

III	7	13	30	3	29
V	8	9	5	21	39
VIII	15	22	35	24	27
IV	23	31	40	4	25
VII	38	12	34	23	11
II	20	2	33	32	36
VI	17	14	26	19	6
I	37	16	18	10	1

Man hat hier bereits eine Indextafel, geordnet nach den Werten 1 bis 40 des Index zur Basis 7.

Ist die Periode zu lang, so muß man zwar die übrigen Perioden durch Zerlegung berechnen, bei der Ausfüllung der Zeilen genügt es dagegen, einige Anfangsziffern im Schaltwerk einzustellen. So hätte man z. B. (vergl. Nr. 2) beim Modul 17: $\frac{1}{17} = 0,0588235294117647$ und würde nun etwa die Vielfachen von 05882 bestimmen:

05882, 11764, 17646 usw.

Sie zeigen deutlich genug, wie die Perioden von $\frac{2}{17}$, $\frac{3}{17}$ usw. lauten. Man erhält demnach folgende Zusammenstellung:

1	10	15	14	4	6	9	5	16	7	2	8	13	11	3	12
---	----	----	----	---	---	---	---	----	---	---	---	----	----	---	----

und dies ist schon eine Indextafel für die Basis 10. Uebrigens enthält die zweite Hälfte die Ergänzungen der Zahlen der ersten Hälfte auf 17. 20.

Wien, am 29. Juni 1918.

Über die Korrelationsmethode.

Von H. GRAVELIUS in Dresden.

Für eine zunehmende Reihe von Anwendungen der hydrographischen Praxis kommt die aus der Variationsstatistik stammende Korrelationsmethode in Betracht. Sie betrifft, wie bekannt, die Aufgabe, aus einem großen statistischen Material über zusammengehörige Wertepaare zweier Veränderlichen auf den funktionalen Zusammenhang zwischen diesen zu schließen. Auf die Anwendung der Korrelationsmethode in der Meteorologie hat F. Exner vor 10 Jahren, meines Wissens zum ersten Male in Deutschland, in der Meteorologischen Zeitschrift nach einer englischen Quelle aufmerksam gemacht¹⁾. Seine lediglich als Referat gegebene, auf Herleitung und Beweis nicht eingehende Darstellung beschränkt sich aber auf eine Form des »Korrelationskoeffizienten«, die dessen Wesen nicht erkennen läßt und für die Anwendung keineswegs immer bequem ist. Die Exnersche Quelle scheint, nach meinen fruchtlosen Bemühungen um sie, in Deutschland zurzeit nicht zugänglich zu sein. Ich will deshalb die Theorie der Korrelation, deren rationelle Anwendung mir fördernd erscheint, hier auf dem Wege entwickeln, zu dem mich ein spezielles Problem geführt hat, umsomehr als dabei Gesichtspunkte zur Sprache kommen, die Exner fremd geblieben zu sein scheinen und die ich auch bei anderen, die die Methode anwandten, namentlich in Amerika und Schweden, nicht berührt gefunden habe.

¹⁾ Einer freundlichen Mitteilung des Herrn Herausgebers entnehme ich, daß bereits 1899 in Deutschland über die Methode geschrieben worden ist (Duncker, die Variationsstatistik, Leipzig 1898). Das Buch ist aber selbst beim Verlag nicht mehr zu erhalten. Es war ein Abdruck aus Archiv für Entwicklungsmechanik Bd. 8, das aber auch nicht aufzutreiben war. Lediglich einer brieflichen Mitteilung entnahm ich, daß sich in dem Dunckerschen Buch die hier gegebene Entwicklung nicht finde.

Exner hat den Korrelationskoeffizienten so definiert. Es seien zwei Reihen von Beobachtungen gegeben X_i ($i = 1, 2, \dots, n$), Y_i ($i = 1, 2, \dots, n$). Ihre arithmetischen Mittel seien X_0 und Y_0 . Man bilde die Abweichungen

$$x_i = X_i - X_0, y_i = Y_i - Y_0,$$

dann ist

$$r = \frac{\sum x y}{\sqrt{\sum x^2} \sqrt{\sum y^2}}$$

eine zwischen -1 und $+1$ schwankende Zahl; und es ist um so eher ein Zusammenhang (eine Korrelation) zwischen den Erscheinungen X und Y anzunehmen, je größer der absolute Betrag von r ist. Daß r zwischen jenen Grenzen schwankt, ist leicht einzusehen; wie man aber gerade auf diese Form eines Maßes für den Zusammenhang zweier Erscheinungen kommt, wird nicht klar. Es erscheint willkürlich gewählt; und dieser Eindruck wird bestätigt und erhöht durch eine spätere Darlegung Exners, die 1913 erschienen ist (Über die Korrelationsmethode. Jena bei Gustav Fischer). In Wirklichkeit ist aber r gar nichts willkürliches, auch nichts neues, sondern ein ganz alter Bekannter, der sich ganz organisch aus einer sinngemäßen Anwendung der Methode der kleinsten Quadrate ergibt.

1. Ableitung des „Korrelationskoeffizienten“. Hydrographisch-klimatologische Forschungen hatten mich dazu geführt; den Niederschlag in Sachsen durch eine Tabelle mit doppeltem Eingang darzustellen, deren vertikales Argument die Meereshöhe in Stufen von je 50 m und deren horizontales Argument die jährliche Regenmenge in Stufen von je 50 mm ist. Die Tabelle gibt die Anzahl der Stationen, die in der Höhenstufe (X_i, X_{i+1}) eine Regenmenge der Stufe (Y_i, Y_{i+1}) haben.

Es lag nahe, diese Darstellung zur Gewinnung eines Urteils über die Abhängigkeit des Niederschlages von der Meereshöhe zu benutzen. Zu dem Zwecke waren die Mittel der einzelnen Horizontalreihen (d. h. das mittlere Y für eine gegebene Stufe X_i, X_{i+1}) zu bilden. Bei der graphischen Abbildung der Tabelle ergaben die durch diese mittleren Y gegebenen Punkte einen Schwarm, der sich einer Geraden anschloß. Nun wurde die Aufgabe umgekehrt, d. h. es wurden die Mittel der einzelnen Vertikalreihen gebildet (mittleres X zu einer gegebenen Stufe Y_i, Y_{i+1} oder mittlere Meereshöhe einer Regenstufe). Dann ergab sich wieder ein Punktschwarm mit zugeordneter Geraden. Aber beide Geraden fielen nicht zusammen, sondern machten, roh gemessen, einen Winkel von nahezu 10° .

Mit dieser Beobachtung war der Zugang zur »Korrelationsmethode« von selbst gegeben.

Wir wollen eine Tabelle der angegebenen Art graphisch übertragen. Die Punkte 1, 2, ..., n auf den Achsen sind die Endpunkte der Stufen. Wir wollen gleich so transformieren, daß wir durch diejenigen Punkte der Achsen, welche den Mittelwerten der ganzen Beobachtungsreihen X und Y entsprechen, Parallelen X_0, Y_0 zu den anfänglichen Achsen ziehen. Werden diese Geraden X_0, Y_0 nun als Achsen genommen, so ist also jeder Punkt der Ebene der Träger eines Paares von Abweichungen x, y ; und die in unserer Tabelle eingetragenen Zahlen geben die Häufigkeit des Vorkommens eines solchen Paares x, y an.

Die Methode der kleinsten Quadrate zeigt nun sofort, daß die beiden Geraden, die durch die Punktschwärme der Mittelwerte der Reihen und Kolonnen sich legen lassen, durch den Ursprung des gewählten Koordinatensystems gehen, da sowohl $\sum x$ wie $\sum y$ nach Definition verschwinden. Die Normalgleichungen dieser Geraden sind also

$$\sum x y = b_1 \sum y^2, \quad \sum x y = b_2 \sum x^2,$$

also

$$b_1 = \frac{\sum x y}{\sum y^2}, \quad b_2 = \frac{\sum x y}{\sum x^2}.$$

Definieren wir noch die Streuungen σ_1^2 und σ_2^2 durch

$$n \sigma_1^2 = \sum x^2, \quad n \sigma_2^2 = \sum y^2,$$

so ist

$$b_1 = \frac{\sum x y}{n \sigma_2^2}, \quad b_2 = \frac{\sum x y}{n \sigma_1^2}$$

oder auch

$$b_1 = \frac{\sum x y}{n \sigma_1 \sigma_2} \frac{\sigma_1}{\sigma_2}, \quad b_2 = \frac{\sum x y}{n \sigma_1 \sigma_2} \frac{\sigma_2}{\sigma_1} \quad \dots \quad (1).$$

Hier ist nun

$$r = \frac{\sum xy}{n \sigma_1 \sigma_2} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (2).$$

Man hat also für die beiden Geraden

$$x = r \frac{\sigma_1}{\sigma_2} y, \quad y = r \frac{\sigma_2}{\sigma_1} x \quad (3),$$

und die Größen b_1, b_2 hängen mit r durch

$$r^2 = b_1 b_2 \quad (4)$$

zusammen. Der Winkel φ der beiden Geraden (3) ist gegeben durch

$$\operatorname{tg} \varphi = \frac{\frac{1}{b_1} - b_2}{1 + \frac{b_2}{b_1}} = \frac{1 - r^2}{b_1 + b_2} = \frac{1 - r^2}{r} \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} \quad (5).$$

Man erkennt also, daß $0 \leq r \leq 1$, und daß, wenn r sein Maximum erreicht, $\varphi = 0$ wird, also die Geraden (3) zusammenfallen, oder, mit anderen Worten, die Beziehung $x = b_1 y$ umkehrbar ist. In diesem Falle ist also der Zusammenhang zwischen x und y ein vollkommen fester: es besteht ein mathematisches Gesetz.

Für abnehmende r wächst φ und wird für $r = 0$ zu $\varphi = \frac{\pi}{2}$; die Geraden (3) fallen dann mit den Koordinatenachsen zusammen, es besteht gar kein Zusammenhang zwischen den Erscheinungen X und Y . Die Mittelwerte aller Horizontalreihen liegen dann auf der Ordinatenachse, die Mittelwerte aller Vertikalreihen auf der Abszissenachse. (Es läßt sich dies auch unmittelbar aus der Häufigkeitstheorie nachweisen für von einander unabhängige x, y .)

Die Betrachtungsweise, die eben zur Einführung des Korrelationskoeffizienten führte, stellt insofern einen Fortschritt gegen die älteren Betrachtungen dar, als sie deutlich darauf aufmerksam macht, daß eine empirisch hergeleitete Beziehung nicht ohne weiteres umkehrbar ist —, eine Feststellung, die für die Anwendungen von Bedeutung ist.

Die Korrelationsmethode gibt aber die beiden Seiten einer solchen Beziehung, und gibt in r bezw. dem Winkel φ ein Maß für die Annäherung zwischen ihnen.

2. Der mittlere Fehler der Darstellung. Sind nun x, y durch die Ausgleichung gefunden, so leiten sich aus

$$v_1 = x - b_1 y \\ v_2 = y - b_2 x$$

die mittleren Fehler s_1, s_2 der Darstellungen x, y nach der Ausgleichung ab.

Man hat

$$n s_1^2 = \sum v_1^2 = \sum x^2 - 2 b_1 \sum xy + b_1^2 \sum y^2,$$

was nach kurzer Reduktion wird

$$s_1^2 = \sigma_1^2 (1 - r^2) \quad (6)$$

und entsprechend

$$s_2^2 = \sigma_2^2 (1 - r^2) \quad (6).$$

Da nun ferner

$$b_1 = \frac{\sum xy}{\sum y^2}$$

war, so ist nach einem bekannten Satz über das letzte Ergebnis bei der Gaußschen Auflösung der Normalgleichungen, der mittlere Fehler von b_1

$$\sigma_{b_1} = \frac{s_1}{\sqrt{\sum y^2}} = \frac{s_1}{\sigma_2 \sqrt{n}}$$

und entsprechend

$$\sigma_{b_2} = \frac{s_2}{\sigma_1 \sqrt{n}}$$

oder

$$\sigma_{b_1} = \frac{\sigma_1 \sqrt{1 - r^2}}{\sigma_2 \sqrt{n}}, \quad \sigma_{b_2} = \frac{\sigma_2 \sqrt{1 - r^2}}{\sigma_1 \sqrt{n}} \quad (7).$$

3. Zusammenfassung der Ergebnisse. Wir wollen zur bequemerem Uebersicht bei der praktischen Anwendung die gewonnenen Ergebnisse kurz zusammenstellen und

dabei in Hinblick auf die dann vorzunehmende Erweiterung der Aufgabe die Bezeichnungen etwas ändern.

Die als Beobachtungen zugrunde liegenden Abweichungen seien mit x_1 und x_2 bezeichnet. Wir führen dann ein

$$\begin{aligned} p_{12} &= p_{21} = \frac{\sum x_1 x_2}{n}, \quad n \sigma_1^2 = \sum x_1^2, \quad n \sigma_2^2 = \sum x_2^2, \\ r_{12} &= r_{21} = \frac{p_{12}}{\sigma_1 \sigma_2}, \quad b_{12} = \frac{p_{12}}{\sigma_2^2}, \quad b_{21} = \frac{p_{21}}{\sigma_1^2}, \\ b_{12} &= r_{12} \frac{\sigma_1}{\sigma_2}, \quad b_{21} = r_{21} \frac{\sigma_2}{\sigma_1}, \quad r_{12} = \sqrt{b_{12} b_{21}}, \\ x_1 &= b_{12} x_2, \quad x_2 = b_{21} x_1, \\ s_1 &= \sigma_1 \sqrt{1 - r_{12}^2}, \quad s_2 = \sigma_2 \sqrt{1 - r_{21}^2}, \\ \sigma_{b_{12}} &= \frac{\sigma_1}{\sigma_2} \frac{\sqrt{1 - r_{12}^2}}{\sqrt{n}}, \quad \sigma_{b_{21}} = \frac{\sigma_2}{\sigma_1} \frac{\sqrt{1 - r_{21}^2}}{\sqrt{n}}. \end{aligned}$$

Mit diesen Formeln hat man alles beisammen, was zur Darstellung des Zusammenhanges zweier Erscheinungen nötig ist. Es fehlt nur eine Angabe über σ_r , den mittleren Fehler des Korrelationskoeffizienten r . Diese Größe läßt sich aber mit elementaren Betrachtungen nicht erlangen. Wir kommen auf sie in Abschnitt 5 zu sprechen.

Hat man die Beziehungsgleichungen zwischen den Abweichungen x_1 und x_2 erlangt, so läßt sich leicht zu den Erscheinungen selbst übergehen. Denn es ist

$$x_i = X_i - X_i^0,$$

wenn mit X^0 der Mittelwert der betreffenden Beobachtungsreihe bezeichnet wird. Aus

$$x_1 = b_{12} x_2$$

folgt also sofort

$$\begin{aligned} X_1 - X_1^0 &= b_{12} (X_2 - X_2^0) \\ X_1 &= b_{12} X_2 + (X_1^0 - b_{12} X_2^0). \end{aligned}$$

und entsprechend

$$X_2 = b_{21} X_1 + (X_2^0 - b_{21} X_1^0).$$

4. Korrelation „zu Dreien“. Wir wenden uns nun der nächsten Hauptaufgabe zu, der Darstellung einer »Korrelation zu Dreien«, um es kurz auszudrücken. Untersuchungen dieser Art haben besondere Wichtigkeit für die angewandte Klimatologie. Die Ernteergebnisse (x_1) hängen immer von wenigstens zwei klimatologischen Elementen, z. B. Temperatur (x_2) und Niederschlag (x_3) ab. Analoga aus der Hydrographie (Wasserstandsprognose etc.) liegen auf der Hand.

Wenn also die Erscheinungen X_1, X_2, X_3 untereinander zusammenhängen und x_1, x_2, x_3 wieder die Abweichungen $X_1 - X_1^0$ usw. bedeuten, dann bestehen zunächst die drei Beziehungsgleichungen

$$\left. \begin{aligned} x_1 &= b_{123} x_2 + b_{132} x_3 \\ x_2 &= b_{213} x_1 + b_{231} x_3 \\ x_3 &= b_{312} x_1 + b_{321} x_2 \end{aligned} \right\} \dots \dots \dots ().$$

wo das Anordnungsgesetz der Indices klar ist.

Diese Korrelation zu Dreien hat aber zur allgemeinen Voraussetzung, daß auch unter den einzelnen Paaren Beziehung besteht, so daß also noch die Gleichungen bestehen

$$\left. \begin{aligned} x_1 &= b_{12} x_2 & x_2 &= b_{21} x_1 \\ x_1 &= b_{13} x_3 & x_3 &= b_{31} x_1 \\ x_2 &= b_{23} x_3 & x_3 &= b_{32} x_2 \end{aligned} \right\} \dots \dots \dots (9).$$

Die Größen b sind nach der Methode der kleinsten Quadrate zu bestimmen. Die Fehlergleichungen sind

$$\left. \begin{aligned} v_{123} &= x_1 - b_{123} x_2 - b_{132} x_3 \\ v_{213} &= x_2 - b_{213} x_1 - b_{231} x_3 \\ v_{312} &= x_3 - b_{312} x_1 - b_{321} x_2 \\ v_{12} &= x_1 - b_{12} x_2 & v_{21} &= x_2 - b_{21} x_1 \\ v_{13} &= x_1 - b_{13} x_3 & v_{31} &= x_3 - b_{31} x_1 \\ v_{23} &= x_2 - b_{23} x_3 & v_{32} &= x_3 - b_{32} x_2 \end{aligned} \right\} \dots \dots \dots (10)$$

und somit die Normalgleichungen

$$\left. \begin{aligned} \sum x_2 v_{123} &= 0 & \sum x_3 v_{123} &= 0 & \sum x_2 v_{12} &= 0 & \sum x_1 v_{21} &= 0 \\ \sum x_1 v_{213} &= 0 & \sum x_3 v_{213} &= 0 & \sum x_3 v_{13} &= 0 & \sum x_1 v_{31} &= 0 \\ \sum x_1 v_{312} &= 0 & \sum x_2 v_{312} &= 0 & \sum x_3 v_{23} &= 0 & \sum x_2 v_{32} &= 0 \end{aligned} \right\} \quad (11),$$

oder allgemein geschrieben

$$\sum x_i v_{hik} = 0 \quad \sum x_i v_{hi} = 0 \quad \sum x_k v_{hik} = 0 \quad \dots \quad (11a)$$

(der Faktor x hat niemals den ersten Index des Faktors v).

Diese Normalgleichungen wollen wir nun nicht nach der gewöhnlichen Methode explicite aufstellen und auflösen, sondern beachten, daß alle Summen von der Form

$$\sum v_{ik} v_{hik} = 0$$

sind, so daß aus der Entwicklung einer solchen Summe unter steter Berücksichtigung der Gleichungen (11) bzw. (11a) sich Relationen für die Größen b_{hik} gewinnen lassen müssen. Die Bestimmung der b_{hi} kann nach früherem als erledigt gelten, wird uns also nicht mehr beschäftigen.

Wir haben nun

$$0 = \sum v_{ik} v_{hik} = \sum (x_i - b_{ik} x_k) (x_h - b_{hik} x_i - b_{hki} x_k) = \sum (x_h v_{ik} - b_{hik} x_i v_{ik}) \quad (a);$$

da aber

$$\sum x_k v_{ih} = 0, \quad \sum x_k v_{hk} = 0,$$

so kann man in der vorhergehenden Gleichung (a)

$$\begin{aligned} x_h &\text{ durch } v_{hk} = x_h - b_{hk} x_k \\ x_i &\text{ » } v_{ik} = x_i - b_{ik} x_k \end{aligned}$$

ersetzen, wodurch wird

$$\sum v_{ik} v_{hik} = \sum v_{ik} v_{hk} - b_{hik} \sum v_{ik}^2,$$

also, da die linke Seite Null ist,

$$b_{hik} = \frac{\sum v_{ik} v_{hk}}{\sum v_{ik}^2} \quad \dots \quad (12).$$

Vertauschung von h und i ändert den Zähler nicht. Wir schreiben also

$$b_{ihk} = \frac{\sum v_{ik} v_{hk}}{\sum v_{hk}^2} \quad \dots \quad (12a).$$

Definieren wir

$$\begin{aligned} n \sigma_{ik}^2 &= \sum v_{ik}^2, \quad n \sigma_{hk}^2 = \sum v_{hk}^2 \\ \frac{\sum v_{ik} v_{hk}}{n} &= p_{hik} = p_{ihk} \quad \dots \quad (13), \end{aligned}$$

so wird

$$\begin{aligned} b_{hik} &= \frac{p_{hik}}{\sigma_{ik}^2} = \frac{p_{hik} \sigma_{hk}}{\sigma_{ik} \sigma_{hk} \sigma_{ik}} \\ b_{ihk} &= \frac{p_{ihk}}{\sigma_{ik}^2} = \frac{p_{ihk} \sigma_{ik}}{\sigma_{ik} \sigma_{hk} \sigma_{hk}} \quad \dots \quad (14). \end{aligned}$$

Setzt man jetzt

$$r_{hik} = \frac{p_{hik}}{\sigma_{ik} \sigma_{hk}}, \quad r_{ihk} = \frac{p_{ihk}}{\sigma_{ik} \sigma_{hk}} \quad \dots \quad (15),$$

wo nach (13)

$$r_{hik} = r_{ihk} \quad \dots \quad (16),$$

so wird

$$b_{hik} = r_{hik} \frac{\sigma_{hk}}{\sigma_{ik}}, \quad b_{ihk} = r_{ihk} \frac{\sigma_{ik}}{\sigma_{hk}}, \quad r_{hik}^2 = b_{hik} b_{ihk} \quad \dots \quad (17).$$

Nun war

$$n \sigma_{ik}^2 = \sum v_{ik}^2$$

$$\sum v_{ik}^2 = \sum (x_i - b_{ik} x_k) (x_i - b_{ik} x_k) = \sum x_i^2 - b_{ik} \sum x_i x_k,$$

aber

$$\sum x_i^2 = n \sigma_i^2 \sum x_i x_k = n b_{ki} \sigma_i^2;$$

daher ist

$$\begin{aligned} n \sigma_{ik}^2 &= n \sigma_i^2 - n b_{ik} b_{ki} \sigma_i^2, \\ \sigma_{ik}^2 &= \sigma_i^2 (1 - r_{ik}^2) \quad \dots \quad (18), \end{aligned}$$

und entsprechend

$$\sigma_{ki}^2 = \sigma_k^2 (1 - r_{ik}^2) \quad \dots \quad (18a).$$

Vorhin hatten wir (Gl. (12))

$$\sum v_{hk} v_{ik} = b_{hik} \sum v_{ik}^2 \quad \dots \quad (b).$$

Mit Rücksicht auf Gl. (11a) ist aber

$$\sum v_{hk} v_{ik} = \sum x_i x_h - b_{hk} \sum x_i x_k,$$

wodurch (b) wird

$$\begin{aligned} p_{ih} - b_{hk} p_{ik} &= b_{hk} \sigma_{ik}^2; \\ \text{aber} \quad p_{ih} &= b_{hi} \sigma_i^2, \quad p_{ik} = b_{ki} \sigma_i^2, \\ \sigma_{ih}^2 &= \sigma_i^2 (1 - r_{ik}^2), \quad r_{ik}^2 = b_{ik} b_{ki}, \\ \text{also endlich} \quad b_{hk} &= \frac{b_{hi} - b_{hk} b_{ki}}{1 - b_{ik} b_{ki}} \dots \dots \dots (18). \end{aligned}$$

Nun hatten wir gefunden

$$\begin{aligned} b_{h+k} &= r_{h+k} \frac{\sigma_{hk}}{\sigma_{ik}} \\ b_{hk} &= r_{hk} \frac{\sigma_h}{\sigma_k}, \quad b_{ki} = r_{ki} \frac{\sigma_k}{\sigma_i}, \quad b_{hi} = r_{hi} \frac{\sigma_h}{\sigma_i}, \end{aligned}$$

so daß man zunächst aus (18) erhält

$$r_{h+k} = \frac{r_{hi} - r_{hk} r_{ki} \frac{\sigma_h \sigma_k}{\sigma_i \sigma_{hk}}}{1 - r_{ik}^2}$$

und wegen

$$\sigma_{ik} = \sigma_i \sqrt{1 - r_{ik}^2}, \quad \sigma_{hk} = \sigma_h \sqrt{1 - r_{hk}^2}$$

endgültig

$$r_{h+k} = \frac{r_{hi} - r_{hk} r_{ki}}{\sqrt{1 - r_{ik}^2} \sqrt{1 - r_{hk}^2}} \dots \dots \dots (19).$$

5. Der mittlere Fehler bei Korrelation „zu Dreien“. Es erübrigt nun noch die Darstellung des mittleren Fehlers einer Beobachtung x_h nach der Ausgleichung. Zu dem Zweck betrachten wir

$$\sum v_{hik}^2 = n \sigma_{hik}^2.$$

Unter wiederholter Anwendung der Beziehungen (11a) (Normalgleichungen) erhalten wir

$$\begin{aligned} \sum^2 v_{hik} &= \sum x_h v_{hik} = \sum v_{hi} v_{hik} = \sum x_h v_{hi} - b_{hki} \sum x_k v_{hi} \\ &= \sum x_h v_{hi} - b_{hki} \sum v_{hi} (x_{ki} - b_{ki} x_i) = \sum x_h v_{hi} - b_{hki} \sum v_{hi} v_{ki} = \sum v_{hi}^2 - b_{hki} \sum v_{hi} v_{ki}. \end{aligned}$$

Nun ist aber (Gl. (12))

$$\begin{aligned} \sum v_{hi} v_{ki} &= b_{hki} \sum v_{hi}^2, \\ \text{also} \quad \sum v_{hik}^2 &= \sum v_{hi}^2 - b_{hki} b_{hki} \sum v_{hi}^2 = (1 - b_{hki} b_{hki}) \sum v_{hi}^2 = (1 - r_{hki}^2) \sum v_{hi}^2 \\ \sigma_{hik}^2 &= \sigma_{hi}^2 (1 - r_{hki}^2) \end{aligned}$$

oder endlich, wegen

$$\begin{aligned} \sigma_{hi}^2 &= \sigma_h^2 (1 - r_{hi}^2) \\ \sigma_{hik}^2 &= \sigma_h^2 (1 - r_{hi}^2) (1 - r_{hki}^2) \dots \dots \dots (20). \end{aligned}$$

Nach Gl. (12) ist der Gewichtungsfaktor für b_{hik}

$$\frac{1}{n \sigma_{ik}^2} = \frac{1}{n \sigma_i^2 (1 - r_{ik}^2)},$$

so daß man erhält

$$b_{hik} = \frac{\sigma_h}{\sigma_i \sqrt{n}} \frac{\sqrt{(1 - r_{hi}^2) (1 - r_{hki}^2)}}{\sqrt{1 - r_{ik}^2}} \dots \dots \dots (21).$$

Die Darstellung ist absichtlich auf den, soweit ich sehe, zunächst allein für die gewässerkundliche Arbeit in Betracht kommenden Fall von drei Variablen beschränkt worden. Man sieht aber, wie man auf dem gleichen Wege einer zweckmäßigen Benutzung der Eigenschaften der v zur Verallgemeinerung für beliebig viele Elemente x gelangen kann.

Es möge aber noch eine kurze Notiz angefügt werden, die sich auf den, offenbar nach englischem Vorbild, in den Anwendungen vielfach gegebenen, »wahrscheinlichen Fehler« von r bezieht. Man hat

$$r = \frac{1}{n \sigma_1 \sigma_2} \sum x y.$$

Der mittlere Fehler σ_r läßt sich nun, soweit ich sehe, nur durch eine tiefer gehende und auch viel Raum in Anspruch nehmende fehlertheoretische Untersuchung erhalten. Wenigstens habe ich nur diesen Weg gefunden. Auf seine Wiedergabe in diesem Zusammenhang muß daher wohl verzichtet werden. Vielleicht ist es zulässig, in einem besonderen Aufsatz darauf einzugehen.

Aber es kann auf elementarem Wege wenigstens eine Plausibelmachung des Ausdrucks für σ_r erreicht werden, die freilich ihre Schwächen hat und nicht den Anspruch

erhebt, ein Beweis zu sein. Es mag nämlich r als eine lineare Funktion der x_i mit den Koeffizienten y_i angesehen werden. Dann ist

$$\sigma_r^2 = \frac{1}{n^2 \sigma_1^2 \sigma_2^2} (y_1^2 \sigma_{x_0}^2 + \dots + y_n^2 \sigma_{x_n}^2).$$

Die σ_x sind aber untereinander gleich und zwar

$$\sigma_x = s_1 = \sigma_1 \sqrt{1 - r^2},$$

der mittlere Fehler eines x nach der Ausgleichung. Also wird

$$\sigma_r^2 = \frac{1}{n^2 \sigma_1^2 \sigma_2^2} s_1^2 \sum y^2,$$

wo nun $\sum y^2$ wieder nichts anderes ist als die Summe der Quadrate der Abweichungen y nach der Ausgleichung, d. h.

$$\sum y^2 = n s_2^2.$$

Damit

$$\sigma_r^2 = \frac{(1 - r^2)^2}{n},$$

dies ist aber nun tatsächlich der in der Literatur auftretende Wert, denn es wird als »wahrscheinliche Fehler des Korrelationskoeffizienten r « angegeben

$$f = 0,676 \frac{1 - r^2}{\sqrt{n}}.$$

Halten wir aber fest, daß das hier gegebene kein Beweis ist, da wir soeben stillschweigend »für einen Augenblick« angenommen, die y seien unabhängig von den x . Es ist übrigens selbstverständlich, daß das gleiche Ergebnis erlangt worden wäre, wenn $\sum x y$ als lineare Funktion der y mit den Koeffizienten x aufgefaßt worden wäre.

Ich komme aber hier nur deshalb auf dieses σ_r zu sprechen, weil ihm eine Bedeutung zugesprochen worden ist, die es nicht hat. Es wird von einem Autor zum andern der Grundsatz übernommen, eine Korrelation dürfe als befriedigend festgestellt gelten, wenn

$$r : f = 6$$

sei. In seinem trefflichen Buch über Wetter und Wettervorhersage spricht sich Defant sogar dahin aus, daß ein Verhältnis $r : f = 5$ genüge, »um das Bestehen einer Beziehung zwischen zwei Elementen zu beweisen«.

Eine solche Auffassung von der Bedeutung des Verhältnisses $r : f$ für die Bewertung einer Korrelation erscheint mir mißverständlich. Nicht dieser Quotient, sondern der absolute Wert von r ist maßgebend für die Enge oder Stärke eines Zusammenhanges. Man scheint zu übersehen, daß

$$\frac{\sigma_r}{r} = \frac{1 - r^2}{r \sqrt{n}}$$

noch von n abhängig ist. Der vorhin erwähnte kritische Wert $r : f = 5$ entspricht in runder Zahl dem Verhältnis $\sigma_r : r = 0,3$. Legen wir dies zugrunde, so ergeben sich folgende zusammenhängende Werte von n und r

$n = 36$	100	400
$r = 0,445$	0,303	0,160.

Man wird nun schwerlich eine Korrelation für $r = 0,16$ noch als diskutabel erachten wollen, selbst wenn $\sigma_r : r$ den angeblich beweisenden Wert 0,3 hat. Und die Zahl von 400 Elementenpaaren läßt sich leicht genug erreichen: Denn wir können doch, wenn wir zu neuen Einsichten gelangen wollen, nicht immer an den Monats- und Jahresmitteln kleben bleiben bei der hydrographischen und meteorologischen Forschung, sondern müssen vielmehr die Einzelbeobachtung als Element einführen.

Die Berechnung von σ_r und seine Wiedergabe in der Form von f ist zwar üblich geworden, hat aber wenig schlüssigen Wert. Es ist allein der absolute Betrag von r , auf den es hier ankommt. An den idealen Wert $|r| = 1$ wird man sich bei der Bearbeitung empirischen Materials freilich nicht klammern, aber auch nicht zu weit von ihm abweichen. Man beachte, daß für $r = 0,5$ und den günstigen Fall $\sigma_1 = \sigma_2$

$$\operatorname{tg} \varphi = 0,75 \quad \varphi = 37^\circ$$

wird, also eine schon sehr beträchtliche Divergenz der beiden darstellenden Geraden sich ergibt.