die Schätzgleichung

$$\hat{z}_{ic} = 0.48 \cdot z_{i1} - 0.30 \cdot z_{i2} + 0.18 \cdot z_{i3},$$

---

bei $r_{12} \neq 0$ kleiner als 1 ist. Wenn sich dagegen eigenständiger Vorhersagebeitrag und Suppressor-Funktion beim zweiten Prädiktor überlagern (also bei $r_{2c} \neq 0$ und $r_{12} \neq 0$), dann wird auch der Zähler $r_{1c} - r_{12} \cdot r_{2c}$ in der Formel für b_1 durch die Anwesenheit des zweiten Prädiktors verändert, und nur in dem zuvor behandelten "Umschlagspunkt" zwischen dem Überwiegen des eigenständigen Vorhersagebeitrags und der Suppressor-Funktion (also bei $r_{12} = r_{2c} / r_{1c}$) entsteht für den ersten Prädiktor dieselbe Gewichtszahl, als wenn der zweite Prädiktor gar nicht vorhanden wäre.

[165] Wie bereits an früherer Stelle (Fußnote 155) erwähnt, ergeben sich wesentlich einfachere Formeln bei Verwendung der sog. Matrix-Algebra. Grundsätzlich läßt sich das Verfahren auch ohne Matrix-Algebra darstellen (vgl. etwa Gaensslen & Schubö, 1976, S. 96ff.). Die Darstellung dieses Ansatzes ist aber recht komplex und würde auch wenig zum Verständnis der folgenden Überlegungen beitragen.

allerdings nur unter der den Formeln zugrundeliegenden Voraussetzung daß die übrigen Korrelationen r_{2c}, r_{3c} und r_{13} alle null sind. Unter dieser Voraussetzung ist vor allem bemerkenswert, daß der dritte Prädiktor mit einem Gewicht von 0.18 in die Regressionsgleichung eingeht, obwohl die einzige Variable, mit der er korreliert, der zweite Prädiktor ist. Da die Korrelation r_{2c} aber ebenfalls null ist, ist der dritte Prädiktor auch kein Suppressor, jedenfalls nicht im bisher bekannten Sinn.

Die Bedeutung des dritten Prädiktors in dieser Regressionsgleichung wird verständlich, wenn man sich noch einmal die Funktion eines Suppressors vergegenwärtigt. Er verbessert die Vorhersagefähigkeit eines anderen Prädiktores, indem er diesen von Beimischungen bereinigt, die nichts mit dem Kriterium zu tun haben. Man kann das auch folgendermaßen formulieren: Bei $r_{1c} \neq 0$, $r_{2c} = 0$ und $r_{12} \neq 0$ verbessert sich die Vorhersage, wenn statt des ersten Prädiktors seine Abweichung von dem Wert in die Regressionsgleichung eingeht, der aufgrund des zweiten Prädiktors zu erwarten wäre. In der obigen Situation verbessert sich aber die Vorhersage des ersten Prädiktors, wenn man zusätzlich zum zweiten Prädiktor auch noch den dritten heranzieht; denn dieser kann bei der Vorhersage des ersten Prädiktors als Suppressor eingesetzt werden. (Es ist $r_{13} = 0$, aber $r_{12} \neq 0$ und $r_{23} \neq 0$.) In dieser Situation werden also durch die Einbeziehung des dritten Prädiktors in die Vorhersage des zweiten noch mehr Kriteriums-fremde Beimischungen des ersten Prädiktors unterdrückt. Man kann diese Funktion des dritten Prädiktors also folgendermaßen zusammenfassen: Indem er als Suppressor die Vorhersage des ersten Prädiktors verbessert, trägt er als "Suppressor zweiten Grades" auch zur Verbesserung der Vorhersage des ersten Prädiktors bei.

Das ist unter den Voraussetzungen der obigen Formeln (also bei $r_{2c} = r_{3c} = r_{13} = 0$) noch gerade durchschaubar; aber diese Funktion als Suppressor zweiten Grades kann sich nun auch wieder auf unterschiedlichste Weise mit eigenständigen Beiträgen des dritten Prädiktors zur Vorhersage des Kriteriums oder des ersten Prädiktors überlagern, und dann ist es kaum noch durchschaubar, wie die Gewichtszahlen in der multiplen Regressionsgleichung mit den Korrelationen zusammenhängen. Es braucht kaum noch erwähnt zu werden, daß diese Komplexität der Überlagerung von eigenständigen Vorhersagebeiträgen und Suppressor-Funktionen (einschließlich Suppressoren höheren Grades) mit jedem weiteren Prädiktor zunimmt.

>>

Ergänzend ist noch darauf hinzuweisen, daß die hier besprochene multiple Regression und Korrelation nur die linearen Abhängigkeiten der Prädiktoren und des Kriteriums auswerten. Es ist jedoch möglich, auch nichtlineare Zusammenhänge mit ähnlichen Techniken zu verrechnen.

<<

Viele Statistik-Software-Pakete für Computer enthalten Programme, mit denen man sogenannte[166] nicht-lineare Regression berechnen kann. Dabei kann man als Schätzgleichung für ein Kriterium eine nahezu beliebige Funktion der Prädiktoren mit einer Formel angeben, in der außer den Prädiktoren auch weitere Größen wie die *b*-Gewichte in der multiplen Regression auftauchen. Oft werden diese Größen als "Regressionsparameter" o.ä. bezeichnet. Dann berechnet der Computer

[166] Das Wort "sogenannte" spielt darauf an, daß das, wonach die Regression eigentlich benannt ist - der Regressionseffekt - bei der nicht-linearen Regression nicht immer auftritt.

diejenigen Werte für diese Regressionsparameter, bei denen die Schätzwerte am besten mit den tatsächlichen Kriteriumswerten übereinstimmen. Die meisten derartigen Programme ermöglichen es auch, anstelle der quadrierten Abweichung zwischen tatsächlichen und geschätzten Werten des Kriteriums eine andere Funktion vorzugeben, deren Summe über alle "Individuen" durch geeignete Wahl der Regressionsparameter möglichst klein gemacht werden soll.

Bei einigen Typen nicht-linearer Schätzgleichungen ist es aber auch möglich, die optimalen Werte der Regressionsparameter mit Hilfe der für lineare multiple Regressionsgleichungen entwickelten Formeln bzw. Programme zu bestimmen. Zur Vorbereitung eines Beispiels ist zunächst auf eine Eigenschaft der linearen Regressionsgleichungen mit zwei Prädiktoren hinzuweisen. Bei einer Schätzgleichung wie $\hat{Y}_i = 0.60 \cdot X_{i1} + 0.25 \cdot X_{i2} + a$ ergeben sich hohe Schätzwerte zwar vor allem dann, wenn die Prädiktoren-Werte X_{i1} und X_{i2} beide groß sind; doch können niedrige Werte eines Prädiktors durch entsprechend höhere Werte des anderen Prädiktors kompensiert werden. In manchen Fällen ist es aber auch sinnvoll, davon auszugehen, daß eine solche Kompensation gar nicht (oder nur eingeschränkt) möglich ist, daß also hohe Werte des Kriteriums nur (oder fast nur) vorkommen, wenn die Werte beider Prädiktoren hoch sind. Das läßt sich mit einer Schätzgleichung der Form

$$\hat{Y}_i = b_1 \cdot X_{i1} + b_2 \cdot X_{i2} + b_3 \cdot X_{i1} \cdot X_{i2} + a$$

erreichen.[167] Wenn in dieser Schätzgleichung die Gewichtszahl b_3, mit der das Produkt $X_{i1} \cdot X_{i2}$ in die Schätzgleichung eingeht, positiv und hinreichen groß (im Vergleich zu b_1 und b_2) ist, dann können niedrige Werte eines Prädiktors nicht mehr einfach durch hohe Werte des anderen Prädiktors kompensiert werden. Mit der Wahl eines geeigneten negativen b_3 in Verbindung mit positiven Gewichtszahlen b_1 und b_2 erreicht man andererseits, daß hohe Werte in einem Prädiktor genügen, um einen hohen Schätzwert zu erhalten, und daß hohe Werte in beiden Prädiktoren keine höheren Schätzwerte ergeben als hohe Werte in einem Prädiktor. Mit positivem b_3 ergeben sich also hohe Schätzwerte, wenn X_{i1} und X_{i2} beide hoch sind, während negatives b_3 in Verbindung mit positiven Gewichtszahlen b_1 und b_2 zu hohen Schätzwerten führt, wenn X_{i1} oder X_{i2} hoch ist. Beides kann in bestimmten Fällen sinnvoll sein.

Ob ein solcher Fall vorliegt, kann man nun einfach feststellen, und das Verfahren dazu ist durch die Bezeichnung b_3 für die Gewichtszahl des Produkts $X_{i1} \cdot X_{i2}$ bereits angedeutet: Man bildet nach der Definition

$$X_{i3} := X_{i1} \cdot X_{i2}$$

---

[167] Die additive Konstante a gehört grundsätzlich zu jeder linearen multiplen Regressionsgleichung. Für den bisher angenommenen Fall, daß Kriterium und Prädiktoren z-Werte sind, konnte auf diese Konstante verzichtet werden, weil sie in diesem Fall null ist. Das gilt auch noch, wenn wenigstens die Mittelwerte des Kriteriums und der Prädiktoren null sind. Für die Vorhersage eines Kriteriums aus m Prädiktoren gilt allgemein die Formel $a = \overline{y} - \sum_{j=1..m} b_j \cdot \overline{x}_j$. Wenn aber in einer nicht-linearen Schätzgleichung auch Ausdrücke wie das Produkt $X_{i1} \cdot X_{i2}$ auftauchen, dann ist das arithmetische dieser Produkte im allgemeinen selbst dann nicht null, wenn die Mittelwerte der Prädiktoren null sind. Wollen Sie nachweisen, daß das arithmetische Mittel dieser Produkt in diesem Fall die Kovarianz der beiden Prädiktoren ist?

einen zusätzlichen Prädiktor X_{i3} und berechnet eine multiple Regressionsgleichung mit drei Prädiktoren. Dann kann man sehen, ob sich die Güte der Vorhersage - also die multiple Korrelation - durch die Hinzunahme dieses dritten Prädiktors nennenswert erhöht und ob das Vorzeichen der Gewichtszahl b_3 positiv oder negativ ist.

Solche nicht-linearen Schätzgleichungen verwendet man beispielsweise in der Diagnostik, wenn es darum geht, die Gedankengänge, mit denen ein eher intuitiv arbeitender Diagnostiker seine Informationen kombiniert, mit einer nicht-linearen Regressionsgleichung möglichst gut nachzuvollziehen. In solchen Gedankengängen spielen dann oft die oben genannten und- bzw. oder-Verbindungen zweier Merkmalsausprägungen eine Rolle. Der Vorteil des Nachvollziehens solcher Gedankengänge mit einer multiplen Regressionsgleichung besteht darin, daß dabei die optimale Gewichtung solcher Verknüpfungen empirisch gefunden wird; denn natürlich ergeben sich dann wieder diejenigen Gewichtszahlen b_j und diejenige Zahl a, bei denen die Summe der quadrierten Abweichungen zwischen tatsächlichen geschätzten Kriteriums-Werten möglichst klein ist.

In ähnlicher Weise lassen sich natürlich auch die Koeffizienten in Schätzgleichungen wie

$$\hat{Y}_i = b_1 \cdot X_{i1} + b_2 \cdot X_{i2} + b_3 \cdot X_{i1} \cdot X_{i2} + b_4 \cdot X_{i1}{}^2 + b_5 \cdot X_{i2}2 + a$$

oder auch

$$\hat{Y}_i = b_1 \cdot X_{i1} + b_2 \cdot X_{i2} + b_3 \cdot (X_{i1}{}^2 + X_{i2}{}^2)^{½} + a$$

bestimmen - allgemein: Schätzgleichungen, bei denen der Schätzwert eine Linearkombination der Prädiktoren und von Funktionen ist, in die ein Prädiktor oder mehrere eingehen. Diese Funktionen berechnet man und behandelt sie als zusätzliche Prädiktoren in einer linearen multiplen Regressionsgleichung, die nach den üblichen Verfahren erstellt wird.

>>

5) Kennziffern von Summen und Linearkombinationen in mehrdimensionalen Verteilungen

Gegenstand dieses Abschnitts sind einige Eigenschaften die vor allem als Hilfen für mathematische Herleitungen praktisch sind. Sie sind vor allem "zum Nachschlagen" da, also keine unmittelbaren Lernziele und insofern "für Spezialisten" markiert. Das gilt jedoch nicht für ein Beispiel mit unmittelbarer Anwendungsrelevanz: Die Varianz einer Summe.

<<

a) Das arithmetische Mittel einer Linearkombination

Bei den eindimensionalen Verteilungen haben wir festgestellt, daß der Durchschnitt Lineartransformationen "mitmacht". Entsprechendes gilt auch bei Linearkombinationen. Bilden wir (wie z.B. bei der multiplen Regression) eine Linearkombination verschiedener Variablen, so ist das arithmetische Mittel der so ermittelten Werte gleich der Linearkombination der Durchschnitte der