Zeitschrift: Mitteilungen der Thurgauischen Naturforschenden Gesellschaft
Herausgeber: Thurgauische Naturforschende Gesellschaft
Band: 32 (1940)

Artikel: Korrelationsrechnung
Autor: Früh, E.
DOI: https://doi.org/10.5169/seals-593767

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 22.01.2026

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Korrelationsrechnung

Von **E. Früh**, Kradolf

Streuung und Abhängigkeit

Ist eine Größe p so definiert, daß ihr zwischen den Grenzen a und b ein wahrscheinlichster Wert m_p zukommt, dann muß bei n Messungen die Häufigkeit eines Wertes p_i um so größer sein, je kleiner $p_i - m_p$ ist. Nach dem Gaußschen Verteilungsgesetz ist m_p so zu bestimmen, daß

$$n\sigma^2 = \Sigma (p_i - m_p)^2$$

zu einem Minimum wird. Diese Bedingung ist erfüllt, wenn

$$m_p = \Sigma \frac{p_i}{n}$$

also gleich dem arithmetischen Mittel aller p_i-Werte ist. Es leuchtet ein, daß bei gegebener Meßgenauigkeit nur bei hinreichend großem n ($n \rightarrow \infty$) der „wahre" Mittelwert erhalten werden kann; das aus einer beschränkten Zahl von Messungen sich ergebende arithmetische Mittel ist als immerhin bester Annäherungswert noch mit einem von n abhängigen Fehler behaftet.

Sind $v_i = p_i - m_p$ die wahren Fehler, oder Abweichungen, bezogen auf das wahre Mittel m_p, dann wird σ als mittlerer Fehler, mittlere Abweichung oder auch mittlere Streuung bezeichnet. Ist m'_p das angenäherte Mittel, dann erhält man als mittlere Streuung, oder kurz Streuung

$$\sigma = \sqrt{\frac{\Sigma (p_i - m'_p)^2}{n-1}} = \sqrt{\frac{[v\,v]}{n-1}},$$

wenn unter [vv] nach Gauß die Summe aller Fehlerquadrate verstanden wird. Der Fehler des Mittelwertes wird dann

$$\sigma_m = \frac{\sigma}{\sqrt{n}}$$

10741
12JJJJ

(vgl. S. 59). Ist für einen Zahlenkomplex nicht ein wahres Mittel definiert, dann kann auch die Streuung nicht auf ein solches bezogen werden. Es ist in diesem Fall

$$\sigma = \sqrt{\frac{[vv]}{n}}$$

Weist zum Beispiel ein von a bis b reichendes Intervall keine Häufungsstellen der Werte auf, dann ist das Mittel nur von den Intervallgrenzen abhängig. Die Dichte der Besetzung des Intervalls spielt bei der Berechnung des Mittels keine Rolle; die Streuung strebt mit wachsendem n einem Grenzwert zu:

$$\sigma_\infty = \frac{a - b}{6} \sqrt{3},$$

ist also bei homogener Besetzung des Intervalls ebenfalls unabhängig von n. — Vor der Berechnung einer Streuung haben wir also stets zu untersuchen, auf was für eine Art von Mittel diese zu beziehen ist.

Liegen zwei innerhalb beliebiger Grenzen veränderliche Größen y und x vor, zwischen denen durch das Experiment, oder auf dem Wege der Erfahrung eine Bindung irgendwelcher Art festgestellt worden ist, dann wird einem bestimmten Wert der einen Größe eine innerhalb eines gewissen Bereiches bleibende Wertegruppe der andern Größe zuzuordnen sein. Zu einem $x = X_i$ gehören also n_i Werte y_i, deren Mittelwert m_y^i und Streuung, bezogen auf m_y^i gewissermaßen durch X_i bedingt sind. Ebenso wird für $y = Y_i$ die Größe x innerhalb eines beschränkten Bereiches variieren und somit die Streuung und der Mittelwert m_x^i der x_i-Werte als durch Y_i bedingt zu bezeichnen sein. Damit ist aber über die Art der Beziehung zwischen y und x nicht eine Aussage gemacht so, als ob die eine der Größen eine unabhängig, die andere eine abhängig Veränderliche sei und man weitere Möglichkeiten, wie etwa die der Beeinflussung beider Größen durch dritte auszuschließen habe. Die Ausdrücke: bedingter Streuungsbereich, bedingtes Mittel, usw. haben lediglich den Sinn einer durch die Existenz einer Bindung gerechtfertigten Zuordnung.

Wir haben Zusammenhänge zu unterscheiden:

A. Nach der Art

a. Kausalzusammenhänge: Beziehungen zwischen Grund und Folge, zwischen Ursache und Wirkung. Nach dem Kausalitätsgesetz ist jede Veränderung durch eine andere bedingt; Beziehungen dieser

Art feststellen heißt nichts anderes, als (Natur-) Gesetze entdecken. Es leuchtet ein, daß diese Zusammenhänge außerordentlich wichtig sind.

b. Symptomatische Zusammenhänge. Folgen, die auf die nämliche Ursache zurückzuführen sind. Sind Ereignisse das Ergebnis der Einflußnahme eines übergeordneten Erscheinungskomplexes, dann wird man unter Umständen auch bei Unkenntnis desselben vom Eintreffen des einen Ereignisses auf das Eintreffen des oder der andern schließen können.

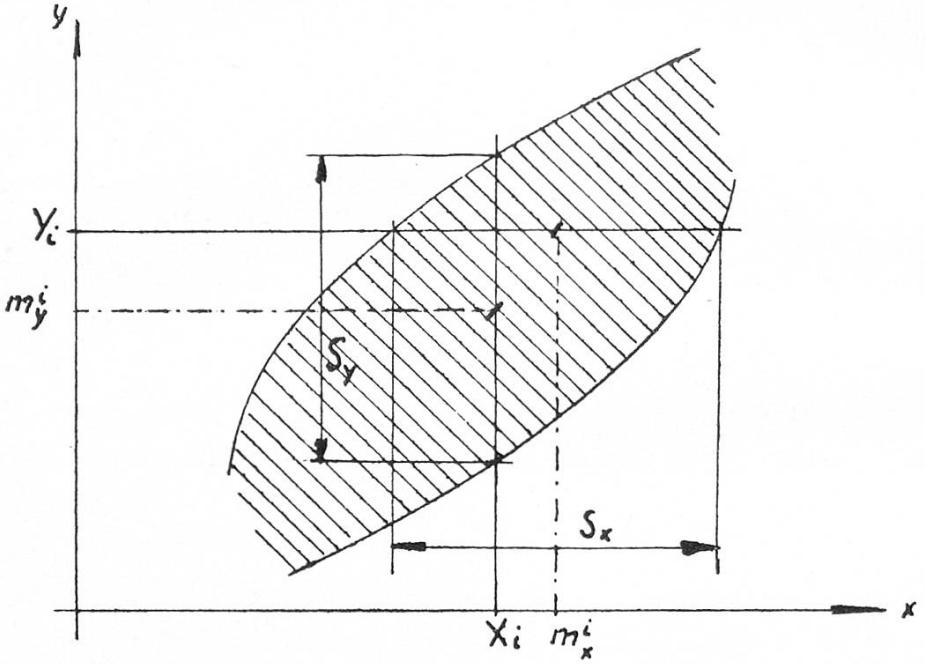

Fig. 1

S_y: Streuungsbereich von y für $x = X_i$
S_x: Streuungsbereich von x für $y = Y_i$
S_x setzt S_y voraus und umgekehrt.

B. Nach der Enge der Bindung

a. Funktionelle Zusammenhänge: Zu bestimmten Werten einer oder mehrerer bedingenden Größen gehören bestimmte Werte der bedingten Größe. Eine Streuung ist also weder bei abhängig, noch bei unabhängig Veränderlichen vorhanden. Funktionen sind durchaus nicht nur rein theoretische Zuordnungen; man kennt eine ganze Reihe empirisch festgestellter funktioneller Beziehungen, obwohl ja zugegeben werden muß, daß alles nicht a priori Gültige nur mit einer gewissen Wahrscheinlichkeit ausgesprochen werden kann. Diese Wahrscheinlichkeit ist aber von völliger Sicherheit kaum zu unterscheiden.

b. Korrelative Zusammenhänge (Ko-Relation = Mitbeziehung). Die Streuung von y (x, x_i) ist eine durch x oder x_i (y) bedingte. (x_i, $i = 1 - n$ bedeutet, daß y unter Umständen nicht nur von einem einzigen x abhängig ist.) Je nach der Größe der bedingten Streuung muß die Abhängigkeit als mehr oder weniger eng bezeichnet werden. Die Untersuchung solcher korrelativer Beziehungen macht es notwendig, diese rein qualitative Aussage durch eine quantitative zu ersetzen: Die Abhängigkeit muß gemessen werden können. Das Verdienst, ein Abhängigkeitsmaß geschaffen und damit die Korrelationsrechnung begründet zu haben, kommt dem Engländer Galton zu („Correlations and their measurement" 1888).

Das Abhängigkeitsmaß

Theoretisch muß es möglich sein, jede an einer Größe y auftretende Veränderung auf Veränderungen einer Reihe von x_i-Werten zurückzuführen, von denen y funktionell abhängig ist:

$$y = f\ (x_1,\ x_2,\ \ldots\ x_n).$$

Von den n möglichen Teilbeziehungen zwischen

y und x_1, y und x_2, usw.

wird der zwischen y und einem bestimmten $x = x_i$ am ehesten festzustellen sein, wenn man alle übrigen Größen, die von Einfluß sind, konstant hält. Das ist nun vielfach nicht möglich und dieser Umstand bedingt den korrelativen Charakter solcher Teilbeziehungen. Anders gesagt: Korrelative Beziehungen zwischen zwei Größen lassen auf andere, vielleicht unbekannte Einflüsse schließen. Eine Voraussetzung, auf der die Korrelationsrechnung grundsätzlich aufbaut, kann unter Umständen einen Zusammenhang als loser kennzeichnen, als er möglicherweise ist: sie nimmt eine lineare Beziehung zwischen x und y an. Das schränkt die Anwendungsmöglichkeit der Korrelationsrechnung nicht in so großem Maße ein, wie man es vielleicht glauben möchte; abgesehen davon, daß lineare Beziehungen sehr häufig sind, hat man oft die Möglichkeit, durch Logarithmierung nicht lineare Zusammenhänge in lineare umzuwandeln. Ausdrücke von der Form:

$$y^c = ax^b$$

$$y = a^x, \text{ usw.}$$

das heißt Zusammenhänge, die sich in dieser Weise ausdrücken lassen, sind also ohne weiteres auswertbar. Dazu kommt nun noch eine weitere Methode, die an dieser Stelle etwas eingehender dargestellt werden soll.

Es sei festgestellt worden — vielleicht mit Hilfe der graphischen Darstellung —, daß zwischen den einander zugeordneten Größen x und y eine Beziehung bestehe, die vielleicht am besten durch eine Funktion n-ten Grades:

$$y = a + bx + cx^2 + dx^3 + \ldots + kx^n$$

dargestellt werden könnte. Untersucht man nun, in welcher Weise y wächst, wenn man x immer um gleiche Beträge zunehmen läßt, dann stellt man fest, daß die „Steigung" von y wieder durch eine Funktion von x ausgedrückt werden kann, wobei aber als höchste Potenz die (n — 1)-te von x auftritt. Von dieser Funktion ermitteln wir wieder die „Steigung"; sie ergibt eine Funktion mit x^{n-2} als höchster Potenz — und wenden dieses Verfahren solange an, bis eine k-te Differenzenreihe nur noch eine lineare Änderung zeigt. Man strebt mit dem geschilderten Verfahren eine Beziehung an, die durch k-fache Differentiation der Funktion f(x) erhalten werden kann, wenn gilt:

$$f^{(k)} = a' + b'x \qquad \text{(vgl. S. 51—56).}$$

Hat man mit Hilfe der Korrelationsrechnung den Zusammenhang zwischen $f^{(k)}$ und x festgestellt, dann kann man rückschreitend durch Integration die ursprüngliche Funktion, das heißt jene, welche dem festgestellten Zusammenhang am besten gerecht wird, wieder erhalten.

Das numerische Verfahren ist nun folgendes: Wir bilden die Differenzen aufeinanderfolgender Werte von y, die x-Werten zugehören, zwischen denen das Intervall stets dasselbe bleibt. Diese regelmäßige Zunahme von x kann im praktischen Fall durch eine passende Klasseneinteilung erreicht werden, wobei dann an Stelle eines y der Mittelwert m_y^i aller der der Klasse angehörenden y_i tritt. Die Differenzbildung wird nun solange durchgeführt, bis man annähernd konstante Werte erhält, oder aber, bis diese Differenzwerte weder eine mit x parallel gehende Zunahme, noch eine Abnahme zeigen. An dieser Stelle setzt dann die Korrelationsrechnung ein. Ein Beispiel soll das Gesagte erläutern.

x	y	y'	y''	y'
0	2			4
		7		
1	9		6	10
		13		
2	22		6	16
		19		
3	41		6	22
		25		
4	66		6	28
		31		
5	97		6	34
		37		
6	134			40

Der Einfachheit halber wurde ein funktioneller Zusammenhang gewählt. Da die zweite Differenz y'' den konstanten Wert 6 hat, ist leicht zu berechnen, was für Werte y' für x = 0, 1, 2, ... annimmt. Sie sind in der hintersten Kolonne angegeben.

Man findet:

$$y' = 4 + 6x$$

und daraus

$$y = 2 + 4x + 3x^2$$

Wie wir sehen, ist die Beziehung zwischen x und y' linear. Hier hätte man also für unbestimmt variable Werte von y'' mit Hilfe der Korrelationsrechnung den Grad der Abhängigkeit zu bestimmen. So würde folgendes Beispiel die Anwendung der Korrelationsrechnung notwendig machen:

x	y	y'	y''	y'''
0	1,8			
		7,4		
1	9,2		5,0	
		12,4		2,2
2	21,6		7,2	
		19,6		—2,3
3	41,2		4,9	
		24,5		—2,7
4	65,7		7,6	
		32,1		—4,0
5	97,8		3,6	
		35,7		5,0
6	133,5		8,6	
		44,3		
7	177,8			

Bei y'' finden wir bereits Werte, die keinen Zusammenhang mit x mehr erkennen lassen. Wir haben also y' mit x in Beziehung zu setzen. Da die y' Werte der ersten Differenzreihe sind, haben wir ihnen als Argumente zuzuordnen: 0,5, 1,5, 2,5 usw., so daß das der Korrelationsrechnung zugrunde zu legende Zahlenmaterial folgendes wäre:

x:	0,5	1,5	2,5	3,5	4,5	5,5	6,5
y':	7,4	12,4	19,6	24,5	32,1	35,7	44,3

Wir werden später noch darauf zurückkommen. Es genügt vorläufig, gezeigt zu haben, daß dann nichtlineare Zusammenhänge, die durch Logarithmierung nicht in eine für die Korrelationsrechnung auswertbare Form gebracht werden können, einer Untersuchung zugänglich zu machen sind, wenn es durch Bildung von Differenzreihen gelingt, eine lineare Beziehung zu schaffen.

Bei geeigneter Koordinatenwahl läßt sich eine lineare Bindung zwischen den beiden Größen y und x in folgender Weise darstellen:

$$\frac{y}{x} = c;$$

sie ist also durch proportionale Änderung von y und x gekennzeichnet. In einem rechtwinkligen Koordinatensystem entspricht der Gl. $y = cx$ eine durch den Nullpunkt des Systems führende Gerade. Für die Ableitung des Abhängigkeitsmaßes, oder, wie es auch genannt wird, des Korrelationskoeffizienten, benützen wir eine andere bildliche Darstellung, welche die Bedeutung der Streuung deutlicher erkennen läßt. Setzen wir

$$c = \frac{b}{a},$$

dann ergibt sich die Porportion

$$y : x = b : a$$

aus folgender Figur:

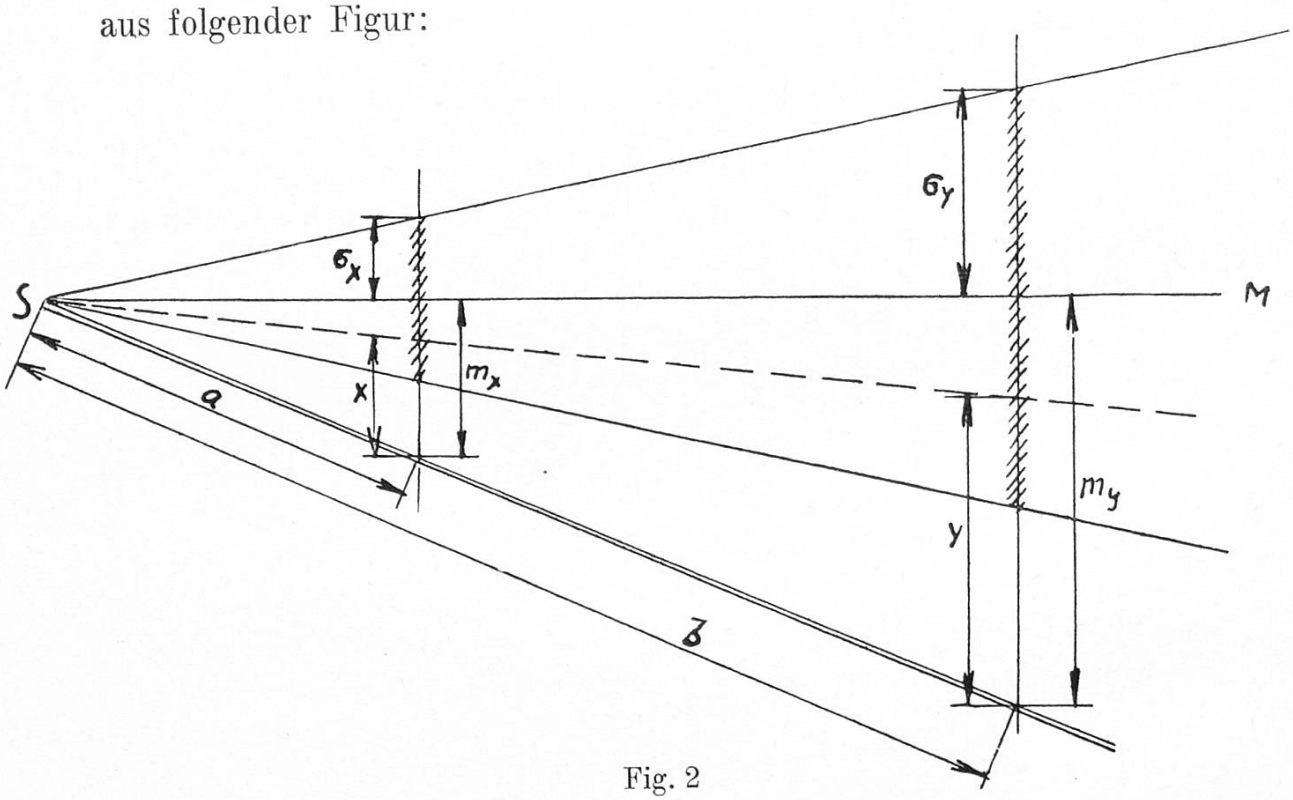

Fig. 2

Jeder von S ausgehende Strahl grenzt auf den Senkrechten durch A und B (es brauchen keine Senkrechte zu sein; sie wurden nur gewählt, um die Symmetrie der Streuung in bezug auf SM besser darstellen zu können) zu SM Strecken x und y ab, die die gegebene Proportion erfüllen. m_x und m_y sind Mittelwerte eines von x_1 bis x_2

(y_1 bis y_2) reichenden Intervalls. Der Streuungsbereich ist durch Schraffierung hervorgehoben; man sieht — was übrigens noch bewiesen wird —, daß

$$y : x = \sigma_y : \sigma_x \text{ ist.}$$

Denkt man sich beispielsweise von einem in S aufgestellten Gewehr, dessen Visierlinie mit SM zusammenfällt, Schüsse auf in A und B senkrecht zu SM aufgestellte Scheiben abgegeben, dann werden die Schußbilder in diesen Scheiben Streuungen aufweisen, deren Querschnittsmaße annähernd im Verhältnis a : b stehen. Die Streuung — als Längenmaß — ist proportional dem Abstand von S, hängt also in gleicher Weise von diesem Abstand ab, wie x oder y, so daß die oben angeführte Beziehung eingentlich ohne weiteres verständlich ist.

Teilt man den Intervallbereich von x_1 bis x_2 in 2n Teile ein, so daß also ein einzelner Teil

$$\Delta x = \frac{x_2 - x_1}{2n} = \frac{m_x - x_1}{n}$$

ist und die Abstände vom Mittelwert absolut genommen der Reihe nach

$$1\Delta x,\ 2\Delta x,\ 3\Delta x,\ \ldots \text{ usw.}$$

sind, dann erhalten wir

$$\sigma_x^2 = \frac{1}{n} \Sigma \Delta x_i^2 = \frac{1}{n} \cdot \frac{(m_x - x_1)^2}{n^2} \cdot (1^2 + 2^2 + \ldots)$$

$$= \frac{1}{n} \cdot \frac{(m_x - x_1)^2}{n^2} \cdot \frac{n}{6} \cdot (n + 1) \cdot (2n + 1)$$

$$\text{und} \quad \sigma_x = \frac{m_x - x_1}{n} \cdot \sqrt{\frac{(n + 1)\ (2n + 1)}{6}}$$

Entsprechend erhält man für

$$\sigma_y = \frac{m_y - y_1}{n} \cdot \sqrt{\frac{(n + 1)\ (2n + 1)}{6}}$$

und daraus

$$\frac{\sigma_y}{\sigma_x} = \frac{m_y - y_1}{m_x - x_1} = \frac{b}{a}$$

Die Ableitung des Korrelationskoeffizienten r ergibt sich nun aus folgendem Ansatz:

$$\frac{\Delta y}{\Delta x} = r \frac{\sigma_y}{\sigma_x} \qquad 1)$$

Der Sinn von r ist also folgender: Je mehr sich der Wert des Abhängigkeitsmaßes der Zahl 1 nähert, um so mehr nähert sich der Zusammenhang zwischen x und y einem streng linearen. Starke Abweichungen von 1 zeigen also nicht immer lose Zusammenhänge, sondern oft auch nichtlineare an.

Die Ableitung des Abhängigkeitsmaßes ist nun leicht möglich auf Grund der Bedingung, daß die Summe aller Fehlerquadrate, die man durch die Addition aller Gleichungen

$$\left(\frac{\Delta y_i}{\sigma_y} - r \frac{\Delta x_i}{\sigma_x}\right)^2 = v_i^2 \qquad i = 1, 2, \ldots n$$

erhält, zu einem Minimum wird. Differenzieren wir

$$\frac{\Sigma \Delta y_i^2}{\sigma_y^2} - 2r \frac{\Sigma \Delta x_i \Delta y_i}{\sigma_x \sigma_y} + r^2 \frac{\Sigma \Delta x_i^2}{\sigma_x^2} = \Sigma v_i^2,$$

oder

$$n - 2r \frac{\Sigma \Delta x_i \Delta y_i}{\sigma_x \sigma_y} + n r^2 = \Sigma v_i^2$$

nach r, dann ergibt sich

$$r = \frac{\Sigma \Delta x_i \Delta y_i}{n \sigma_x \sigma_y} \qquad 2)$$

Die Grenzen für r erhalten wir aus der Minimumsbedingung: Man findet

$$1 - r^2 = \frac{1}{n} \Sigma v_i^2 \qquad 3)$$

Da die rechte Seite positiv ist, muß es auch die linke sein; $|r|$ kann also nur Werte zwischen 0 und 1 annehmen. Wegen

$$\sigma_x = \sqrt{\frac{\Sigma \Delta x^2}{n}} \text{ und } \sigma_y = \sqrt{\frac{\Sigma \Delta y^2}{n}}$$

kann man dem Abhängigkeitsmaß auch folgende Form geben:

$$r = \frac{\Sigma \Delta x \Delta y}{\sqrt{\Sigma \Delta x^2 \Sigma \Delta y^2}} \qquad 4)$$

Bezeichnenderweise ergibt sich derselbe Wert für r aus dem Ansatz

$$\frac{\Delta x}{\Delta y} = r \frac{\sigma_x}{\sigma_y} \qquad 5)$$

Da er, wie der erste, zur Bildung einer Beziehungsgleichung zwischen Δx und Δy führt:

$$\Delta x = r \frac{\sigma_x}{\sigma_y} \Delta y \quad \text{und} \qquad 6)$$

$$\Delta y = r \frac{\sigma_y}{\sigma_x} \Delta x \quad \text{(nach dem ersten Ansatz),}$$

können, wegen fehlender Reziprozität der Koeffizienten $r \frac{\sigma_x}{\sigma_y}$ und $r \frac{\sigma_y}{\sigma_x}$ diese beiden Gleichungen unmöglich Gleichungen ein- und derselben Geraden sein. Sie entsprechen vielmehr zwei Geraden, die einen um so größeren Winkel miteinander einschließen, je kleiner r wird. Weil für $\Delta x = 0$ auch $\Delta y = 0$ wird, schneiden sich beide im Punkt (m_x, m_y), der übrigens wegen der vorgelegten Form der Gleichungen zum 0-Punkt des Koordinatensystems wird. Für $r = 0$ fallen die Geraden mit den Achsen zusammen, während für den Maximalwert von r (1) nur eine Gerade sich ergibt. Es soll später noch gezeigt werden, wie r funktionell mit dem Winkel zwischen beiden Geraden zusammenhängt. Das Vorzeichen des Abhängigkeitsmaßes richtet sich, wie die Beziehungsgleichungen zeigen, darnach, ob x und y gleichzeitig zunehmen, oder abnehmen, oder aber, ob die Änderung der beiden Größen eine ungleichsinnige ist. Im ersten Falle wird r positiv, im zweiten negativ. Das läßt sich übrigens auch aus dem für das Abhängigkeitsmaß gewonnenen Ausdruck herauslesen. Machen wir den Punkt (m_x, m_y) zum Nullpunkt des Koordinatensystems, dann müssen bei gleichsinniger Änderung von x und y die Punkte P_i (x_i, y_i) vorwiegend im ersten und dritten Quadranten liegen; $\Sigma \Delta x \Delta y$ wird also positiv. Nimmt y bei zunehmendem x ab, dann wird die Punktschar den zweiten und den vierten Quadranten besetzen; wegen ungleichem Vorzeichen von Δx und Δy wird die Summe aller $\Delta x_i \Delta y_i$ negativ. Zur Illustration des Gesagten mögen die folgenden drei Figuren dienen:

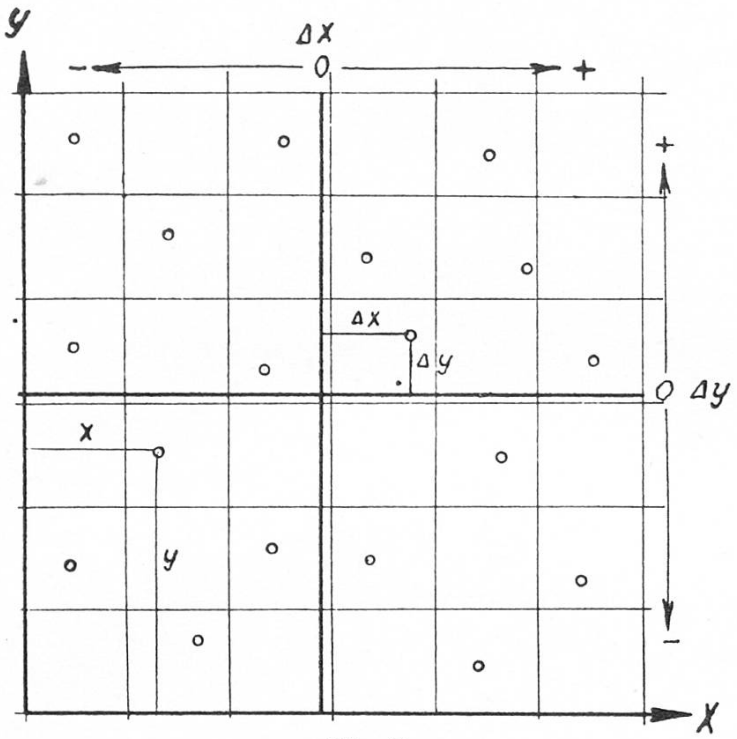

Fig. 3

1. Fall: Unregelmäßige Verteilung der Punkte P_i, deren Koordinaten zusammenhängende Werte von x und y sind. Das Zusammentreffen ist als irgendwie Zufälliges zu bezeichnen. Nach Bildung der Mittel

$$m_x = \frac{\Sigma x_i}{n} \qquad m_y = \frac{\Sigma y_i}{n}$$

wurde durch P (m_x, m_y) ein neues System gelegt; es ist leicht einzusehen, daß das Produkt der auf dieses System bezogenen Koordinaten ($\Delta x_i \cdot \Delta y_i$) in annähernd gleicher Häufigkeit positive und negative Werte ergeben wird. Da auch die absoluten Werte, im gesamten genommen, kaum stark voneinander abweichen werden, erhält man für $\Sigma \Delta x \Delta y$ einen von Null nur wenig abweichenden Wert. r ist also auch klein und bestätigt damit den losen Zusammenhang zwischen x und y.

Man kann von vorneherein auf Beziehungslosigkeit zwischen x und y schließen, wenn

a. homogene Besetzung des Feldes mit Punkten P_i (x_i, y_i) vorliegt;
b. die Punkte längs den Achsen, oder auf Geraden parallel zu diesen sich vorfinden;
c. eine Häufung der Punkte in einem bestimmt zu umschreibenden Bereich eintritt, durch den sich keine wegen spezieller Lagerung

der Pi bevorzugte Achse legen läßt. Weist die Ebene mehr als einen Häufungsbereich auf, dann müssen diese die in a und b für die P_i angegebenen Bedingungen erfüllen.

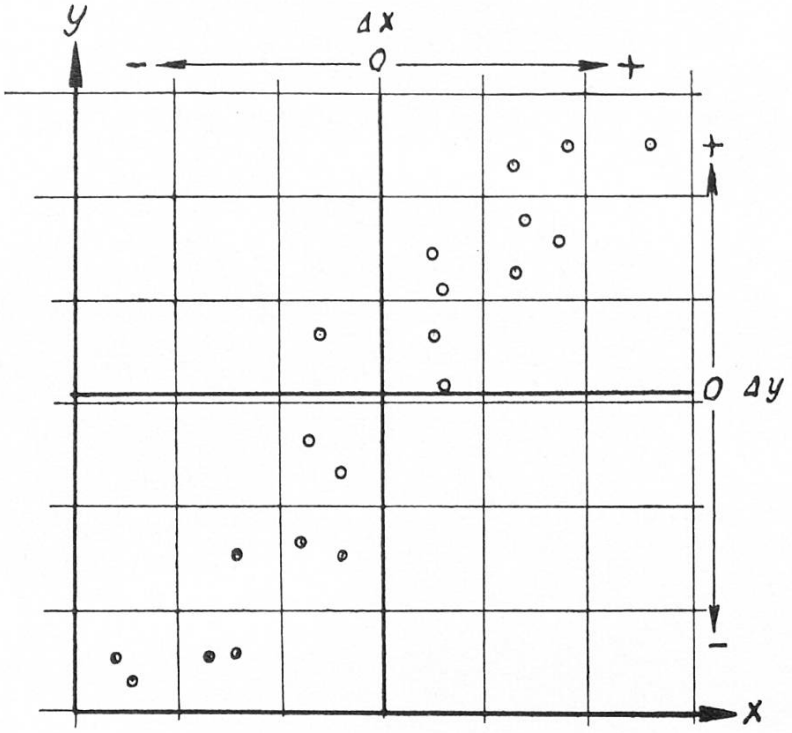

Fig. 4

2. Fall: Zunehmenden x-Werten entsprechen im allgemeinen zunehmende y-Werte. Eine Beziehung zwischen den beiden Größen besteht also zweifellos. Das durch P_m (m_x, m_y) gelegte Koordinatensystem teilt die ganze Fläche in vier Quadranten auf, von denen aber, im Gegensatz zum vorigen Fall, hauptsächlich der erste und der dritte mit Punkten besetzt sind. Die Summe aller $\Delta x \Delta y$ wird positiv und damit auch der Korrelationskoeffizient r.

3. Fall: Die Beziehung zwischen x und y ist ebenfalls deutlich zu erkennen. Der Abnahme von y bei gleichzeitig zunehmendem x entsprechend finden sich die Punkte hauptsächlich im zweiten und vierten Quadranten (vgl. Fig. 5). Für r muß sich also ein negativer Wert ergeben.

Aus der Lage der Punkte in den letzten beiden Fällen läßt sich die voraussichtlich lineare Abhängigkeit aus dem mehr oder weniger gerade gerichteten Streuungs-„band" erkennen. Es besteht also kein Anlaß, eine der im Anfang dieses Abschnittes angedeuteten Methoden zur „Geraderichtung" der Bindung anzuwenden.

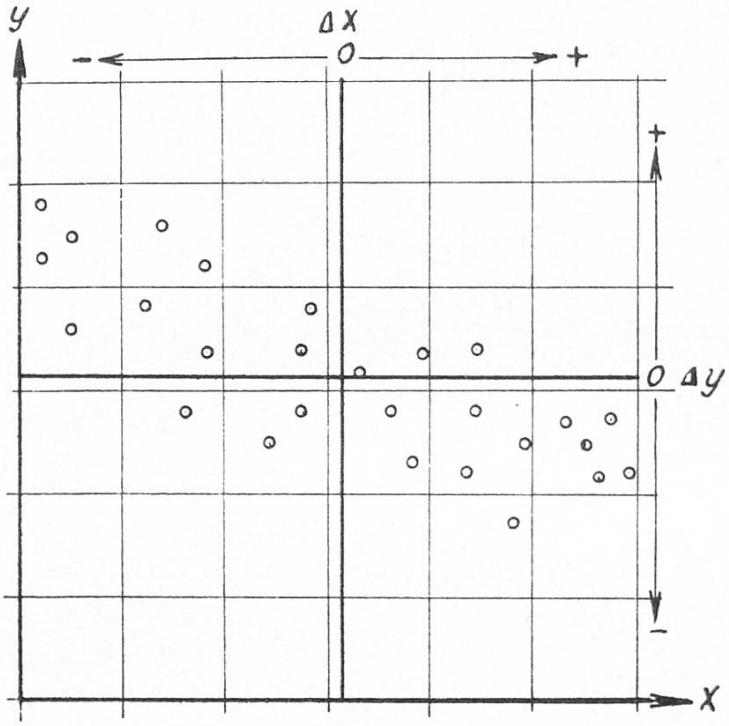

Fig. 5

Die Beziehungsgleichung

Hat man auf die Frage nach der Möglichkeit eines Zusammenhangs zwischen zwei Größen die Antwort erhalten, daß — beispielsweise — für den Korrelationskoeffizienten der Wert von — 0,935 sich ergeben habe und daß auf Grund dieses Resultates

1. die Existenz eines Zusammenhangs bejaht werden müsse,
2. die Bindung eine sehr enge und an ihrem linearen Charakter nicht zu zweifeln sei,
3. die eine Größe zunehme, wenn die andere abnehme,

dann können wir diese Auskunft kaum als erschöpfende bezeichnen. Wir begnügen uns nicht damit, zu wissen, daß mit der Veränderung einer Größe die Veränderung einer andern parallel geht; wir möchten darüber hinaus das Wieviel in Erfahrung bringen. Das führt zur Aufstellung einer Beziehungsgleichung zwischen x und y, der wir die Form geben wollen:

$$y_i = bx_i + a_i \qquad (i = 1 \rightarrow n) \qquad 6)$$

Unter (x_i, y_i) sind dabei alle möglichen Paare von einander zugeordneten Werten von x und y zu verstehen. Summieren wir über alle i von 1 bis n, und teilen durch n, dann erhalten wir

$$m_y = bm_x + m_a \qquad (m_a = \frac{1}{n}\sum_1^n a_i) \qquad 7)$$

Subtrahieren wir 7) von allen Gleichungen 6) und summieren neuerdings über alle i, dann muß wegen

$$\Sigma \Delta y = 0, \quad \Sigma \Delta x = 0, \quad \Sigma \Delta a_i = 0$$

die gesuchte Gerade durch $P_m (m_x, m_y)$ gehen. Ihre Gleichung lautet also

$$y_i - m_y = b \cdot (x_i - m_x) \quad \text{oder} \qquad 8)$$

$$\Delta y_i = b \cdot \Delta x_i$$

Es muß nun festgestellt werden, ob diese Beziehungsgleichung mit der bereits früher aufgestellten (6)) identisch ist.

Für jedes Wertepaar $(\Delta x_i, \Delta y_i)$ wird des nicht funktionellen, sondern korrelativen Charakters der Bindung wegen b von Fall zu Fall einen andern Wert annehmen. Setzen wir dagegen einen bestimmten Koeffizienten b in alle Gleichungen ein, dann werden wir bei jeder mittels eines Summanden v_i eine Korrektur vorzunehmen haben, so daß die Gleichungen dann folgendes Aussehen bekommen:

$$\Delta y_i - b \cdot \Delta x_i = v_i$$

Weil die Wahl eines bestimmten b die Korrektur durch v_i notwendig macht, wird man füglich v_i als den durch b bedingten Fehler der Gleichung bezeichnen können. Der beste Wert für b wird dann nach der Methode der kleinsten Quadrate derjenige sein, für den die Summe aller Fehlerquadrate zu einem Minimum wird. Quadrieren wir alle Gleichungen und summieren über alle i, dann ergibt sich:

$$\Sigma \Delta y^2 - 2b \, \Sigma \Delta x \Delta y + b^2 \, \Sigma \Delta x^2 = \Sigma v^2 = \text{Min.}$$

Wir differenzieren nach b und erhalten

$$-2 \, \Sigma \Delta x \Delta y + 2b \, \Sigma \Delta x^2 = 0$$

$$b = \frac{\Sigma \Delta x \Delta y}{\Sigma \Delta x^2} \qquad 9)$$

Drückt man Δx durch Δy aus:

$$x_i - m_x = b^* \cdot (y_i - m_y) \quad \text{oder}$$

$$\Delta x_i = b^* \cdot \Delta y_i \qquad 10)$$

dann erhält man in ähnlicher Weise

$$b^* = \frac{\Sigma \Delta x \Delta y}{\Sigma \Delta y^2} \qquad 11)$$

Diese neuen Maße b werden als Regressionskoeffizienten bezeichnet. Zur bessern Unterscheidung kennzeichnen wir sie mit besondern Indizes: b_{21} gibt an, um wieviel Δy zu- oder abnimmt, wenn Δx sich um die Einheit ändert; b_{12} drückt entsprechend die Änderung von Δx durch die von Δy aus. Also

$$\begin{aligned} \Delta y &= b_{21} \Delta x \\ \Delta x &= b_{12} \Delta y \end{aligned} \qquad 12)$$

Infolge der korrelativen Beziehung zwischen x und y sind 12) nicht Gleichungen ein und derselben Geraden, denn das Produkt $b_{21} \cdot b_{12}$ wird nicht zu eins, wie dies bei funktioneller Bindung der Fall sein müßte. Wir erhalten vielmehr

$$b_{21} \cdot b_{12} = r^2$$

Damit werden die Ergebnisse, die bei der Ableitung des Abhängigkeitsmaßes erzielt worden sind, bestätigt; die Regressionskoeffizienten stimmen mit den bei 6) angegebenen Koeffizienten von Δx und Δy überein:

$$\begin{aligned} b_{21} &= \frac{\Sigma \Delta x \Delta y}{\Sigma \Delta x^2} = r \cdot \frac{\sigma_y}{\sigma_x} \\ b_{12} &= \frac{\Sigma \Delta x \Delta y}{\Sigma \Delta y^2} = r \cdot \frac{\sigma_x}{\sigma_y} \end{aligned} \qquad 13)$$

Bevor wir an die praktische Durchführung der Korrelationsrechnung herangehen, haben wir noch die Verläßlichkeit der Rechnungsergebnisse zu prüfen. Da diese sich auf ein dem Umfang nach beschränktes Beobachtungsmaterial stützen, ist es ohne weiteres denkbar, daß zusätzliche Beobachtungen ein bereits erzieltes Ergebnis nicht, oder wenigstens nicht in vollem Umfange bestätigen werden. Diesem Umstand muß in der Weise Rechnung getragen werden, daß man zu der quantitativen Aussage einen möglichen Variabilitätsbereich beifügt, dessen Grenzen durch einen zu definierenden mittleren Fehler — auch mittlere Abweichung oder Streuung genannt, wie dies früher schon erwähnt wurde — bestimmt sind. Wenn dieser mittlere Fehler im allgemeinen mit wachsendem n (mit n ist der Umfang des der Rechnung zugrunde gelegten Zahlenmaterials bezeichnet) abnimmt, dann kommt beim Korrelationskoeffizienten noch hinzu, daß er, wie sein m. Fehler über die Sicherheit oder Unsicherheit einer Be-

ziehung Aufschluß zu geben hat. r ist also selber eine Art Fehlermaß und es ist deshalb verständlich, daß man sich bei der Definition des mittleren Fehlers den Korrelationskoeffizienten zur Grundlage genommen hat. Nach Pearson ist

$$f_r = \frac{1 - r^2}{\sqrt{n}} \qquad 14)$$

Eine Ableitung dieses Fehlermaßes kommt hier nicht in Frage; wir haben aber die Möglichkeit, anhand der in 3) ausgedrückten Beziehung f durch das Mittel der Fehlerquadrate auszudrücken. Dieses ist gleich dem Betrag, um den r^2 von 1 abweicht. Da f den Ausdruck $1 - r^2$ ebenfalls im Zähler aufweist, kann dieser als Fehler einer Teilbestimmung von r mittels eines Wertepaares (x_i, y_i) aufgefaßt werden. Wir hätten also:

$$f_i = \frac{1}{n} \Sigma v_i^2 = 1 - r^2$$

Nun ist aber bekanntlich der Fehler eines Mittels von Werten, denen die Fehler f_i zukommen, in folgender Weise zu berechnen:

$$f_m = \frac{1}{n} \sqrt{\Sigma f_i^2}$$

In unserem Falle sind alle Fehler unter sich gleich $(1 - r^2)$, so daß sich ergibt:

$$f_r = \frac{1}{n} \sqrt{n \cdot (1 - r^2)^2}$$

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

Für die Regressionskoeffizienten ergeben sich die Fehler als Proportionalwerte zu f_r.

$$f(b_{21}) = f_r \cdot \frac{\sigma_y}{\sigma_x}$$

$$f(b_{12}) = f_r \cdot \frac{\sigma_x}{\sigma_y} \qquad 15)$$

Die Durchführung der Korrelationsrechnung

Der rechnerischen Auswertung des Materials sollte grundsätzlich immer die graphische Darstellung vorangehen. Man trägt auf der horizontalen Achse eines Koordinatensystems die Werte der nicht bedingten Veränderlichen und auf der senkrechten die der bedingten Veränderlichen ab. Jedes Wertepaar ergibt im System einen Punkt; die Lagerung der Punkteschar hat uns darüber Aufschluß zu geben, ob sich die Durchführung der Rechnung überhaupt lohnt und — wenn dies der Fall sein sollte — um was für eine Art von Beziehung es sich handelt. Zeitigt diese erste Prüfung ein positives Ergebnis und hat man das Material in auswertbare Form gebracht (in welcher Weise das bei nichtlinearen Bindungen geschieht, ist bereits früher dargestellt worden), dann wäre der Arbeitsgang etwa folgender:

1. Ermittlung der arithmetischen Mittel der x_i- und der y_i-Werte.
2. Bestimmung der Abweichungen der x_i- und der y_i-Werte von deren Mittelwert.
3. Bildung der Produkte $\Delta x_i \cdot \Delta y_i$ und Summierung derselben.
4. Quadrierung der Δx_i und Δy_i und Summierung der Quadrate.

Damit sind alle Elemente für die endgültige Berechnung des Korrelations- und der Regressionskoeffizienten ermittelt. Es erweist sich nun aber als zweckmäßiger, statt mit den Abweichungen vom arithmetischen Mittel mit solchen von einem angenäherten Mittel zu rechnen. Man kann sich dann den ersten Teil der Arbeit — das Berechnen des Mittels — ersparen.

Nennt man das angenäherte Mittel m_a, das arithmetische m, die Einzelwerte a_i, deren Abweichungen von m v_{wi}, die Abweichungen von m_a v_i, dann gilt:

$$\begin{array}{l}
a_1 = m + v_{w1} = m_a + v_1 \\
a_2 = m + v_{w2} = m_a + v_2 \\
a_3 = m + v_{w3} = m_a + v_3 \\
\cdot \\
\cdot \\
\hline
\sum\limits_1^n a_i = n \cdot m + \sum\limits_1^n v_{wi} = n \cdot m_a + \sum\limits_1^n v_i
\end{array}$$

Definitionsgemäß wird m so gewählt, daß $\Sigma v_{wi} = 0$ wird. Daraus folgt:

$$m = m_a + \frac{\Sigma v_i}{n} \qquad 16)$$

Das wahre Mittel ist also gleich der Summe aus angenähertem Mittel und dem Mittel aller Abweichungen gegenüber m_a. Weiter ergibt sich:

$$v_{w1} = m_a - m + v_1 = v_1 - \frac{\Sigma v_i}{n}$$

also

$$v_{w1}^2 = v_1^2 - 2v_1 \frac{\Sigma v_i}{n} + \left(\frac{\Sigma v_i}{n}\right)^2$$

$$v_{w2}^2 = v_2^2 - 2v_2 \frac{\Sigma v_i}{n} + \left(\frac{\Sigma v_i}{n}\right)^2$$

$$\vdots$$

$$\Sigma v_{wi}^2 = \Sigma v_i^2 - \frac{1}{n}(\Sigma v_i)^2 \qquad 17)$$

Die Formel gestattet die direkte Berechnung der Summe der Quadrate der Abweichungen aus den Abweichungen gegenüber dem angenäherten Mittel. Schließlich ist es auch möglich, den Ausdruck $\Sigma v_{wi} v'_{wi}$ ohne Kenntnis der Mittel m und m', auf die sich die Abweichungen v_w und v'_w beziehen, zu berechnen. Werden wieder die Abweichungen gegenüber den angenäherten Mitteln m_a und m'_a mit v und v' bezeichnet, dann gilt:

$$v_{w1} \cdot v'_{w1} = \left(v_1 - \frac{\Sigma v_i}{n}\right) \cdot \left(v'_1 - \frac{\Sigma v'_i}{n}\right)$$

$$= v_1 v'_1 - v_1 \frac{\Sigma v'_i}{n} - v'_i \frac{\Sigma v_i}{n} + \frac{\Sigma v_i \Sigma v'_i}{n^2}$$

$$v_{w2} \cdot v'_w = v_2 v'_2 - v_2 \frac{\Sigma v'_i}{n} - v'_2 \frac{\Sigma v_i}{n} + \frac{\Sigma v_i \Sigma v'_i}{n^2}$$

$$\vdots$$

$$\Sigma v_{wi} \cdot v'_{wi} = \Sigma v_i v'_i - \frac{1}{n} \Sigma v_i \Sigma v'_i \qquad 18)$$

Nach diesen Ausführungen soll nun die vollständige numerische Rechnung an einigen Beispielen dargestellt werden.

1. Beispiel: Die korrelativen Eigenschaften der Punkteschar in Fig. 3. Angenommene Mittelwerte: Für x $m_a = 30$; für y $m'_a = 30$.

x	y	Δ x —	Δ x +	Δ y —	Δ y +	Δ x Δy —	Δ x Δy +	Δ x²	Δ y²
5	14	25		16			400	625	256
5	55	25			25	625		625	625
5	35	25			5	125		625	25
13	25	17		5			85	289	25
14	46	16			16	256		256	256
17	7	13		23			299	169	529
23	33	7			3	21		49	9
24	16	6		14			84	36	196
25	55	5			25	125		25	625
33	44		3		14		42	9	196
34	15		4	15		60		16	225
37	36		7		6		42	49	36
44	5		14	25		100		196	625
45	54		15		24		360	225	576
46	25		16	5		90		256	25
49	43		19		13		247	361	169
54	13		24	17		408		576	289
55	34		25		4		100	625	16
n = 18 Summen		139	127	120	135	1810	1659	5012	5003
		— 12		+ 15		—151			

$$m_x = 30 - \frac{12}{18} = 29{,}33\ldots \quad \Sigma \Delta x^2 = 5012 - \frac{144}{18} = 5004$$

$$m_y = 30 + \frac{15}{18} = 30{,}583\ldots \quad \Sigma \Delta y^2 = 5003 - \frac{225}{18} = 4990{,}5$$

$$\Sigma \Delta x \Delta y = -151 + \frac{180}{18} = -141$$

Als Korrelationskoeffizient ergibt sich:

$$r = -\frac{141}{\sqrt{5004 \cdot 4990{,}5}} = -0{,}0282$$

Fehler des Korrelationskoeffizienten:

$$f_r = \frac{1 - 0{,}0282^2}{\sqrt{18}} = \pm\, 0{,}235_7$$

r hat einen erwartungsgemäß von 0 sehr wenig abweichenden Wert; zwischen x und y besteht also kein Zusammenhang. Diese Tatsache wird noch unterstrichen durch f_r: Der Fehler des Korrelationskoeffizienten ist nahezu zehnmal so groß, wie dieser selbst, so daß also für r ein positiver Wert ebenso wahrscheinlich wäre, wie ein negativer. Für die Regressionskoeffizienten erhalten wir:

$$b_{21} = -\frac{141}{5004} = -\,0{,}0282 \pm 0{,}23_6$$

$$b_{12} = -\frac{141}{4990{,}5} = -\,0{,}0282 \pm 0{,}23_6$$

Da $b_{21} = \mathrm{tg}\alpha$ und $b_{12} = \mathrm{tg}\alpha' = \cot(\alpha + \delta)$ zu setzen ist, wobei mit δ der Winkel zwischen beiden Geraden bezeichnet wird, ist es uns möglich, aus den Regressionskoeffizienten δ zu berechnen.

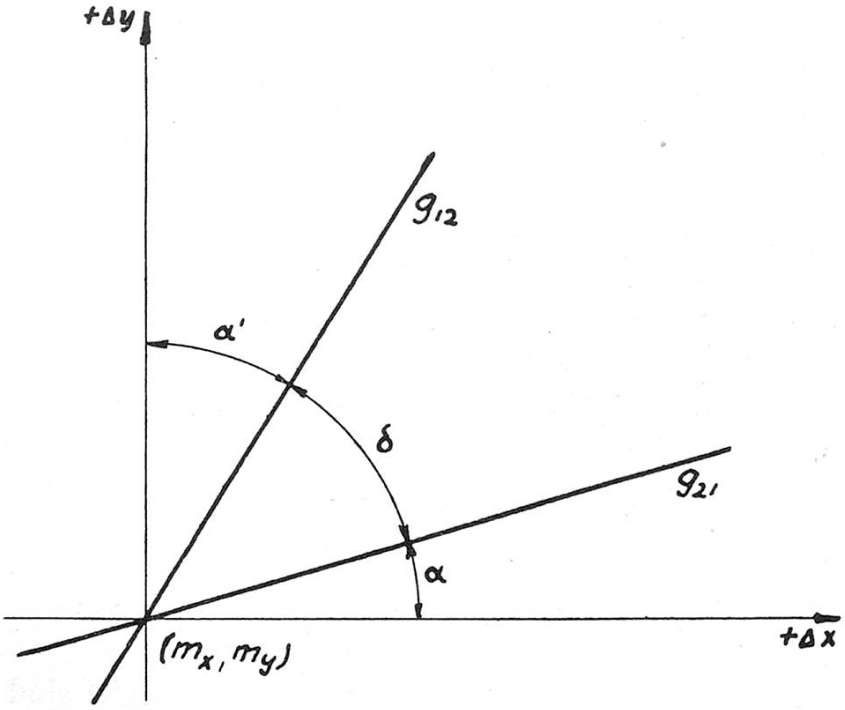

Fig. 6

Es ist

$$\operatorname{tg} \delta = \frac{\operatorname{tg}(\alpha + \delta) - \operatorname{tg} \alpha}{1 + \operatorname{tg}\alpha \operatorname{tg}(\alpha + \delta)}$$

$$= \frac{\frac{1}{b_{12}} - b_{21}}{1 + \frac{b_{21}}{b_{12}}} =$$

$$= \frac{1 - r^2}{b_{12} + b_{21}} \qquad 19)$$

Für unser Beispiel ergibt sich:

$$\operatorname{tg} \delta = \frac{1 - 0{,}0282^2}{-2 \cdot 0{,}0282} = -\frac{0{,}9992}{0{,}0564} = -17{,}71$$

$$\delta \sim -87^0$$

Die beiden Geraden stehen also nahezu senkrecht aufeinander. Weil das Vorzeichen negativ ist, müssen sie durch den II. und IV. Quadranten laufen. Die Abweichung von Abszisse und Ordinate beträgt je etwa 1,5°. Bei einer von englischen und amerikanischen Statistikern bevorzugten Form der Beziehungsgleichung 8), oder 6), die man erhält, wenn man für Δx σ_x und für Δy σ_y als Maßeinheit wählt, werden die Regressionskoeffizienten unter sich gleich und damit gleich r. Dann gilt:

$$\operatorname{tg} \delta_0 = \frac{1 - r^2}{2r} \qquad 20)$$

Es ist dies, wie später zu zeigen sein wird, der maximale Wert, der bei keiner andern Wahl der Maßstäbe mehr erreicht wird. Er soll inskünftig als neuer Parameter mit p bezeichnet werden. Drücken wir umgekehrt r durch eine Funktion des Maximalwinkels aus, dann erhalten wir:

$$r = \frac{1 - \sin \delta_0}{\cos \delta_0} \qquad 21)$$

Der Maximalwinkel ist für die Enge der Bindung ebenso charakteristisch, wie der Korrelationskoeffizient selbst.

Die Beziehungsgeraden zum vorstehenden Beispiel sind in der folgenden Figur dargestellt.

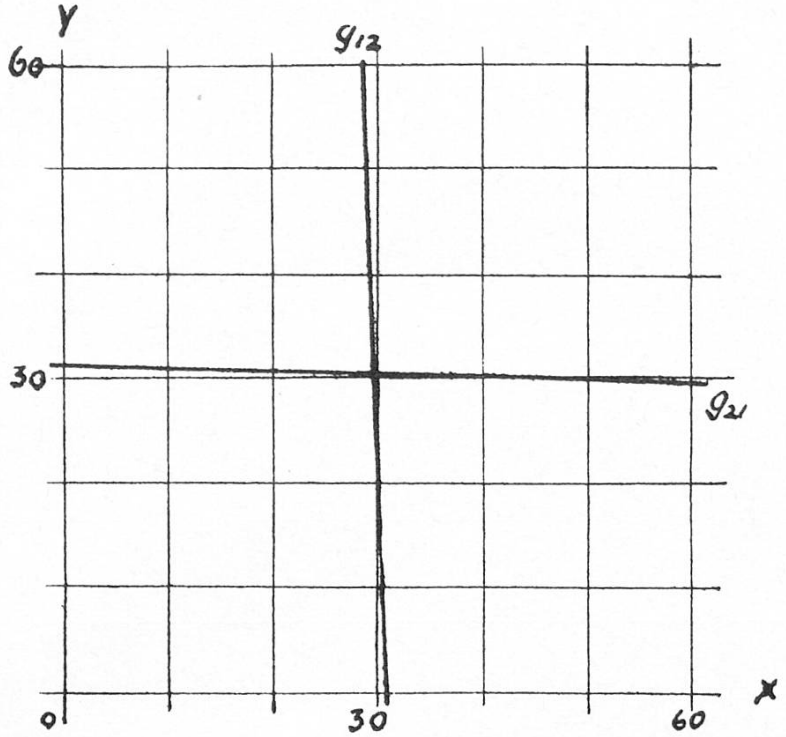

Fig. 7

2. Beispiel. Punktsystem der Fig. 4. Als angenäherte Mittelwerte für x und y wurden wieder je 30 gewählt ($m_x = 30$; $m_y = 30$). Die Ergebnisse sind hier vorausgenommen.

$$m_x = 30 - \frac{1}{20} = 29{,}95 \qquad \Sigma \Delta x^2 = 4059 - \frac{1}{20} = 4058{,}95$$

$$m_y = 30 + \frac{15}{20} = 30{,}75 \qquad \Sigma \Delta y^2 = 5939 - \frac{225}{20} = 5927{,}75$$

$$\Sigma \Delta x \Delta y = +4605 + \frac{15}{20} = +4605{,}75$$

$$r = +\frac{4606}{\sqrt{4059 \cdot 5928}} = +0{,}940 \quad f_r = \frac{1 - 0{,}94^2}{\sqrt{20}} = \pm 0{,}026$$

$$b_{21} = +\frac{4606}{4059} = +1{,}134 \pm 0{,}037$$

$$b_{12} = + \frac{4606}{5928} = + 0{,}777 \pm 0{,}022$$

$$p' = \frac{1 - 0{,}940^2}{1{,}134 + 0{,}777} = \frac{0{,}1165}{1{,}911} = 0{,}061$$

$$\delta' = 3^0\, 30'$$

$$p = \frac{0{,}1165}{1{,}880} = 0{,}062 \quad \delta_0 = 3^0\, 34'$$

n	x	y	Δx —	Δx +	Δy —	Δy +	ΔxΔy —	ΔxΔy +	Δx^2	Δy^2
1	4	7	26		23			598	676	529
2	6	4	24		26			624	576	676
3	13	6	17		24			408	289	576
4	16	6	14		24			336	196	576
5	16	16	14		14			196	196	196
6	22	17	8		13			104	64	169
7	26	16	4		14			56	16	196
8	26	24	4		6			24	16	36
9	23	27	7		3			21	49	9
10	24	37	6			7	42		36	49
11	35	37		5		7		35	25	49
12	36	32		6		2		12	36	4
13	36	41		6		11		66	36	121
14	35	45		5		15		75	25	225
15	43	43		13		13		169	169	169
16	47	46		17		16		272	289	256
17	43	53		13		23		299	169	529
18	44	48		14		18		252	196	324
19	48	55		18		25		450	324	625
20	56	55		26		25		650	676	625
Summen			124	123	147	162	42	4647	4059	5939
			— 1		+ 15		+ 4605			

Der Korrelationskoeffizient hat einen hohen, positiven Wert; er bestätigt den engen Zusammenhang, der sich schon aus der graphischen Darstellung ergibt und die gleichsinnige Änderung von x und y.

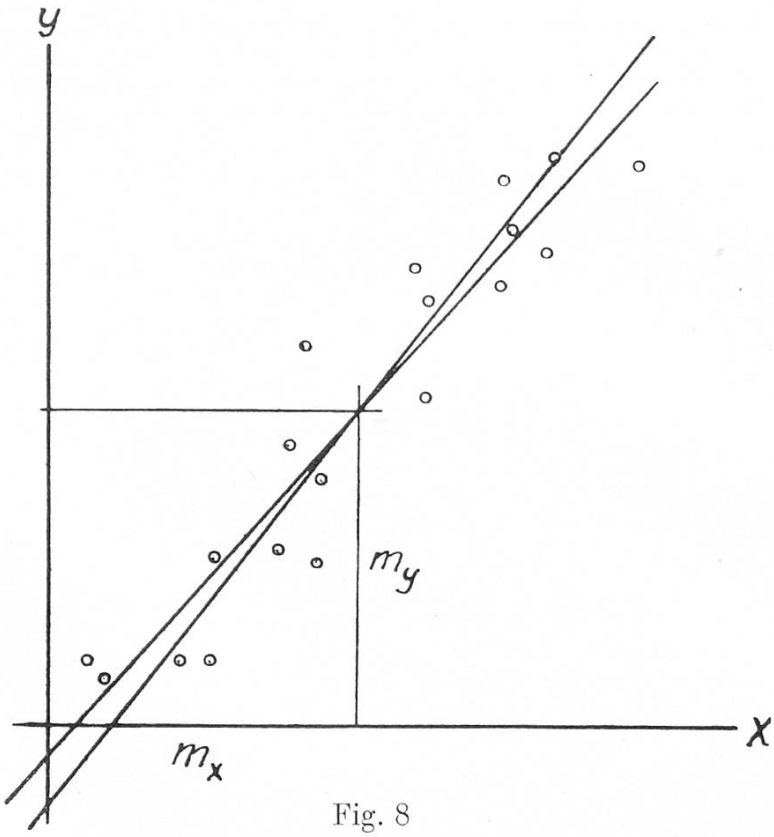

Fig. 8

Die Geraden g_{21} und g_{12}, deren Gleichungen

$$\Delta y = +1{,}134\ \Delta x$$

und

$$\Delta x = +0{,}777\ \Delta y$$

heißen, erhält man sehr einfach: Man berechnet aus den Regressionskoeffizienten die Winkel, die die Geraden mit den Achsen einschließen, und legt die g dann in der bestimmten Richtung durch P (m_x, m_y). Die Richtung ergibt sich auch durch Konstruktion des rechtwinkligen Dreiecks mit dem dem Regressionskoeffizienten entsprechenden Kathetenverhältnis. Im vorliegenden Beispiel ist g_{21} um 48° 40' gegen die Horizontale und g_{12} 37° 50' gegen die Vertikale geneigt. Der Winkel zwischen beiden Geraden ergibt sich, wie bei der direkten Berechnung (vergleiche p') zu 3° 30'; er kann, wenn x und y in „Normalmaß" (σ_x und σ_y) ausgedrückt werden, den maximalen Betrag von 3° 34' erreichen.

3. Beispiel. System der Punkte in Fig. 5. Angenäherte Mittelwerte: $m_x = 30$, $m_y = 30$.

n	x	y	Δx —	Δx +	Δy —	Δy +	ΔxΔy —	ΔxΔy +	Δx^2	Δy^2
1	2	43	28			13	364		784	169
2	2	48	28			18	504		784	324
3	5	36	25			6	150		625	36
4	5	45	25			15	375		625	225
5	11	38	19			8	152		361	64
6	13	46	17			16	272		289	256
7	16	28	14		2			28	196	4
8	18	34	12			4	48		144	16
9	18	42	12			12	144		144	144
10	24	25	6		5			30	36	25
11	27	28	3		2			6	9	4
12	27	34	3			4	12		9	16
13	28	38	2			8	16		4	64
14	33	32		3		2		6	9	4
15	36	28		6	2		12		36	4
16	38	23		8	7		56		64	49
17	39	34		9		4		36	81	16
18	43	22		13	8		104		169	64
19	43	34		13		4		42	169	16
20	44	28		14	2		28		196	4
21	49	25		19	5		95		361	25
22	48	17		18	13		234		324	169
23	53	27		23	3		69		529	9
24	55	25		25	5		125		625	25
25	56	22		26	8		208		676	64
26	57	27		27	3		81		729	9
27	59	22		29	8		232		841	64
Summen			194	233	73	114	3281	148	8819	1869
			+ 39		+ 41		−3133			

$$m_x = 30 + \frac{39}{27} = 31{,}445 \qquad \Sigma \Delta x^2 = 8819 - \frac{1521}{27}$$

$$m_y = 30 + \frac{41}{27} = 31{,}518 \qquad = 8762{,}66..$$

$$\Sigma \Delta x \Delta y = -3133 - \frac{1599}{27} \qquad \Sigma \Delta y^2 = 1869 - \frac{1681}{27}$$

$$= -3192{,}2 \qquad = 1806{,}74$$

$$r = -\frac{3192}{\sqrt{8763 \cdot 1807}} = -0{,}804$$

$$f_r = \frac{1 - 0{,}804^2}{\sqrt{27}} = \pm 0{,}0681$$

$$b_{21} = -\frac{3192}{8763} = -0{,}364$$

$$b_{12} = -\frac{3192}{1807} = -1{,}768$$

$$p' = \frac{1 - 0{,}804^2}{-0{,}364 - 1{,}768} = -\frac{0{,}354}{2{,}132} = -0{,}166$$

$$\delta' = -9^0\,27'$$

$$p = -\frac{0{,}354}{2 \cdot 0{,}804} = -\frac{0{,}354}{1{,}608} = -0{,}220$$

$\delta_0 = -12^0\,24'$ (das Minuszeichen deutet an, daß die Geraden durch den 2. und 4. Quadranten laufen)

Der ungleichsinnigen Änderung von x und y entsprechend ist das Vorzeichen des Korrelationskoeffizienten und damit auch der b- und p-Werte negativ. r ist kleiner als im vorigen Beispiel; ein Vergleich der beiden Punktbilder in den Fig. 4 und 5 läßt leicht erkennen, daß beim zweiten der Streuungsbereich größer, der Zusammenhang zwischen x und y also ein lockerer ist als beim ersten.

Fig. 9 zeigt, daß die Regressionsgeraden sich gut in das Punktsystem einfügen. Der Winkel zwischen beiden beträgt $9^0\,27'$ (max. $12^0\,24'$), ist also wesentlich größer als im 2. Beispiel. Man erkennt hieraus, daß p (δ) ein empfindlicheres Maß der Abhängigkeit darstellt als r. Nehmen wir die Werte des 2. Beispiels zur Grundlage, dann haben sich die entsprechenden des dritten in folgender Weise geändert: r ist, absolut genommen, um 0,136 kleiner; p dagegen um 0,158 größer geworden!

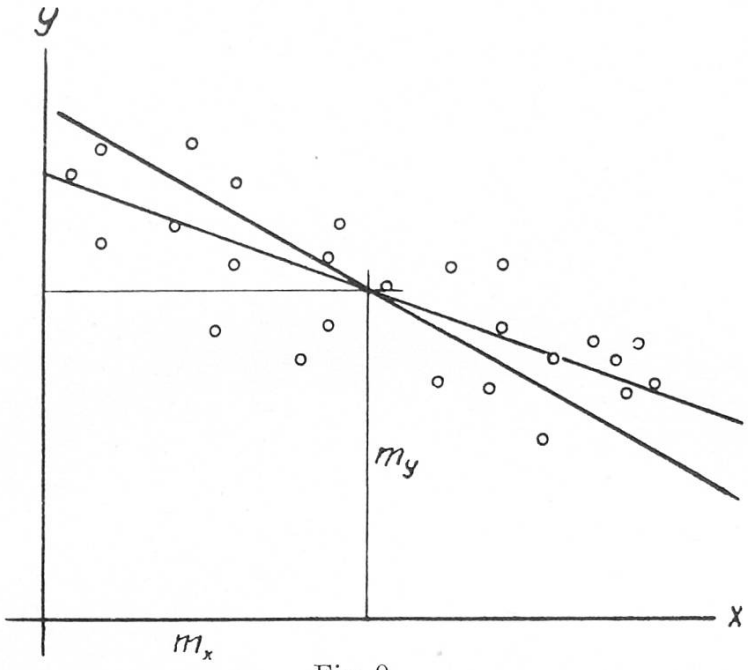

Fig. 9

Fig. 10 stellt p als Funktion von r dar. Für r = 1 ist p = 0; in diesem Punkt ist die Steigung von p absolut genommen = 1, also gleich der von r selbst. Sie nimmt nun aber mit sinkenden Werten von r rasch zu: Für r = 0,9 beträgt sie 1,118; für r = 0,8 bereits 1,282, usw. p selbst erreicht für $r = -1 + \sqrt{2}$ den Wert 1. δ_0 ist dann 45°, also gleich weit von 0° und 90° entfernt. Man kann daher mit gutem Grund sagen, daß der Wert 0,414 für den Korrelationskoeffizienten ungefähr die Mitte hält zwischen völliger Beziehungslosigkeit und streng funktionellem (linearem) Zusammenhang. Qualitativ wären also Zusammenhänge, für die r kleiner als 0,414 ist, als zweifelhafte zu bezeichnen. Eine völlige Beziehungslosigkeit liegt nicht nur dann vor, wenn r = 0 ist, sondern auch, wenn der Fehler von r ebenso groß ist, wie r selbst. Als untere Grenze der Korrelation wäre dann nach 14) festzulegen:

$$r_{mi} = -\frac{\sqrt{n}}{2} + \sqrt{1 + \frac{n}{4}} \qquad 22)$$

Sie ergibt sich als eine Funktion von n, also der Zahl der Wertepaare (x_i, y_i), die der Korrelationsrechnung zugrunde gelegt werden. Mit wachsendem n nähert sich r_{mi} der Grenze 0. Die Sicherheit einer Aussage nimmt mit wachsendem n zu. Die folgende Tabelle enthält die r_{mi} und die zugehörigen n. Man sieht, daß zum Beispiel bei 14 Wertepaaren der Korrelationskoeffizient nicht kleiner als 0,25 sein darf, wenn überhaupt eine Beziehung noch möglich sein soll.

r_{mi}	n
0	∞
0,05	398
0,10	98
0,15	42
0,20	23
0,25	14
0,30	9,2
0,35	6,3
0,40	4,4
0,45	3,2
0,50	2,3
1,00	0

Es hat natürlich nur formale Bedeutung, die Werte für n über r_{mi} = 0,30 hinaus anzugeben. Durch 2 Punkte läßt sich immer eine Gerade, durch 3 immer ein Kegelschnitt legen, so daß man also bei 2 oder 3 Wertepaaren in jedem Fall auf Grund des Rechnungsergebnisses auf einen rein funktionellen Zusammenhang zu schließen hätte. Um Fehlschlüssen zu entgehen, wird man also der Korrelationsrechnung eine möglichst große Zahl von Wertepaaren zugrunde legen — unter Berücksichtigung gewisser Vorbehalte, von denen später noch die Rede sein wird.

Wir wollen an dieser Stelle uns noch jenen Fällen nichtlinearer Beziehungen zuwenden, die nicht durch Logarithmierung einer rechnerischen Auswertung zugänglich zu machen sind. Es handelt sich, wie bereits pag. 18/19 erwähnt wurde, um Bindungen, die sich am besten durch rationale Funktionen n-ten Grades darstellen lassen. Hat die Bildung von Differenzreihen, wie sie auch bei der Interpolationsrechnung gebräuchlich sind, eine k-te Reihe ergeben, die mit dem Argument x sich linear ändert, dann setzen wir diese Reihe mit der der entsprechenden x-Werte in Beziehung. Durch Integration läßt sich die ursprüngliche „Funktion" wieder gewinnen. Für die Ermittlung der Konstanten stehen genügend in diesen lineare Gleichungen zur Verfügung.

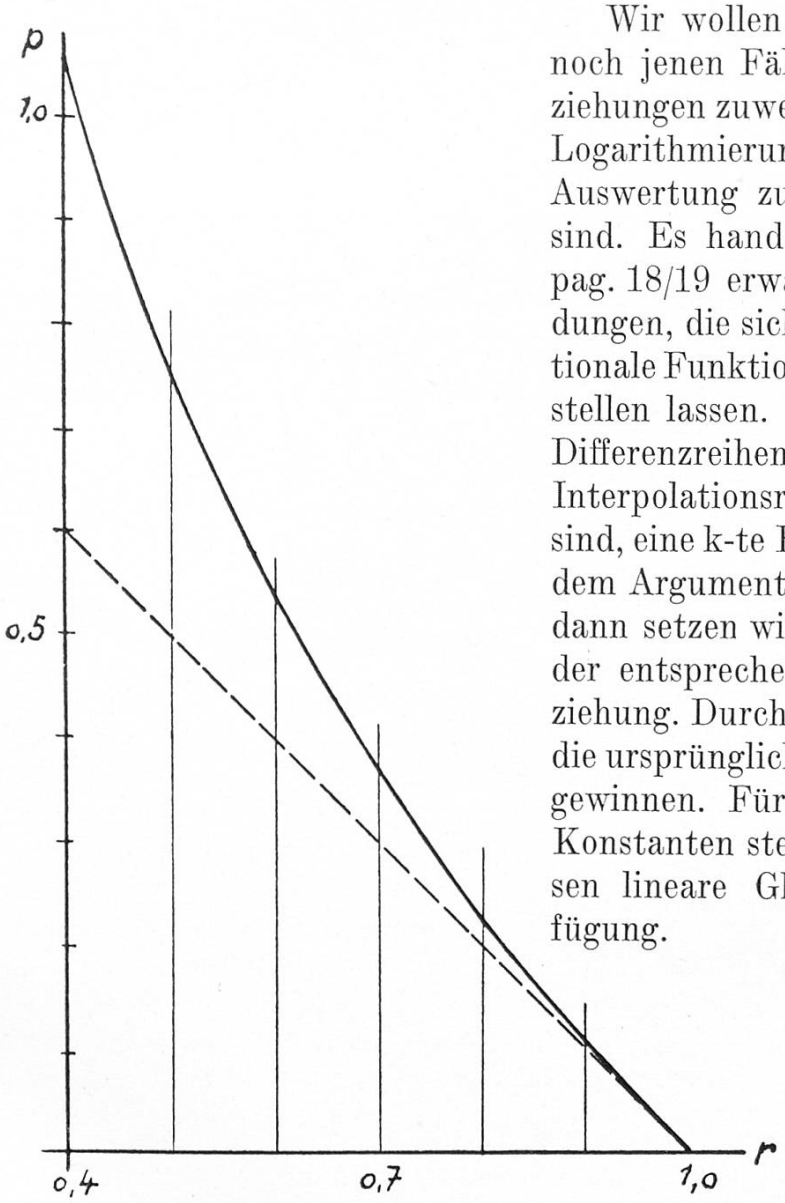

Fig. 10

4. Beispiel: Auswertung des aus dem 2. Differenzschema pag. 19 gewonnenen Zahlenmaterials.

$$m_x' = 3{,}5, \quad m_y' = 25$$

Der Einfachheit halber wurde für y' y gesetzt.

n	x	y	Δx —	Δx +	Δy —	Δy +	ΔxΔy —	ΔxΔy +	Δx^2	Δy^2
1	0,5	7,4	3		17,6			52,8	9	309,8
2	1,5	12,4	2		12,6			25,2	4	158,8
3	2,5	19,6	1		5,4			5,4	1	29,2
4	3,5	24,5	0		0,5			0	0	0,3
5	4,5	32,1		1		7,1		7,1	1	50,4
6	5,5	35,7		2		10,7		21,4	4	114,5
7	6,5	44,3		3		19,3		57,9	9	372,5
Summen			6	6	36,1	37,1		169,8	28	1035,5
			0			1,0				

$m_x = 3{,}5$ $\qquad \Sigma \Delta x^2 = 28$

$m_y = 25 - 0{,}143 = 24{,}857$ $\qquad \Sigma \Delta y^2 = 1035{,}5 - 0{,}143 = 1035{,}4$

$\Sigma \Delta x \Delta y = 169{,}8$

$$r = + \frac{169{,}8}{\sqrt{28 \cdot 1035{,}4}} = + 0{,}997$$

$$f_r = \frac{1 - 0{,}997^2}{\sqrt{7}} = \pm 0{,}002$$

$$b_{21} = \frac{169{,}7}{28} = + 6{,}06 \qquad b_{12} = \frac{169{,}8}{1035{,}4} = + 0{,}164$$

Mit Hilfe der Regressionskoeffizienten lassen sich nun folgende Gleichungen aufstellen:

$$\Delta y' = 6{,}06 \, \Delta x \quad \text{und}$$
$$\Delta x = 0{,}164 \, \Delta y'$$

Durch Umformung erhalten wir aus der ersten:

$$y' - m_y' = 6{,}06 \, (x - m_x)$$
$$y' - 24{,}857 = 6{,}06 \, (x - 3{,}5)$$
$$y' = 6{,}06x + 3{,}65$$

und aus der zweiten:

$$x - 3{,}5 = 0{,}164\ (y' - 24{,}857)$$

$$x = 0{,}164y' - 0{,}5766$$

$$y' = 6{,}10x + 3{,}51$$

Wegen des annähernd funktionellen Zusammenhangs dürfen wir, ohne dabei große Fehler zu begehen, die Differenzen der Koeffizienten ausgleichen. Wir erhalten dann an Stelle zweier Funktionen zwischen x und y' nur eine einzige:

$$y' = 6{,}08x + 3{,}58$$

Durch Integration ergibt sich:

$$y = c + 3{,}58x + 3{,}04x^2$$

Es bleibt nun noch übrig, die Konstante c zu bestimmen. Gehen wir auf die ursprünglich vorgelegten Wertepaare (x_i, y_i) zurück, dann können wir aus jedem einen Wert für c gewinnen; das Mittel aller dieser c_i setzen wir dann in die oben erhaltene Gleichung ein.

Ist zum Beispiel x = o, dann wird y = c = (vergleiche pag. 19) 1,8. In ähnlicher Weise finden wir die übrigen c_i. Sie sind nachstehend für alle Werte von x angegeben.

x	y	c
0	1,80	1,80
1	9,2	2,58
2	21,6	2,28
3	41,2	3,10
4	65,7	2,74
5	97,8	3,90
6	133,5	2,78
7	177,8	3,78
		22,96
	Mittel	2,87

Endgültige Gleichung, die dem Zusammenhang am besten gerecht zu werden vermag:

$$y = 2{,}87 + 3{,}58x + 3{,}04x^2$$

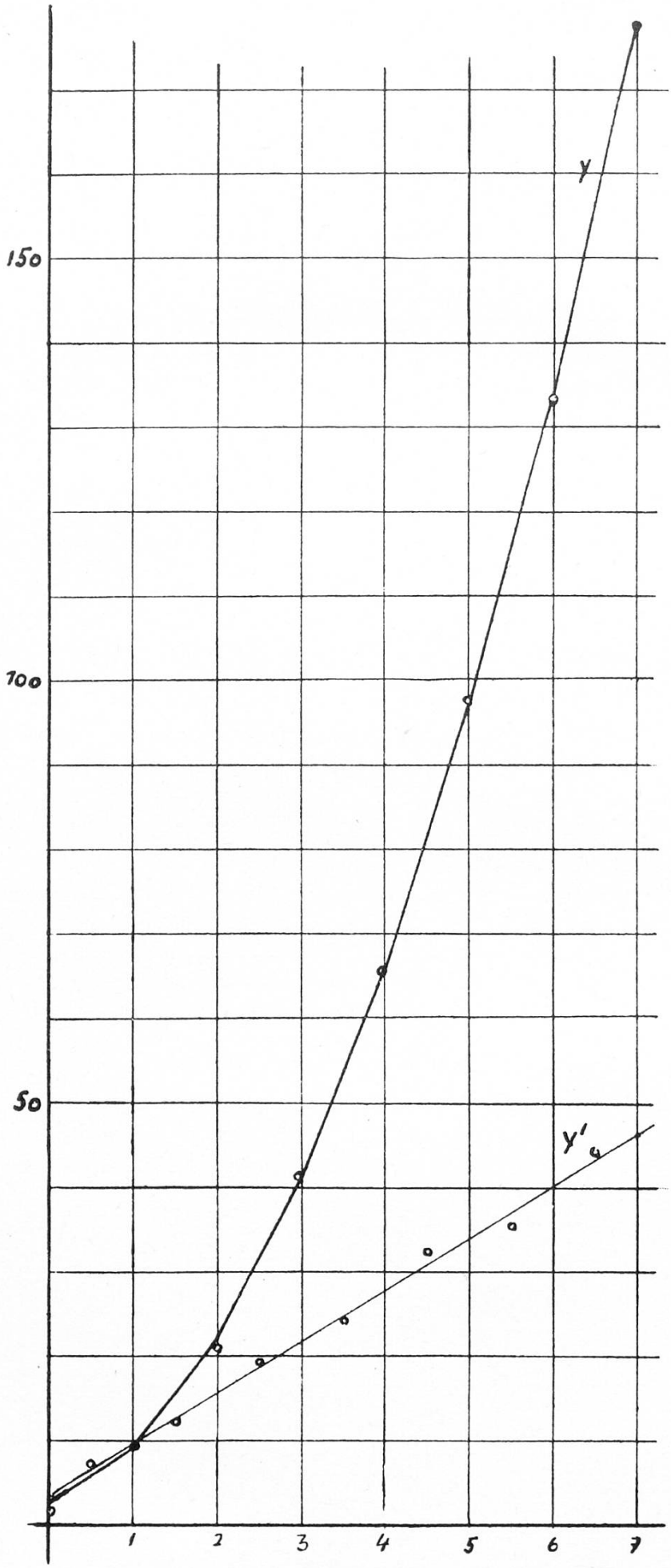

Fig. 11

Berechnete Funktionswerte.

x	y
0	2,87
1	9,49
2	22,19
3	40,97
4	65,83
5	96,77
6	133,59
7	176,89

Der Vergleich mit den „empirisch" ermittelten Werten zeigt, daß die Differenzen geringfügig sind, was beweist, daß die hier dargestellte Methode brauchbar ist. Die nebenstehende Fig. 11 enthält nicht nur die, der oben angegebenen Funktion entsprechende parabolische Kurve, sondern auch die der ersten Ableitung (ersten Differenzreihe) entsprechende Gerade. Die Übereinstimmung der berechneten mit den gegebenen Werten darf auch hier als sehr gute bezeichnet werden. Erhält man für den Korrelationskoeffizienten einen kleinen Wert, dann müssen die beiden Beziehungsgleichungen gesondert ausgewertet werden, das heißt aus jeder ist eine Integral-„Funktion" zu ermitteln. Der Gang der Rechnung bleibt im übrigen derselbe.

5. Beispiel: Das vorgelegte Zahlenmaterial sei folgendes:

x:	0	1	2	3	4	5	6
y:	4,1	— 2,2	— 5,9	— 1,9	16,3	53,8	118,1
x:	7	8	9	10			
y:	213,7	347,8	526,2	753,8			

Es ist zu untersuchen, ob ein Zusammenhang besteht und welcher Art dieser ist.

Schon eine vorläufige Prüfung ergibt, daß eine lineare Beziehung nicht in Frage kommen kann. Während x in gleichen Intervallen von 0 bis 10 fortschreitet, ist die Änderung von y weder in der Richtung, noch im Betrag konstant. Eine logarithmische Behandlung ist der negativen Werte von y wegen nicht möglich. Vielleicht führt aber die Bildung von Differenzreihen zum Ziele.

x	y	y'	y''	y'''
0	4,1			
		— 6,3		
1	— 2,2		2,6	
		— 3,7		5,1
2	— 5,9		7,7	
		4,0		6,5
3	— 1,9		14,2	
		18,2		5,1
4	16,3		19,3	
		37,5		7,5
5	53,8		26,8	
		64,3		4,5
6	118,1		31,3	
		95,6		7,2
7	213,7		38,5	
		134,1		5,8
8	347,8		44,3	
		178,4		4,9
9	526,2		49,2	
		227,6		
10	753,8			

Aus dem vorstehenden Schema ergibt sich eine zu der von x parallele, mehr oder weniger lineare Zunahme der Werte der zweiten Differenzreihe (y''); der Zusammenhang zwischen x und y muß sich also am besten durch eine Gleichung dritten Grades ausdrücken lassen. Wir stellen mit Hilfe der Korrelationsrechnung die Bindung zwischen x und y'' fest und die Regressionsgleichung, die an Stelle der unbestimmt variablen Werte von y''' einen konstanten setzt, kann in zwei Schritten auf die ursprüngliche „Funktion" zurückgeführt werden. — Man beachte, daß die Kolonnenüberschriften im Differenzenschema (y', y'', usw.) nicht mit Ableitungen zu verwechseln sind; nur wo eine k-te Differenzenreihe einen (annähernd) konstanten Wert ergibt, ist sie numerisch der k-ten Ableitung gleich zu setzen.

Der Einfachheit halber setzen wir nachstehend für y'' wieder y. Angenommene Mittelwerte: $m_x' = 5$; $m_y' = 25$.

n	x	y	Δx —	Δx +	Δy —	Δy +	ΔxΔy —	ΔxΔy +	Δx^2	Δy^2
1	1	2,6	4		22,4			89,6	16	501,76
2	2	7,7	3		17,3			51,9	9	299,29
3	3	14,2	2		10,8			21,6	4	116,64
4	4	19,3	1		5,7			5,7	1	32,49
5	5	26,8	0			1,8		0	0	3,24
6	6	31,3		1		6,3		6,3	1	39,69
7	7	38,5		2		13,5		27,0	4	182,25
8	8	44,3		3		19,3		57,9	9	372,49
9	9	49,2		4		24,2		96,8	16	585,64
Summen			10	10	56,2	65,1	0	356,8	60	2133,49
			0			8,9	+ 356,8			

$$m_x = 5 \qquad \Sigma \Delta x^2 = 60$$

$$m_y = 25 + \frac{8,9}{9} = 25,988.. \qquad \Sigma \Delta y^2 = 2133,49 - \frac{79,21}{9}$$

$$= 2124,69$$

$$r = + \frac{356,8}{\sqrt{60 \cdot 2124,7}} \qquad \Sigma \Delta x \Delta y = + 356,8$$

$$= + 0,9993 \pm 0,0007$$

$$b_{21} = + \frac{356,8}{60} = + 5,9466.. \qquad b_{12} = + \frac{356,8}{2124,7} = + 0,16798$$

Regressionsgleichungen:

a) $\Delta y'' = 5,9466..\Delta x$

$y'' = 25,988.. = 5,9466..(x - 5)$

$y'' = - 3,744.. + 5,9466..x$

b) $x = 5 = 0,16798\ (y'' - 25,988..)$

$x = 0,62642 + 0,16798\ y''$

$y'' = - 3,7292 + 5,9530\ x$

Wir erhalten weiter aus a):

$$y' = c_1 - 3{,}744..x + 2{,}9733..x^2$$

$$y = c_0 + c_1 x - 1{,}8722..x^2 + 0{,}9911..x^3$$

und aus b):

$$y' = c_1 - 3{,}7292\,x + 2{,}9765\,x^2$$

$$y = c_0 + c_1 x - 1{,}8646\,x^2 + 0{,}9922\,x^3$$

Es bleibt nun noch übrig, die beiden Konstanten c_0 und c_1 zu bestimmen. Das kann in der Weise geschehen, daß man in den Gleichungen für x der Reihe nach die Werte 0 bis 10 einsetzt, die, weil die zugehörigen Werte von y auch bekannt sind, als einzige Unbekannte noch c_0 und c_1 enthalten. Verlangt man, daß die Konstanten allen Werten von y in gleicher Weise gerecht werden müssen, dann hat man alle 11 Bestimmungsgleichungen zu je zweien zu kombinieren, was im ganzen auf je 55 Bestimmungen von c_0 und c_1 führt. An Stelle dieser, eine ganz erhebliche Rechenarbeit erfordernden Methode soll hier eine andere zur Anwendung kommen, die sich wieder des Differenzschemas bedient, von dem ausgegangen wurde.

Man setzt für y''' den aus der Regressionsgleichung sich ergebenden konstanten Wert ein (bei a) also b_{21}, bei b) den reziproken Wert von von b_{12}) und baut von hier aus das ganze Schema wieder auf, wobei sukzessive die Kolonnenwerte, bei denen der Einfluß vorderhand noch unbekannter Koeffizienten zur Geltung kommt, den gegebenen angeglichen werden. Eine solche Korrektur ist im vorliegenden Fall bei der ersten Differenzreihe (y') notwendig. Setzen wir einen beliebigen ihrer Werte gleich a, dann ist damit auch die ganze Reihe bestimmt und jede Korrektur an a, die durch Beifügung eines additiven Gliedes vorgenommen wird:

$$a' = a + k,$$

bedingt die gleiche Korrektur an allen Gliedern der Reihe. Daraus ergibt sich ohne weiteres das anzuwendende Verfahren: Man stellt eine willkürlich gebildete (das „willkürlich" bezieht sich auf die Wahl eines Ausgangswertes) der Reihe der gegebenen Werte von y' gegenüber und bestimmt den Verbesserungszuschlag k so, daß die Summe der Abweichungen von der korrigierten Reihe gleich 0 wird. Aus Zweckmäßigkeitsgründen wird man den Ausgangswert so wählen, daß er mit einem der gegebenen Werte übereinstimmt.

a) $y'' = -3{,}744.. + 5{,}9466..x$

x	y''	(y')	y'	Abweichungen —	Abweichungen +
0	—3,7444..	—6,3	—6,3		
1	2,2022..	—4,0977..	—3,7		0,3977..
2	8,1488..	4,0511..	4,0	0,0511..	
3	14,0955..	18,1466..	18,2		0,0533..
4	20,0422..	38,1888..	37,5	0,6888..	
5	25,9888..	64,1777..	64,3		0,1222..
6	31,9355..	96,1133..	95,6	0,5133..	
7	37,8822..	133,9955..	134,1		0,1044..
8	43,8288..	177,8244..	178,4		0,5755..
9	49,7755..	227,6000	227,6		
			Summe	—1,2533..	+1,2533..
				0	

x	y'	(y)	y	Abweichungen —	Abweichungen +
0	—6,3	4,1	4,1		
1	—4,0977..	—2,2	—2,2		
2	4,0511..	—6,2977..	—5,9		0,3977..
3	18,1466..	—2,2466..	—1,9		0,3466..
4	38,1888..	15,9000	16,3		0,4000
5	64,1777..	54,0888..	53,8	0,2888..	
6	96,1133..	118,2666..	118,1	0,1666..	
7	133,9955..	214,3800	213,7	0,6800	
8	177,8244..	348,3755..	347,8	0,5755..	
9	227,6000	526,2000	526,2		
10		753,8	753,8		
		Summe der Abweichungen:		—1,7111..	+1,1444..
				= —0,5666..	

$k = -0{,}5666.. : 11 = -0{,}0515..$

Anfangswert der berichtigten Reihe für y:

$$4{,}1 - 0{,}05151 = 4{,}0484..$$

Wie leicht einzusehen ist, entspricht der Rechenvorgang durchaus dem, der bei der Bestimmung eines arithmetischen Mittels aus einem angenäherten gebräuchlich ist. Als Anfangswert der ersten Differenzreihe wurde — 6,3 gewählt. Die Summe der Abweichungen der (y') von den entsprechenden Werten der y' des ersten Schemas wird 0, so daß die vorgegebene Reihe keiner Korrektur mehr bedarf. In gleicher Weise wurde bei der Berechnung von k zur Angleichung der Reihe der y-Werte an die gegebenen verfahren. Die berichtigten Funktionswerte sind nachstehend angeführt.

x	y
0	4,0484..
1	— 2,2515
2	— 6,3493
3	— 2,2981
4	15,8484
5	54,0373
6	118,2151
7	214,3284
8	348,3240
9	526,1484
10	753,7484

Die Bestimmung der Konstanten c_0 und c_1 ist nun einfach. Alle Funktionswerte ergeben dieselben Größen, so daß man sich auf deren Ermittlung an den Stellen 0 und 1 beschränken kann. Man findet:

$$c_0 = 4{,}0484..$$

$$c_1 = -5{,}4188..$$

Die aus der Regressionsgleichung a) sich ergebende beste Beziehungsgleichung lautet also:

$$y = 4{,}0484... - 5{,}4188..x - 1{,}8722..x^2 + 0{,}9911..x^3$$

In ähnlicher Weise würde man aus b) finden:

$$y = 4{,}29 - 5{,}55x - 1{,}865x^2 + 0{,}992x^3$$

Von der guten Übereinstimmung zwischen gegebenen und berechneten Funktionswerten bekommt man einen Begriff, wenn man die Abweichungen als Fehler der bedingt Veränderlichen auffaßt. Bezeichnet man mit μ_x den Fehler von x, mit μ_y den Fehler von y, dann gilt

$$\mu_y = y' \cdot \mu_x$$

An der Stelle 10 ist zum Beispiel $\mu_y = 0{,}0515$, $y' = 254{,}47$, $\mu_x = 0{,}0002$! (Gleichung nach a), also praktisch gleich null.

Der Vollständigkeit halber mag hier noch die Koeffizientenbestimmung mit Hilfe des Gaußschen Eliminationsverfahrens erwähnt werden. F. Baur (Korrelationsrechnung: Math.-Phys. Bibl., Bd. 75) ersetzt x und y durch die entsprechenden Abweichungen von m_x und m_y und erhält so Bestimmungsgleichungen von folgender Form:

$$\Delta y_i = b_0 + b_1 \Delta x_1 + b_2 \Delta x_i^2 + \ldots + b_n \Delta x_i^n$$

Durch Multiplikation mit Δx_i^h und Summation über alle i von 1 bis k ($k > n + 1$) ergibt sich dann:

$$\sum x_i^h \Delta y_i = b_0 \sum \Delta x_i^h + b_1 \sum \Delta x_i^{h+1} + \ldots + b_n \sum \Delta x_i^{h+n}$$

Setzt man nun nacheinander $h = 0, 1, 2, \ldots$ bis n, dann erhält man ein System von $n + 1$ Gleichungen, aus welchem die $n + 1$ Koeffizienten b berechnet werden können. Vorausgesetzt wird, daß man den Grad der Funktion kennt.

Es wurde bereits früher erwähnt, daß die Intervallgleichheit bei der unabhängig Veränderlichen, welche Voraussetzung für die Anwendung des Differenzenschemas ist, gegebenenfalls durch eine passende Klasseneinteilung erreicht werden kann. Die Klassenbreite muß mindestens so groß gewählt werden, daß jede Klasse mit Werten der abhängig Veränderlichen besetzt ist. Den mit den Intervallmitten identischen x_i werden dann die Mittelwerte m_y^i aller der gleichen Klasse angehörenden y_i zugeordnet. Je nach der Zahl der an einem solchen Mittelwert beteiligten Einzelwerte wird den m_y^i grundsätzlich ein von Fall zu Fall verschiedenes Gewicht p_i zukommen und die Frage ist nun, in welcher Weise diese Gewichte im Differenzenschema zu berücksichtigen sind. Anders gesagt: Wie sind die Gewichte (Häufigkeitswerte) der Glieder der einzelnen Differenzenreihen aus denen der gegebenen m_y^i herzuleiten? Um diese Frage beantworten zu können, müssen wir wissen, wie sich die n-te Differenzreihe selbst aus der 0-ten, das heißt aus der Reihe der m_y^i ergibt.

Setzen wir das Intervall i der x-Werte gleich eins (i hat lediglich die Bedeutung eines Proportionalitätsfaktors), dann können wir die m_y^i der Reihe nach als $f_k, f_{k-1}, f_{k-2}, \ldots f_{k-n}$, das heißt als Werte der vorläufig noch unbekannten Funktion f (x) an den Stellen $x_k, x_{k-1}, x_{k-2}, \ldots x_{k-n}$ auffassen, welche mit Hilfe der Korrelationsrechnung dann so bestimmt werden muß, daß sie den gegebenen m_y^i und damit auch den y_i am ehesten gerecht wird.

Da die Gewichte p_i nicht dieser Funktion an und für sich, sondern bestimmten Funktionswerten zukommen, geben wir f (x) eine Form, die die Bedeutung der unabhängig Veränderlichen an vorgegebenen Stellen auch in den Differenzreihen am leichtesten erkennen läßt. Dafür eignet sich am besten die Reihenentwicklung nach Mac Laurin oder nach Taylor. Sie bricht nach unserer Annahme mit $f^{(n)}(x)$ ab, da wir die n-te Ableitung als konstant voraussetzen. Wir erhalten also folgende Aufstellung nach M. L.:

$$f_k = f_0 + \frac{k}{1!} f'_0 + \frac{k^2}{2!} f''_0 + \ldots + \frac{k^n}{n!} f^{(n)}_0$$

$$f_{k-1} = f_0 + \frac{k-1}{1!} f'_0 + \frac{(k-1)^2}{2!} f''_0 + \ldots + \frac{(k-1)^n}{n!} f^{(n)}_0$$

$$f_{k-2} = f_0 + \frac{k-2}{1!} f'_0 + \frac{(k-2)^2}{2!} f''_0 + \ldots + \frac{(k-2)^2}{n!} f^{(n)}_0$$

. . . . . .

$$f_{k-n-1} = f_0 + \frac{k-n-1}{1!} f'_0 + \frac{(k-n-1)^2}{2!} f''_0 + \ldots + \frac{(k-n-1)^n}{n!} f^{(n)}_0$$

Der Einfachheit halber wurde f_k, f_0, usw. für f (k), f (0): Funktionswert an der Stelle k, bzw. 0 gesetzt; entsprechend bedeutet $f^{(i)}_0$ die i-te Ableitung an der Stelle 0. Wegen i (Intervall) = 1 wird $x_k = k$, $x_{k-1} = k - 1$, usw. Die nachträgliche Wiedereinführung von i bereitet keine Schwierigkeiten.

1. Differenzenreihe

Wie leicht einzusehen ist, besteht die Aufgabe im wesentlichen darin, gleiche Potenzen voneinander zu subtrahieren, wobei die Basis des Minuenden stets um eins, d. h. das Intervall größer ist als die des Subtrahenden. Nach dem binomischen Lehrsatz gilt nun

$$a^m - (a-1)^m = \binom{m}{1} a^{m-1} - \binom{m}{2} a^{m-2} + \binom{m}{3} a^{m-3} - \ldots,$$

sodaß wir erhalten:

$$f_a - f_{a-1} = f'_o + \frac{2a-1}{2!} f''_o + \frac{3a^2 - 3a + 1}{3!} f'''_o +$$
$$+ \frac{4a^3 - 6a^2 + 4a - 1}{4!} f''''_o + \ldots\ldots$$

und daraus als erste Differenzenreihe:

$$f'_o + \frac{2k-1}{2!} f''_o + \frac{3k^2 - 3k + 1}{3!} f'''_o + \ldots.$$

$$\triangle_1 \quad f'_o + \frac{2(k-1)-1}{2!} f''_o + \frac{3(k-1)^2 - 3(k-1) + 1}{3!} f'''_o +$$

$$f'_o + \frac{2(k-2)-1}{2!} f''_o + \frac{3(k-2)^2 - 3(k-2) + 1}{3!} f'''_o + \ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Man erkennt, daß $f_k - f_{k-1}$ auf f_k; $f_{k-1} - f_{k-2}$ auf f_{k-1} zurückgeführt ist; Minuend und Differenz sind also gleichgewichtig.

Auch bei der Bildung der zweiten und aller folgenden Differenzen bedienen wir uns der Beziehung, die sich aus der Binomialformel ergibt; die Differenzen gleicher Potenzen werden wieder in Reihen entwickelt, wobei die Zurückführung der Differenz auf den Minuenden ein System von Potenzwerten mit gleicher Basis ergibt, so daß eine weitere Zusammenfassung unter Bildung für die Reihe typischer Koeffizienten möglich ist. Diese werden nach ihrer Zugehörigkeit zur zweiten, dritten, usw. Reihe mit II_i, III_i, usw. bezeichnet. Weitere Erklärungen sind bei der Übersichtlichkeit des Verfahrens nicht notwendig.

2. Differenzenreihe

$$\triangle_{2k} = \triangle_{1k} - \triangle_{1k-1}$$

$$\binom{m}{1} a^{m-1} - \binom{m}{2} a^{m-2} + \binom{m}{3} a^{m-3} - \ldots.$$
$$\binom{m}{1} (a-1)^{m-1} - \binom{m}{2} (a-1)^{m-2} + \binom{m}{3} (a-1)^{m-3} - \ldots.$$

$$\binom{m}{1}\left[\binom{m-1}{1} a^{m-2} - \binom{m-1}{2} a^{m-3} + \binom{m-1}{3} a^{m-4} - \binom{m-1}{4} a^{m-5} + \ldots\right]$$
$$\binom{m}{2}\left[\qquad - \binom{m-2}{1} a^{m-3} + \binom{m-2}{2} a^{m-4} - \binom{m-2}{3} a^{m-5} + \ldots\right]$$
$$\binom{m}{3}\left[\qquad\qquad + \binom{m-3}{1} a^{m-4} - \binom{m-3}{2} a^{m-5} + \ldots\right]$$
$$\binom{m}{4}\left[\qquad\qquad\qquad - \binom{m-4}{1} a^{m-5} + \ldots\right]$$

usw.

$$= \Pi_1 \binom{m}{2} a^{m-2} - \Pi_2 \binom{m}{3} a^{m-3} + \Pi_3 \binom{m}{4} a^{m-4} - \Pi_4 \binom{m}{5} a^{m-5} + \ldots,$$

wobei

$$\Pi_1 = \binom{2}{1} = 2^2 - 2$$

$$\Pi_2 = \binom{3}{1} + \binom{3}{2} = 2^3 - 2$$

$$\Pi_3 = \binom{4}{1} + \binom{4}{2} + \binom{4}{3} = 2^4 - 2$$

$$\Pi_5 = \binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} = 2^5 - 2$$

$$\ldots\ldots$$

$$\Pi_i = \binom{i+1}{1} + \binom{i+2}{2} + \ldots + \binom{i+1}{i} = 2^{i+1} - 2$$

zu setzen ist. Da bei der Bildung der Differenzwerte in erster Linie der Wegfall der Ableitung und die Beziehung der Differenz zu denen der vorhergehenden Reihe, welche an ihrer Bildung beteiligt sind, interessiert, sind nur die beiden ersten Glieder der Reihenentwicklung für irgend ein Δ_2 angegeben. Man findet:

$$f''_0 + (k-1) f'''_0 + \ldots \qquad = f''_0 + \frac{3!k - \Pi_2}{3!} f'''_0 + \ldots$$

$$\triangle_2 \quad f''_0 + ((k-1)-1) f''_0 + \ldots$$

$$f''_0 + ((k-2)-1) f'''_0 + \ldots$$

$$\ldots\ldots$$

3. Differenzenreihe

$$\triangle_{3k} = \triangle_{2k} - \triangle_{2k-1}$$

$$\Pi_1 \binom{m}{2} a^{m-2} \quad - \Pi_2 \binom{m}{3} a^{m-3} \quad + \Pi_3 \binom{m}{4} a^{m-4} \quad - \ldots$$

$$\Pi_1 \binom{m}{2} (a-1)^{m-2} - \Pi_2 \binom{m}{3} (a-1)^{m-3} + \Pi_3 \binom{m}{4} (a-1)^{m-4} - \ldots$$

$$\ldots\ldots\ldots\ldots$$

$$\Pi_1 \binom{m}{2} \left[\binom{m-2}{1} a^{m-3} - \binom{m-2}{2} a^{m-4} + \binom{m-2}{3} a^{m-5} - \binom{m-2}{4} a^{m-6} + \ldots\right]$$

$$\Pi_2 \binom{m}{3} \left[(\qquad - \binom{m-3}{1} a^{m-4} + \binom{m-3}{2} a^{m-5} - \binom{m-3}{3} a^{m-6} + \ldots\right]$$

$$\Pi_3 \binom{m}{4} \left[(\qquad + \binom{m-4}{1} a^{m-5} - \binom{m-4}{2} a^{m-6} + \ldots\right]$$

$$\Pi_4 \binom{m}{5} \left[(\qquad - \binom{m-5}{1} a^{m-6} + \ldots\right]$$

usw.

$$= III_1 \binom{m}{3} a^{m-3} - III_2 \binom{m}{4} a^{m-4} + III_3 \binom{m}{5} a^{m-5} - III_3 \binom{m}{6} a^{m-6} + .$$

$$III_1 = \binom{3}{2} II_1$$

$$III_2 = \binom{4}{2} II_1 + \binom{4}{3} II_2$$

$$III_3 = \binom{5}{2} II_1 + \binom{5}{3} II_2 + \binom{5}{4} II_3$$

$$\ldots\ldots$$

$$III_i = \binom{i+2}{2} II_1 + \binom{i+2}{3} II_2 + \ldots + \binom{i+2}{i+1} II_i$$

Differenzen:

$$f'''_0 + \frac{24k - 36}{4!} f''''_0 + \ldots = f'''_0 + \frac{4!\,k - III_2}{4!} f''''_0 + \ldots$$

$$\triangle_3 \quad f'''_0 + \frac{24(k-1) - 36}{4!} f''''_0 + \ldots$$

$$f'''_0 - \frac{24(k-2) - 36}{4!} f''''_0 + \ldots$$

$$\ldots\ldots$$

4. Differenzenreihe

a. Koeffizienten:

$$IV_1 = \binom{4}{3} III_1$$

$$IV_2 = \binom{5}{3} III_1 + \binom{5}{4} III_2$$

$$IV_3 = \binom{6}{3} III_1 + \binom{6}{4} III_2 + \binom{6}{5} III_3$$

$$\ldots\ldots$$

$$IV_i = \binom{i+3}{3} III_1 + \ldots + \binom{i+3}{i+2} III_i$$

Reihe: $IV_1 \binom{m}{4} a^{m-4} a - IV_2 \binom{m}{5} a^{m-5} + IV_3 \binom{m}{6} a^{m-6} -$

b) Differenzen:

$$f^{IV}_0 + \frac{5!\,k - IV_2}{5!} f^{V}_0 + \ldots$$

$$\triangle_4 \quad f^{IV}_0 + \frac{5!\,(k-1) - IV_2}{5!} f^{V}_0 + \ldots$$

$$f^{IV}_0 + \frac{5!\,(k-2) - IV_2}{5!} f^{V}_0 + \ldots$$

$$\ldots\ldots$$

n. Differenzenreihe

Die Koeffizienten dieser Reihe seien mit N_i, die der vorhergehenden mit M_i bezeichnet.

a) Koeffizienten:

$$N_1 = \binom{n}{n-1} M_1$$

$$N_2 = \binom{n+1}{n-1} M_1 + \binom{n+1}{n} M_2$$

$$N_3 = \binom{n+2}{n-1} M_1 + \binom{n+2}{n} M_2 + \binom{n+2}{n+1} M_3$$

$$\ldots\ldots$$

$$N_i = \binom{n+i-1}{n-1} M_1 + \binom{n+i-1}{n} + \ldots + \binom{n+i-1}{n+i-2} M_i$$

Reihe: $N_1 \binom{m}{n} a^{m-n} - N_2 \binom{m}{n+1} a^{m-n-1} + N_3 \binom{m}{n+2} a^{m-n-2} - \ldots$

b) Differenzen:

Der Koeffizient von $f^{(n)}{}_0$ ergibt sich zu:

$$\frac{1}{n!}\binom{n}{n-1}\binom{n-1}{n-2}\binom{n-2}{n-3}\ldots\binom{4}{3}\binom{3}{2}\binom{2}{1} = 1$$

$$f^{(n)}_0 + \frac{n!\,k - N_2}{n!} f^{(n+1)}{}_0 + \ldots$$

$$\triangle_n \quad f^{(n)}_0 + \frac{n!\,(k-1) - N_2}{n!} f^{(n+1)}_0 + \ldots$$

$$f^{(n)}_0 + \frac{n!\,(k-2) - N_2}{n!} f^{(n+1)}_0 + \ldots$$

$$\ldots\ldots$$

Auf Grund der Ergebnisse ist also folgendes festzustellen:

1. Ist eine Funktion in einem von a bis b reichenden Gebiet stetig und n-mal differenzierbar und ist $f^{(n)}$ konstant, dann entspricht die n-te Reihe eines auf $n + 1$ Funktionswerten aufgebauten Differenzenschemas der n-ten Ableitung der Funktion.

2. Eine arithmetische Reihe n-ter Ordnung läßt sich durch eine ganze rationale Funktion n-ten Grades ausdrücken.

3. Jeder Wert des Differenzenschemas läßt sich auf einen bestimmten Funktionswert zurückführen und aus diesem auch direkt

berechnen, wenn die Funktion bekannt ist. Daraus ergibt sich auch die Gewichtsgleichheit bei empirisch bestimmtem Funktions- und daraus abzuleitendem Differenzwert. Die Bindung gleichgewichtiger Differenzwerte des Schemas ist immer eine Bindung zwischen Minuend und Differenz.

Arg.	Funkt.	Gewicht	Differenzgewichte				
x_1	f_1	p_1					
			p_1				
x_2	f_2	p_2		p_1			
			p_2		p_1		
x_3	f_3	p_3		p_2		p_1	
			p_3		p_2		p_1
x_4	f_4	p_4		p_3		p_2	
			p_4		p_3		
x_5	f_5	p_5		p_4			
			p_5				
x_6	f_6	p_6					

$$x_1 - x_2 = x_2 - x_3 = x_3 - x_4 = \ldots = i$$

Nachdem nun die Frage nach der Herleitung der Differenz- aus den Funktionswerten bzw. der Gewichte der Differenzwerte aus denen der Funktionswerte beantwortet worden ist, haben wir uns noch um die Bildung der Gewichte zu bekümmern. Es ist bereits erwähnt worden, daß die f_i als Mittelwerte aller y_i aufzufassen sind, die je einer Klasse angehören. Definiert man nun die Gewichte lediglich als Häufigkeitswerte, dann würde das Gewicht P_i eines Mittels $f_i = m_y^i$ (Mittelwert aller y, die der i-ten Klasse angehören) gleich n_i, das heißt der Zahl aller an der Bildung dieses Mittelwertes beteiligten Einzelwerte sein. Das geht nun aber nicht an, weil die Bedeutung eines Mittels nicht nur durch n, sondern auch durch die Streuung, oder den mittleren Fehler gekennzeichnet ist, und zwar so, daß mit wachsendem mittlerem Fehler das Gewicht des Mittelwertes abnimmt.

Schreiben wir die gewogenen Mittel Y_i und Y_k in folgender Form:

$$Y_i = \frac{p_{i1} y_{i1} + p_{i2} y_{i2} + \ldots + p_{in} y_{in}}{p_{i1} + p_{i2} + \ldots + p_{in}} = \frac{[p_i y_i]}{[p_i]} = \frac{[p_i y_i]}{P_i}$$

$$Y_k = \frac{p_{k1} y_{k1} + p_{k2} y_{k1} + \ldots + p_{kn} y_{kn}}{p_{k1} + p_{k2} + \ldots + p_{kn}} = \frac{[p_k y_k]}{[p_k]} = \frac{[p_k y_k]}{P_k}$$

dann sind P_i (Summe aller p_i) und P_k (Summe aller p_k) reine Häufigkeitswerte, wie p_i und p_k. In dieser Form hat P_i nur Vergleichswert

den p_i, nicht aber P_k gegenüber und umgekehrt. Diesen Mangel können wir beheben, wenn wir für die p_i und p_k eine „Gewichtseinheit" einführen, die die p mit den mittleren Fehlern m der y_i in Beziehung zu setzen gestattet. Nehmen wir vorerst an, daß jedes y_i sich selber als Mittelwert einer Reihe von Bestimmungen ergeben habe, von denen jede mit dem mittleren Fehler μ zu behaften sei, dann erhält man diesen mittleren Fehler zu

$$m_i = \frac{1}{p_i} \sqrt{p_i \cdot \mu^2} = \frac{\mu}{\sqrt{p_i}},$$

und für das Gewicht von y_i findet man dementsprechend:

$$p_i = \frac{\mu^2}{m_i^2},$$

wobei μ die Gewichtseinheit der p_i ist. Setzen wir nun den für die p_i gefundenen Ausdruck in der Formel für Y_i ein, dann erhalten wir

$$Y_i = \frac{\left[\frac{\mu^2}{m_i^2} y_i\right]}{\left[\frac{\mu^2}{m_i^2}\right]} \quad \text{und entsprechend}$$

$$Y_k = \frac{\left[\frac{\mu^2}{m_k^2} y_k\right]}{\left[\frac{\mu^2}{m_k^2}\right]}$$

Man sieht, daß in beiden Ausdrücken μ^2 herausfällt bzw. daß die Gewichtseinheit μ beliebig gewählt werden kann. Um Y_i und Y_k vergleichbar zu machen, wählen wir für beide die selbe Gewichtseinheit (was oben übrigens bereits vorausgenommen wurde), so daß wir an Stelle von P_i und P_k die modifizierten Gewichte Q_i und Q_k in folgender Form erhalten:

$$Q_i = \mu^2 \left[\frac{1}{m_i^2}\right] = \mu^2 \left(\frac{1}{m_{i1}^2} + \frac{1}{m_{i2}^2} + \ldots + \frac{1}{m_{in}^2}\right)$$

$$Q_k = \mu^2 \left[\frac{1}{m_k^2}\right] = \mu^2 \left(\frac{1}{m_{K1}^2} + \frac{1}{m_{K2}^2} + \ldots + \frac{1}{m_{Kn}^2}\right)$$

Weil nur das Verhältnis von Q_i zu Q_k von Bedeutung ist, kann die Proportionalitätskonstante μ^2 weggelassen werden. Dazu kommt nun

noch, daß alle yi als Einzelwerte, soweit sie derselben Klasse angehören, denselben mittleren Fehler aufweisen, so daß man für die vergleichbaren Gewichte folgende endgültige Form erhält:

$$Q_i = \frac{n_i}{m_i^2} = \frac{1}{M_i^2}$$

$$Q_k = \frac{n_k}{m_k^2} = \frac{1}{M_k^2}$$

n_i (= P_i) ist dabei die Zahl aller y_i, die der i-ten Klasse angehören, m_i der mittlere Fehler dieser y_i und m_i der Fehler des Mittelwertes. Man erkennt, daß die Gewichte Q die an sie gestellten Bedingungen erfüllen: Sie nehmen mit wachsendem n zu und mit wachsendem mittlerem Fehler ab. Zwischen P und Q besteht folgende Beziehung:

$$P = m^2 \cdot Q$$

Bei der vorauszusetzenden korrelativen Bindung zwischen x und y, die jedem x_i eine Schar von y-Werten zuweist, von denen einer, nämlich der Mittelwert, der wahrscheinlichste ist, hat man sich zu merken, daß auch bei einer durch eine Teilung in Intervalle oder Klassen bewirkten Zusammenfassung der y_i je für die, welche derselben Klasse angehören, ein wahres Mittel gemäß dem Gaußschen Verteilungsgesetz zu definieren ist, und zwar um so eher, je kleiner die Intervallbreite i ist. Dieses wahre Mittel ist nun nicht identisch mit dem arithmetischen und darum darf auch nicht der mittlere Fehler als Mittel der Quadrate aller Abweichungen gegenüber dem arithmetischen Mittel berechnet werden.

Bezeichnen wir das nicht bekannte, wahre Mittel mit M, das arithmetische mit M_a, die Einzelwerte, aus denen M_a berechnet wurde, mit a_i, die Abweichungen (Fehler) von M_a mit v_i, von M mit v_{wi}, die Differenz zwischen M_a und M mit m (der wahre Fehler des Mittelwerts wird dabei dem mittleren gleichgesetzt), dann gilt:

$$m = M_a - M$$

$$v_i = a_i - M$$

$$v_{wi} = a_i - M = a_i - M_a + m = v_i + m$$

$$[v_w v_w] = [vv] + 2m[v] + n \cdot m^2$$

$$= [vv] + n \cdot m^2$$

Nun ist ferner

$$v_w v_w = n \cdot \mu^2$$

wenn mit μ der mittlere Fehler eines Einzelwertes bezeichnet wird — und

$$m = \frac{1}{n}\sqrt{n \cdot \mu^2} = \frac{\mu}{\sqrt{n}}$$

Daraus folgt:

$$n \cdot \mu^2 = [vv] + \mu^2$$

$$\mu = \sqrt{\frac{[vv]}{n-1}} \qquad m = \sqrt{\frac{[vv]}{n(n-1)}}$$

m. F. eines Einzel-, des Mittelwertes.

An Stelle von μ und m wäre wie oben m und M zu setzen (m. F. = mittlerer Fehler).

Bei der Bestimmung der Differenzwerte wurde das Intervall $i = 1$ gesetzt. Für $i \neq 1$ ändert sich der Ausdruck für eine n-te Differenzenreihe

$$\triangle_n \quad \begin{array}{ll} \text{von} & f_o^{(n)} + A \cdot f_o^{(n+1)} + B \cdot f_o^{(n+2)} + \ldots \\ \text{in} & i^n f_o^{(n)} + A \cdot i^{n+1} f_o^{(n+1)} + B \cdot i^{n+2} f_o^{(n+2)} \end{array}$$

Entwickelt man die Funktion statt an der Stelle o nach Taylor an der Stelle a, dann findet man:

$$f_{(a+k)} = f_a + \frac{k}{1!} f'_a + \frac{k^2}{2!} f''_a + \ldots$$

$$f_{(a+k-1)} = f_a + \frac{k-1}{1!} f'_a + \frac{(k-1)^2}{2!} f''_a + \ldots$$

Für die Koeffizienten ergeben sich die gleichen Werte wie bei der Entwicklung der Funktion in einer Reihe nach Mac Laurin. Man hat in den Differenzen an Stelle der f^i_o die selben Ableitungen an der Stelle a: f^i_a zu setzen.

Zusammenfassend ist folgendes zu sagen:

Läßt sich eine korrelative Beziehung zwischen zwei Größen am besten durch eine ganze, rationale Funktion n-ten Grades ausdrücken

— was sich entweder aus der Darstellung der Beziehung in einem Koordinatensystem, oder aus einem Differenzenschema, das sich auf den Werten der bedingt Veränderlichen („Funktionswerten") aufbaut, ergibt —, dann ändern sich die Werte der (n — 1)-ten Differenzenreihe mehr oder weniger linear mit dem in gleichen Intervallen fortschreitenden Argument und können daher mit diesem zusammen der Korrelationsrechnung zugrunde gelegt werden.

Aus den Regressionsgleichungen erhält man durch Integration die Funktionen n-ten Grades, die dem Zusammenhang am besten gerecht zu werden vermögen. Bei sehr engem Zusammenhang, d. h. wenn r einen von $|1|$ nur wenig abweichenden Wert hat, wird man die beiden Regressionsgleichungen auch vereinigen können (vgl. 4. Beispiel).

Zur Bestimmung der n — 1 noch unbekannten Koeffizienten kann ebenfalls das Differenzenschema benutzt werden: Man baut es, bei der n-ten Differenzenreihe, die b_{21}, oder b_{12} als konstanten Wert aufweist, beginnend nach rückwärts wieder auf, wobei jede neu aufzustellende Reihe der entsprechenden des gegebenen Schemas so angeglichen wird, daß die Summe der Abweichungen 0 ergibt. Aus beliebigen n — 1 der schließlich so gefundenen Reihe der Funktionswerte können dann die Koeffizienten bestimmt werden.

Hat man zur Gewinnung gleicher Intervalle beim Argument (x) eine passende Klasseneinteilung vorgenommen, dann ordnet man den Intervallmitten das Mittel aller y_i, die dieser Klasse angehören, zu. Jedes dieser Mittel ist gemäß seinem Gewicht zu berücksichtigen. Das Gewicht ist proportional der Anzahl der y_i, aus denen das Mittel berechnet wurde und umgekehrt proportional dem Quadrat des mittleren Fehlers, bzw. der Streuung. Bei der Berechnung des mittleren Fehlers hat man zu berücksichtigen, daß für die derselben Klasse angehörenden y_i ein wahres Mittel existiert. Im Differenzenschema geht das Gewicht gemäß der funktionellen Beziehung immer vom Minuenden auf die Differenz über (vgl. die Darstellung pag. 57).

6. Beispiel: Es betrifft einen Zusammenhang, der durch den Übergang zu den Logarithmen der Auswertung zugänglich gemacht werden kann: Beziehung zwischen Arbeitslosigkeit und Preisgestaltung in Großbritannien, Januar 1920 bis Juni 1921. Da hier nur das Verfahren interessiert, sei als einziger Kommentar ein Passus aus „La Crise de Chômage 1920—1923" (Bureau International du Travail, Genève 1924) angeführt, welcher Arbeit übrigens auch das Zahlenmaterial entnommen ist.

„De même qu'aux Etats-Unis et au Japon, on constate en Grande-Bretagne une relation étroite entre les statistiques des prix de gros et du chômage.

Pendant la période d'activité d'avril 1919 à avril 1920, où les prix sont montés de 95,9 points en douze mois, le pourcentage de chômage est tombé pendant une certaine période au-dessous de 1 pour cent. De plus, malgré l'augmentation considérable du nombre des grèves, la transformation des industries de guerre en industries de paix et la démobilisation rapide, le pourcentage mensuel moyen le plus élevé enregistré par les syndicats pendant toute la période de prospérité n'a été que de 3,2. La baisse des prix survenue en mai 1920 a coïncidé avec l'augmentation du chômage et, en juin 1921, lorsque le chômage atteignait 23,1 pour cent, les prix étaient tombés de 129,8 points." (pag. 21/22.)

	Zeit	P%	A%		Zeit	P%	A%
1920	J	288,6	2,9	1920	O	282,2	5,3
	F	306,3	1,6		N	263,3	3,7
	M	308,0	1,1		D	243,8	6,1
	A	313,1	0,9				
	M	305,9	1,1	1921	J	232,0	6,9
	J	300,8	1,2		F	215,3	8,5
	J	299,5	1,4		M	208,5	10,0
	A	298,2	1,6		A	199,8	17,6
	S	292,6	2,2		M	190,8	22,2
					J	183,3	23,1

Wir stellen also fest, daß die Anwendung der Korrelationsrechnung in diesem Fall sachlich gerechtfertigt ist. Die in der Tabelle angeführten Zahlenwerte (P = Großhandelspreise, A = Arbeitslosenzahl in % der Gewerkschaftsmitglieder) lassen ohne weiteres die Gegensätzlichkeit von Preisbewegung und Änderung der Arbeitslosigkeit erkennen. Noch deutlicher zeigt dies die folgende graphische Darstellung.

Eine Unregelmäßigkeit im Kurvenverlauf (A) ist nach den Angaben der „Crise de Chômage" auf eine Streikbewegung im Oktober 1920, die eine unverhältnismäßig starke Zunahme der Arbeitslosigkeit

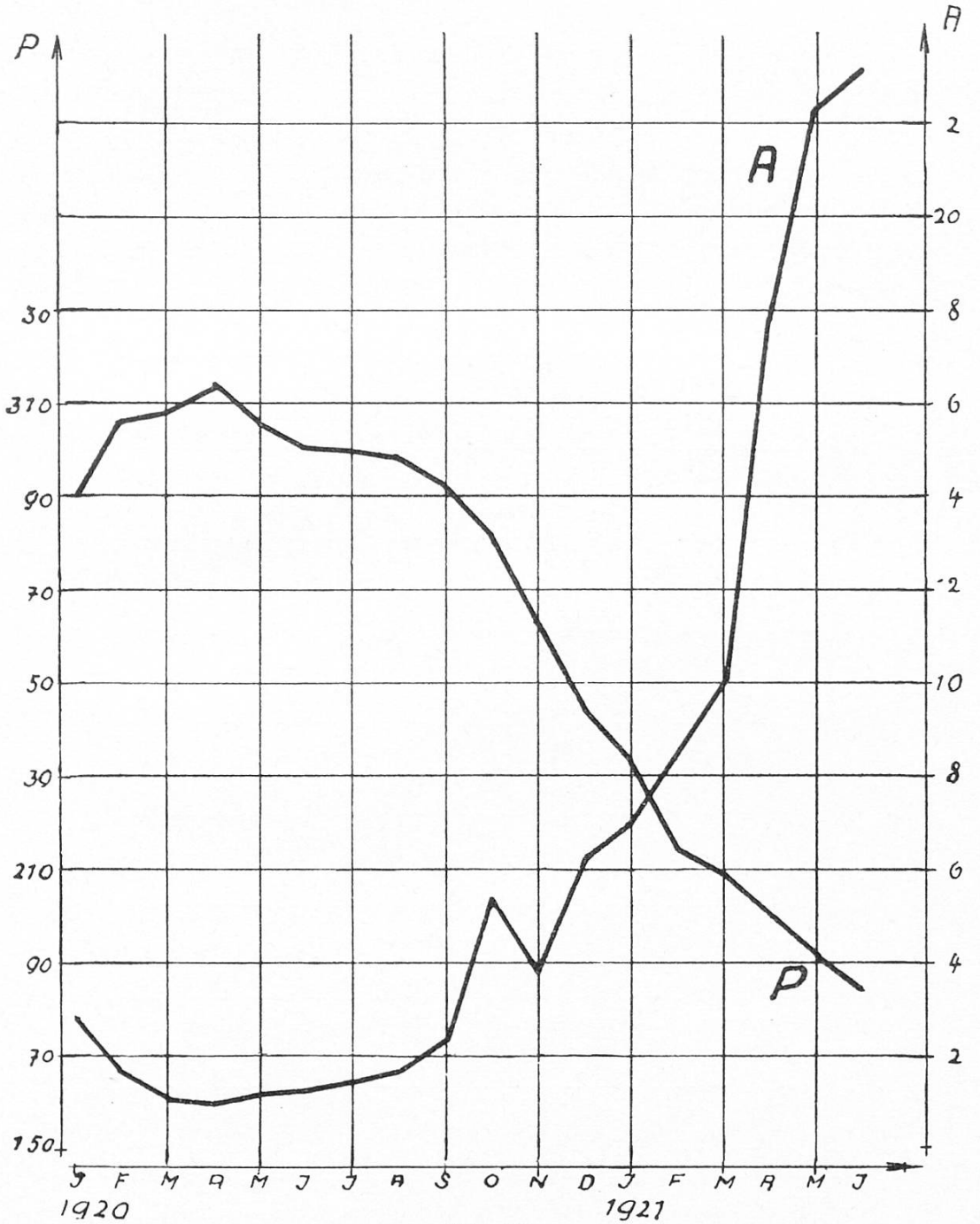

Fig. 12

in diesem Monat, und — nach Beilegung des Arbeitskonflikts — im folgenden eine vorübergehende Erholung auf dem Arbeitsmarkt zur Folge hatte, zurückzuführen.

Im übrigen läßt sich unschwer erkennen, daß die Änderung von P zeitlich der von A vorangeht: Während unter anderm die Preissenkung im April-Mai 1920 einsetzt, ist eine deutliche Zunahme der Arbeitslosigkeit erst im September-Oktober zu beobachten und die vom Dezember desselben Jahres an zu konstatierende Abschwächung

des Preisfalls wirkt sich anscheinend im Mai 1921 auf den Gang der Beschäftigung aus. Aus dieser Priorität der Preisbewegung (sie wird übrigens durch das Int. Arbeitsamt in Cr. d. Ch., pag. 22, 137 u. a. bestätigt) ergibt sich für die Korrelationsrechnung der Schluß, daß P als bedingende und A als bedingte Veränderliche aufzufassen ist.

Es genügt nun aber nicht, zu wissen, daß P und A miteinander in Beziehung stehen und welche Größe durch die andere bedingt ist;

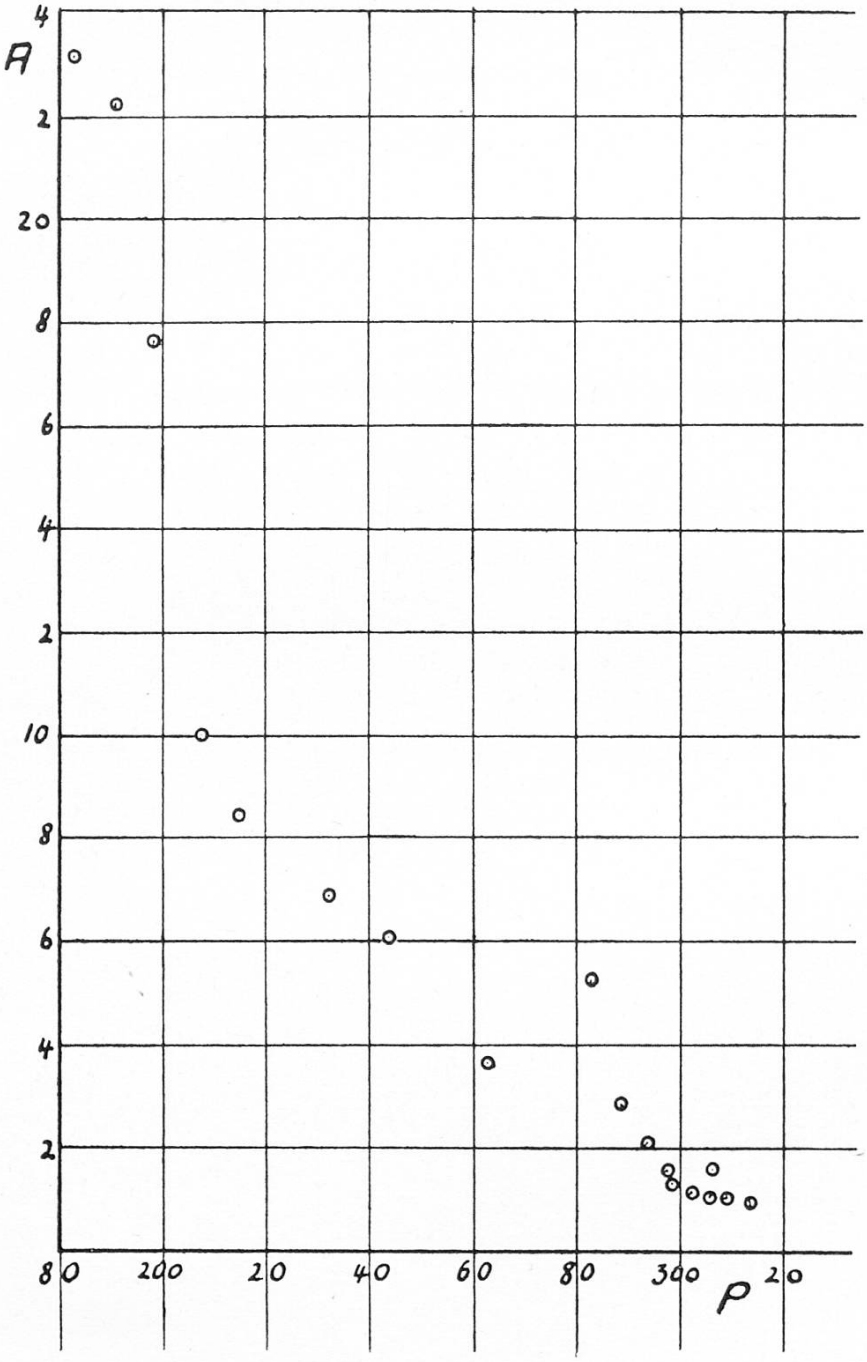

Fig. 13

wir können mit der Korrelationsrechnung erst dann einsetzen, wenn wir uns über die Art des Zusammenhangs orientiert haben — wenigstens so weit, daß wir uns klar sind darüber, ob er linear ist, oder nicht. Fig. 12 zeigt nun deutlich, daß weder in der Änderung von P in der Zeit, noch in der Änderung von A in der Zeit irgendwelche Gesetzmäßigkeit enthalten ist; weder P noch A sind zeitbedingt. Würde man die drei senkrecht aufeinander stehenden Achsen eines Raumkoordinatensystems P, A und der Zeit T zuordnen, dann wäre die durch die drei Größen bedingte Raumkurve also dadurch gekennzeichnet, daß höchstens ihre Projektion auf die PA-Ebene (Ebene durch die P- und die A-Achse) einen irgendwie gesetzmäßigen Verlauf zeigen würde. Man erhält sie, wenn man den Kurvenbildern von P und A in Fig. 12, welche den Projektionen der Raumkurve in der PZ- und der AZ-Ebene entsprechen, zusammengehörende Werte entnimmt. Sie sind die Koordinaten der Punkte in der PA-Ebene.

In der nebenstehenden Figur 13 ist die Zeitkomponente ausgeschaltet. Man erkennt, daß die Punkte auch nicht annähernd auf einer geraden Linie liegen und daß infolgedessen eine direkte Auswertung des Zahlenmaterials gar nicht in Frage kommen kann. Rechnet man nun mit der Möglichkeit, daß eine Art hyperbolischer Kurve der Lage der Punkte im System eher gerecht zu werden vermöchte, dann könnte man den Zusammenhang im allgemeinsten Fall etwa in folgender Form ausdrücken:

$$P^a \cdot A^b = C, \quad \text{oder}$$

$$P \cdot A^{b/a} = C',$$

wobei a, b, C, bzw. C' beliebige Konstanten wären. Eine solche Beziehung wird nun linear, wenn man zu den Logarithmen übergeht. Wir erhalten dann:

$$\log P + \frac{b}{a} \log A = C''$$

Die Rechnung wird nun zeigen, ob und wie weit unsere Vermutung gerechtfertigt ist. In der nachfolgenden Tabelle ist x für log P und y für log A gesetzt. Angenommene Mittelwerte:

$$m'_x = 2{,}420 \quad m'_y = 0{,}814$$

	x	y	Δ x —	Δ x +	Δ y —	Δ y +	Δ x²	Δ y²
1920								
J	2,461	0,462		0,041	0,352		0,00168	0,123
F	486	204		066	610		436	372
M	489	041		069	773		476	598
A	496	9,954		076	860		578	740
M	486	0,041		066	773		436	598
J	479	079		059	735		348	540
J	476	146		056	668		314	446
A	474	204		054	610		292	372
S	466	333		046	481		212	231
O	450	724		030	090		090	008
N	420	568		000	246		000	061
D	387	786	0,033		028		109	001
1921								
J	365	839	055			0,025	302	001
F	333	929	087			115	757	013
M	320	1,000	100			186	1000	035
A	301	246	119			432	1416	187
M	281	346	139			532	1932	284
J	264	364	156			550	2434	303
	Summen		0,689	0,563	6,226	1,840	0,11298	4,910
			0,126		4,386			

$$m_x = 2{,}420 - \frac{1}{18} \cdot 0{,}063 = 2{,}413 \qquad \Sigma\Delta x^2 = 0{,}11298 - \frac{0{,}01588}{18}$$

$$= 0{,}11210$$

$$m_y = 0{,}814 - \frac{1}{18} \cdot 4{,}386 = 0{,}570 \qquad \Sigma\Delta y^2 = 4{,}910 - \frac{19{,}237}{18}$$

$$= 3{,}841$$

	$\Delta x \Delta y$ —	+		$\Delta x \Delta y$ —	+		$\Delta x \Delta y$ —	+
1	0,0144		7	0,0374		13	0,0014	
2	403		8	329		14	100	
3	533		9	221		15	186	
4	654		10	027		16	514	
5	510		11	000		17	739	
6	434		12		0,0009	18	858	

Summe: — 0,6041
+ 0,0009 — 0,6032

$$\Sigma \Delta x \Delta y = -0{,}6032 - \frac{0{,}5526}{18} = -0{,}6339$$

$$r = -\frac{0{,}6339}{\sqrt{0{,}1121 \cdot 3{,}841}} = -0{,}967 \pm 0{,}016$$

$$b_{21} = -\frac{0{,}6339}{0{,}1121} = -5{,}654 \qquad p' = -\frac{0{,}0688}{5{,}81} = -0{,}01185$$

$$\delta' = 40{,}'8$$

$$b_{12} = -\frac{0{,}6339}{3{,}841} = -0{,}1650 \qquad p = -\frac{0{,}0688}{1{,}930} = -0{,}0357$$

$$\delta = 2^0\ 18'$$

Aus den rechnerischen Ergebnissen ist zu entnehmen, daß durch den Übergang zu den Logarithmen es tatsächlich gelungen ist, die Beziehung zu „strecken", das heißt sie in eine lineare umzuwandeln. Der hohe Wert des Korrelationskoeffizienten wäre sonst nicht denkbar. Der Tatbestand wäre also für Großbritannien und für den Zeitabschnitt vom Januar 1920 bis Juni 1921 folgendermaßen auszudrücken:

Trägt man auf der Abszisse eines Koordinatensystems die Logarithmen der Großhandels-Indexziffern und auf der Ordinate die Logarithmen der Arbeitslosenziffern ab, dann liegen die Punkte, die zusammengehörende Werte als Koordinaten haben, annähernd auf einer Geraden.

Der Richtungskoeffizient hat einen zwischen — 5,654 und — 6,060 liegenden Wert: Einer Abnahme von P entspricht eine Zunahme von A; die Änderung des Logarithmus von A, gemessen an der des Logarithmus von P, ist 5,654, bzw. 6,060mal größer.

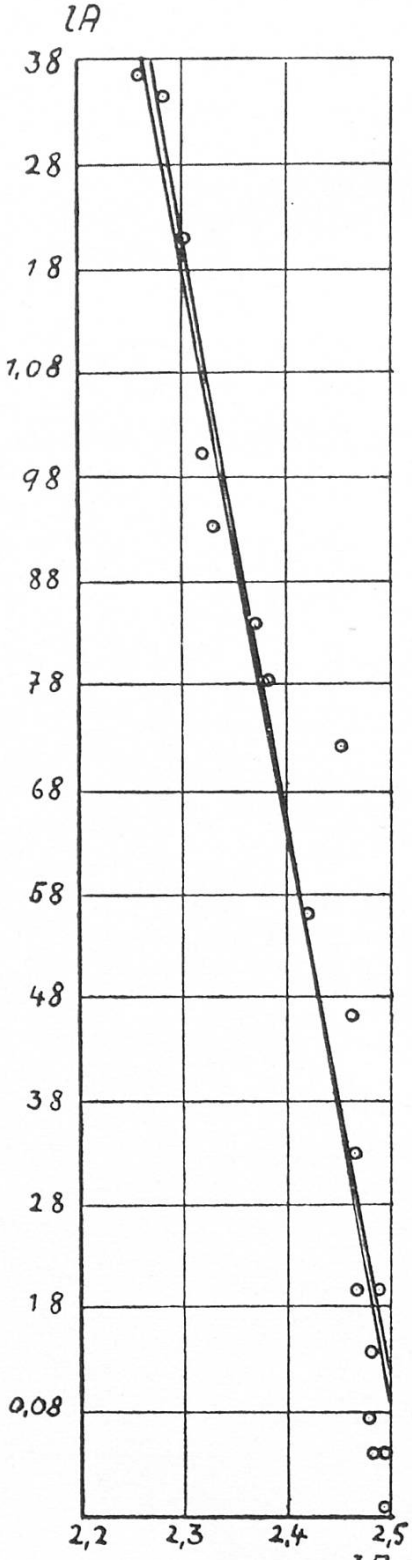

Fig. 14

Die Formulierung ist eine ähnliche, wie in dem nach Pareto benannten Gesetz, das über die korrelativen Beziehungen zwischen Einkommenskategorien und zugehörigen Personenzahlen eine entsprechende Aussage macht. Im vorliegenden Fall aber handelt es sich, wenn auch an der Beziehung als solcher nicht zu zweifeln ist, nicht um ein Ergebnis von allgemeiner Bedeutung. Die Aufgabe bestand nach der sachlichen Rechtfertigung der Anwendung der Korrelationsmethode (vgl. die Einleitung zu diesem Beispiel) in erster Linie darin, die Beziehung in eine lineare umzuformen. Aus dem hier erzielten positiven Ergebnis ist nicht einfach der Schluß zu ziehen, daß bei weiteren Untersuchungen dieser Art das Resultat — soweit dies die Art des Zusammenhangs betrifft — das gleiche sein müßte; es ist wohl möglich, aber nicht sicher, daß man auch für andere Staaten und für andere Zeitabschnitte zu entsprechenden Feststellungen gelangt. Für die Korrelationsrechnung heißt dies aber, daß ihre Anwendung im Dienste der wissenschaftlichen Forschung nur in einer möglichst großen Zahl gleichartiger Fälle sinnvoll und zweckmäßig sein kann.

Die beiden folgenden bildlichen Darstellungen zeigen die Regressionsgeraden und die beim Übergang zum Numerus aus ihnen sich ergebenden Kurven. In beiden Fällen sind auch die Wertepunkte eingetragen. Man sieht, daß die Streuung, das heißt die Abweichung von den errechneten Beziehungslinien, verhältnismäßig gering ist. Nur der früher schon erwähnte Oktoberwert macht eine Ausnahme. Die beiden Geraden (Fig. 14) weichen nur um

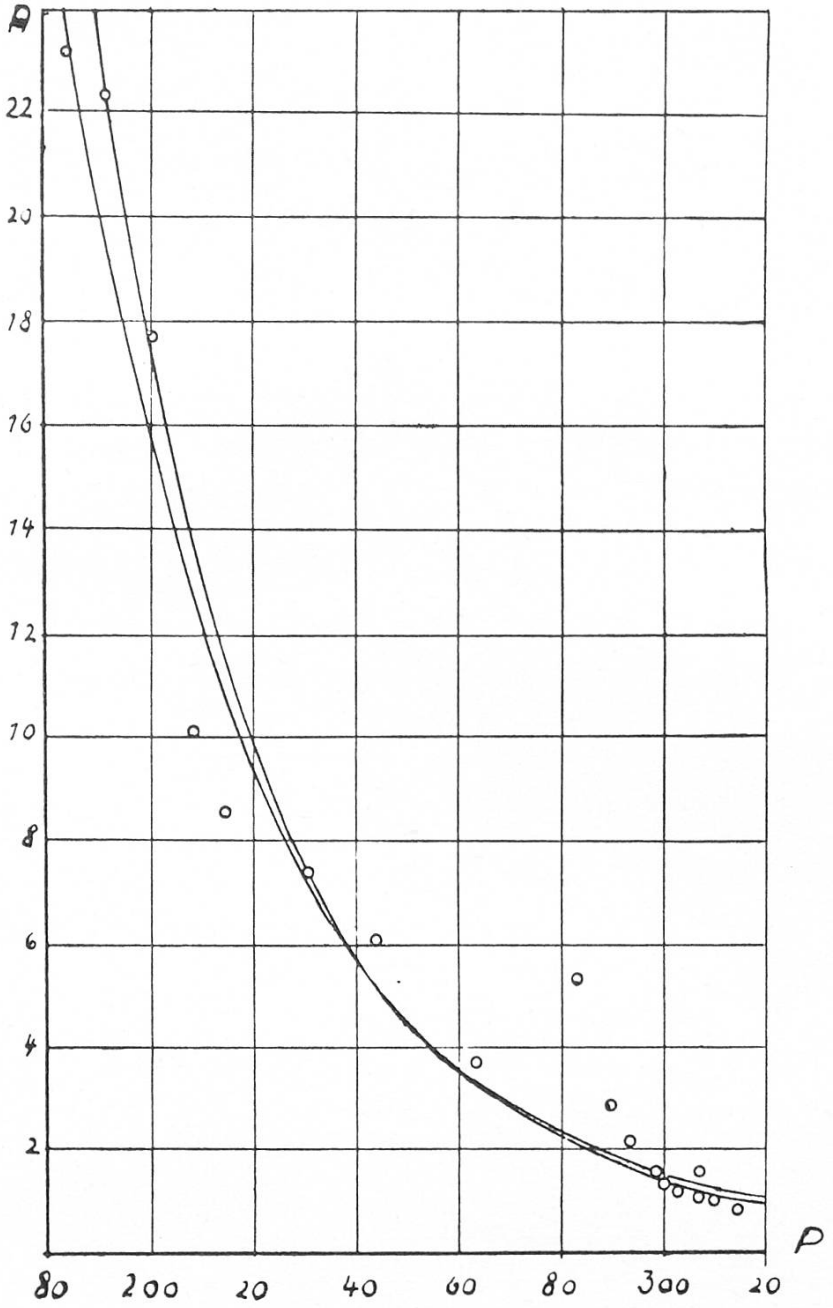

Fig. 15

41' voneinander ab — eine selbstverständliche Folge der aus der Rechnung sich ergebenden engen Beziehung. Die Kurven lassen sich für Indexwerte über 260 nicht mehr gut auseinanderhalten. Aus den Regressionskoeffizienten ergeben sich die Beziehungen folgendermaßen:

$$\text{a)}\quad \log A - 0{,}570 = -5{,}654\,(\log P - 2{,}413)$$
$$\log A = -5{,}654 \log P + 14{,}213$$
$$\text{b)}\quad \log P - 2{,}413 = -0{,}165\,(\log A - 0{,}570)$$
$$\log P = -0{,}165 \log A + 2{,}507$$

Weiter erhalten wir:

$$\text{a)}\quad A = 1{,}634 \cdot 10^{14} \cdot P^{-5{,}654} \quad (K_1)$$

$$\text{b)}\quad P = 3{,}216 \cdot 10^{2} \cdot A^{-0{,}165}$$

$$\text{oder}\quad A = 1{,}550 \cdot 10^{15} \cdot P^{-6{,}061} \quad (K_2)$$

An dieser Stelle mag noch erwähnt werden, daß der hier untersuchte wenigstens keinen Einzelfall darstellt: Für die Schweiz konnten für die Zeit vom Januar 1921 bis Februar 1922 (Material aus dem Stat. Jahrb. der Schweiz; do. Cr. d. Ch.) für den Zusammenhang zwischen log A und log P ein Korrelationskoeffizient von ebenfalls 0,967 und Regressionskoeffizienten von — 5,22 und — 0,18, also Werten, die nur wenig von denen des vorliegenden Beispiels abweichen, festgestellt werden.

Die Wahl anderer Maßstäbe

Man müßte der Korrelationsrechnung die Objektivität absprechen, wenn ihr Resultat durch Zufälligkeiten, wie u. a. die der Wahl anderer Maßstäbe beeinflußt würde. Es ist aber leicht einzusehen, daß dies für den wichtigsten Parameter, nämlich r, nicht der Fall ist. Soll eine Größe k, die mit der Maßeinheit a gemessen wird, durch das Maß b ausgedrückt werden, dann erhalten wir an Stelle von k:

$$k' = \frac{a}{b} k = ck$$

Einen andern Maßstab wählen heißt also nichts anderes, als mit einem konstanten Faktor, nämlich mit dem Maßverhältnis (altes M.: neues M.) multiplizieren. Wird nun mit Hilfe der Korrelationsrechnung die Beziehung zwischen den Zahlenfolgen $x_1, x_2, \ldots x_n$ und $y_1, y_2, \ldots y_n$ untersucht und messen wir sowohl die x-, als auch die y-Werte mit andern Maßen, dann tritt an Stelle der x_i der neue Wert $c_x \cdot x_i$ und an Stelle der y_i die c_y, y_i, wobei die c für die Maßverhältnisse

$$c = \frac{\text{altes Maß}}{\text{neues Maß}}$$

gesetzt werden. Entsprechend werden alle für die Berechnung der Korrelationsparameter in Betracht fallenden Größen mit den Konstanten behaftet. Wir erhalten also:

$$r = \frac{c_x c_y \Sigma \Delta x \Delta y}{\sqrt{c_x^2 \Sigma \Delta x^2 \cdot c_y^2 \Sigma \Delta y^2}} = \frac{\Sigma \Delta x \Delta y}{\sqrt{\Sigma \Delta x^2 \Sigma \Delta y^2}}$$

$$b_{21} = \frac{c_x c_y \Sigma \Delta x \Delta y}{c_x^2 \Sigma \Delta x^2} = \frac{c_y}{c_x} \cdot \frac{\Sigma \Delta x \Delta y}{\Sigma \Delta x^2}$$

$$b_{12} = \frac{c_x c_y \Sigma \Delta x \Delta y}{c_y^2 \Sigma \Delta y^2} = \frac{c_x}{c_y} \cdot \frac{\Sigma \Delta x \Delta y}{\Sigma \Delta y^2}$$

Während sich also bei der Wahl anderer Maßstäbe an der Abhängigkeit zwischen x und y als solcher nichts ändert, erhalten die für die bildliche Darstellung maßgebenden Koeffizienten b neue Werte. Nimmt b_{21} zu, dann nimmt b_{12} im gleichen Verhältnis ab und umgekehrt. Wir haben es also beispielsweise in der Hand, durch bestimmte Wahl der c die Regressionskoeffizienten gleich groß zu machen. Dann ergibt sich:

$$c_x^2 : c_y^2 = \frac{1}{\Sigma \Delta x^2} : \frac{1}{\Sigma \Delta y^2}$$

$$c_x : c_y = \frac{1}{\sigma_x} : \frac{1}{\sigma_y} \text{ und}$$

$$b_{21} = b_{12} = r$$

Dieser Wahl der Reziprokwerte der Streuungen als Maßverhältnisse kommt deshalb eine besondere Bedeutung zu, weil auf diese Weise die Beziehungsgleichung in der von englischen und amerikanischen Statistikern bevorzugten Form („Normalform") erhalten wird (vgl. pag. 34). Setzt man

$$X \text{ für } \frac{x}{\sigma_x}, \qquad Y \text{ für } \frac{y}{\sigma_y},$$

dann lauten die linearen Beziehungsgleichungen:

$$\begin{aligned} Y - \frac{m_y}{\sigma_y} &= r \cdot \left(X - \frac{m_x}{\sigma_y}\right) \\ X - \frac{m_x}{\sigma_x} &= r \cdot \left(Y - \frac{m_y}{\sigma_y}\right) \end{aligned} \qquad 23)$$

Alle Größen in diesen Gleichungen sind unbenannte Zahlen. Die beiden Regressionsgeraden haben symmetrische Lage zur Winkelhalbierenden des durch P (m_x, m_y) gelegten Systems.

Ist allgemein $y = f(x)$, (kor. Bez.)

dann geht bei Einführung von Normalkoordinaten diese Beziehungsgleichung über in

$$Y\sigma_y = f(X\sigma_x)$$

$$\text{oder} \quad Y = \frac{1}{\sigma_y} f(X\sigma_x)$$

Der Parameter p [19]) muß in der Normalform aus Symmetriegründen einen Extremwert (Maximum oder Minimum) erhalten. Durch eine kleine Umformung geht 19) über in

$$p = \frac{1 - r^2}{b + \frac{r^2}{b}},$$

wobei b für b_{21}, oder b_{12} gesetzt werden kann.

Da die Wahl anderer Maße den Zähler nicht beeinflußt, haben wir lediglich festzustellen, für was für ein b die Summe

$$s = b + \frac{r^2}{b}$$

am kleinsten wird. Daß s kein Maximum haben kann, ist ohne weiteres einzusehen. Wir finden:

$$\frac{ds}{db} = 1 - \frac{r^2}{b^2} = o \qquad b = r$$

Die zweite Ableitung wird positiv (an der Stelle $b = r$: $2/r$), wie dies zu erwarten war. b hat gleiches Vorzeichen wie r, so daß nur die positive Wurzel in Betracht fällt. Der Übergang zu Normalkoordinaten bedingt also ein minimales s und damit ein maximales p:

$$p = \frac{1 - r^2}{2r} \text{ in Übereinstimmung mit 20)}$$

Von den mathematischen Eigenschaften der Funktion p (für veränderliches r!) ist bereits früher schon (pag. 23) die Rede gewesen. Ergänzend sei noch nachgetragen, daß sich die Kurve (Fig. 10) asymptotisch der Ordinate und der Geraden

$$p = -\frac{r}{2}$$

nähert. Wählen wir die letztere als Abszisse, dann wird

$$p' r' = \frac{1}{4}\sqrt{5}$$

Zusammenfassend läßt sich sagen, daß

1. das Abhängigkeitsmaß als solches unbeeinflußt bleibt von der Wahl der Maßstäbe, in denen die zueinander in Beziehung gesetzten Größen gemessen werden;
2. die für die bildliche Darstellung maßgebenden Regressionskoeffizienten abhängig sind von den Maßen und daß deren Wahl so getroffen werden kann, daß die b-Werte gleich dem Abhängigkeitsmaß r werden;
3. der Winkel zwischen den Regressionsgeraden sich nicht nur mit der Abhängigkeit, sondern auch mit den Maßen der einander bedingenden Größen ändert und daß er einen für den Grad der Abhängigkeit kennzeichnenden Maximalwert bei der Messung in Normalmaß annimmt.

Die partielle Korrelationsrechnung

Bereits früher wurde schon erwähnt, daß der korrelative Charakter einer Beziehung zwischen zwei Größen auf den möglichen Einfluß noch weiterer zurückgeführt werden könnte. Neben das wichtige Problem der „Streckung" nichtlinearer Beziehungen tritt also das andere, die Bindung auch von mehr als zwei Größen mit Hilfe der Korrelationsrechnung zu untersuchen. Dieser Aufgabe widmet sich die als partielle bezeichnete: Sie bestimmt den Sondereinfluß jeder „bedingenden" Größe auf die „bedingte" und drückt das Ergebnis der Untersuchung in partiellen Korrelations- und Regressionskoeffizienten aus. Der Ableitung dieser Koeffizienten sei der folgende Abschnitt gewidmet.

Eine Reihe von Größen $x_1, \ldots x_n$, sei linear in der Weise verbunden, daß sich setzen läßt:

$$b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + a = 0$$

Sind die Werte x_i k-mal empirisch bestimmt worden, und verlangt man, daß die Koeffizienten b_i so gewählt werden, daß sie, in den k Bestimmungsgleichungen eingesetzt, für a einen kleinsten Wert ergeben sollen, dann bestimmen wir die b_i folgendermaßen: Wir summieren alle Gleichungen:

$$\begin{array}{l}
b_1 x_{11} + b_2 x_{12} + b_3 x_{13} + \ldots + b_n x_{1n} + a_1 = 0 \\
b_1 x_{21} + b_2 x_{22} + b_3 x_{23} + \ldots + b_n x_{2n} + a_2 = 0 \\
\ldots\ldots \\
b_1 x_{k1} + b_2 x_{k2} + b_3 x_{k3} + \ldots + b_n x_{kn} + a_n = 0 \\
\hline
b_1 \Sigma x_{i1} + b_2 \Sigma x_{i2} + \ldots \quad + b_n \Sigma x_{in} + \Sigma a_i = 0
\end{array}$$

Durch Bildung der Differenzen aus x_{in} und dem Mittel aller x_{ik} ergeben sich k weitere Gleichungen von der Form:

$$b_1 (x_{i_1} - \frac{x_{k_1}}{k}) + b_2 (x_{12} - \frac{x_{k_2}}{k}) + \ldots$$

$$+ b_n (x_{in} - \frac{x_{kn}}{k}) + (a_i - \frac{a_k}{k}) = 0$$

oder kürzer:

$$b_1 \Delta x_{i_1} + b_2 \Delta x_{i_2} + \ldots + b_n \Delta x_{in} + \Delta a_i = 0$$

Summiert man über alle i von 1 bis k, dann muß wegen

$$\Sigma \Delta x_{i_1} = 0, \quad \ldots \Sigma \Delta x_{in} = 0 \text{ auch } \Sigma \Delta a_i = 0 \text{ sein.}$$

Setzen wir also:

$$b_1 \Delta x_{11} + b_2 \Delta x_{12} + b_3 \Delta x_{13} + \ldots + b_n \Delta x_{1n} = v_1$$
$$b_1 \Delta x_{21} + b_2 \Delta x_{22} + b_3 \Delta x_{23} + \ldots + b_n \Delta x_{2n} = v_2$$
$$\ldots\ldots$$
$$b_1 \Delta x_{k_1} + b_2 \Delta x_{k_2} + b_3 \Delta x_{k_3} + \ldots + b_n \Delta x_{kn} = v_n,$$

dann können wir nach der Methode der kleinsten Quadrate die Koeffizienten b_i so bestimmen, daß Σv_i^2 zu einem Minimum wird. Durch Quadrierung und Summierung über alle k ergibt sich:

$$\sum_1^k \sum_1^n b_n^2 \Delta x_{kn}^2 + 2 \sum_1^k \sum_1^n b_n b_m \Delta x_{kn} \Delta x_{km} = \sum v_i^2 = \text{Min.}$$

Differenzieren wir partiell nach allen b_i, dann erhalten wir ein System von n in den Koeffizienten lineare Gleichungen (für Δx ist unten x gesetzt; als Summenzeichen wird [ ] verwendet):

$$b_1 [x_1 x_1] + b_2 [x_1 x_2] + b_3 [x_1 x_3] + \ldots + b_n [x_1 x_n] = 0$$
$$b_1 [x_1 x_2] + b_2 [x_2 x_2[+ b_3 [x_2 x_3] + \ldots + b_n [x_2 x_n] = 0$$
$$b_1 [x_1 x_3] + b_2 [x_2 x_3] + b_3 [x_3 x_3] + \ldots + b_n [x_3 x_n] = 0$$
$$\ldots\ldots$$
$$b_1 [x_1 x_n] + b_2 [x_2 x_n] + b_3 [x_3 x_n] + \ldots + b_n [x_n x_n] = 0$$

Dieses homogene Gleichungssystem gestattet selbstverständlich nicht, die b_i zu berechnen, sondern je nach der Wahl eines der x_i-Werte (hier nicht als Differenzwert aufzufassen!) als abhängige Veränderliche, zu

ihnen proportionale Größen. Dividiert man zum Beispiel mit b_1, dann drücken wir x_1 (urspr. Bedeutg.) durch alle übrigen x_i aus, usw. Wir schreiben inskünftig

$$\text{für } \frac{b_i}{b_1} \text{ nach Yule: } b_{1i,23\ldots n},$$

wobei die hinter dem Komma folgenden Indizes auf alle x_i Bezug haben, denen $b_{1i,\ldots}$ nicht als Koeffizient zugehört. Aus einem berechneten Koeffizienten ergeben sich durch zyklische Vertauschung der Indizes alle übrigen.

a) Drei verbundene Größen.

$$b_1 [x_1 x_1] + b_2 [x_1 x_2] + b_3 [x_1 x_3] = 0$$

$$b_1 [x_1 x_2] + b_2 [x_2 x_2] + b_3 [x_2 x_3] = 0$$

$$b_1 [x_1 x_3] + b_2 [x_2 x_3] + b_3 [x_3 x_3] = 0$$

Wir dividieren die erste Gleichung mit $[x_1 x_1]$, die zweite mit $[x_2 x_2]$, die dritte mit $[x_3 x_3]$; ferner alle, um die Abhängigkeit der ersten Größe von den übrigen zu bestimmen, mit b_1. Dann erhalten wir:

$$1 + b_{12,3} b_{21} + b_{13,2} b_{31} = 0$$

$$b_{12} + b_{12,3} + b_{13,2} b_{32} = 0$$

$$b_{13} + b_{12,3} b_{23} + b_{13,2} = 0$$

Die Regressionskoeffizienten bestimmen wir aus der zweiten und dritten Gleichung.

$$b_{12,3} = -\frac{b_{12} - b_{13} b_{32}}{1 - r_{23}^2}$$

$$b_{13,2} = -\frac{b_{13} - b_{12} b_{23}}{1 - r_{23}^2}$$

Durch eine kleine Umformung gehen diese Ausdrücke über in:

$$b_{12,3} = -\frac{r_{12}}{b_{12}} \cdot \frac{r_{12} - r_{13} r_{23}}{1 - b_{23}^2}$$

$$b_{13,2} = -\frac{r_{13}}{b_{31}} \cdot \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2}$$

$b_{12,3}$ gibt an, um welchen Betrag sich die erste der verbundenen Größen ändert, wenn, bei Konstanthaltung der dritten, die Änderung der

zweiten Größe als 1 angenommen wird. r_{12} mißt die Abhängigkeit zwischen x_1 (urspr. Bed.) und x_2, r_{13} diejenige zwischen x_1 und x_3, usw.

Die übrigen b-Koeffizienten lauten:

$$b_{21,3} = -\frac{b_{21} - b_{23}\,b_{31}}{1 - {}^{2}_{13}} = -\frac{r_{12}}{b_{12}} \cdot \frac{r_{12} - r_{23}\,r_{13}}{1 - r_{13}^2}$$

$$b_{23,1} = -\frac{b_{23} - b_{21}\,b_{13}}{1 - r_{13}^2} = -\frac{r_{23}}{b_{32}} \cdot \frac{r_{23} - r_{12}\,r_{13}}{1 - r_{13}^2}$$

$$b_{31,2} = -\frac{b_{31} - b_{32}\,b_{21}}{1 - r_{12}^2} = -\frac{r_{13}}{b_{13}} \cdot \frac{r_{13} - r_{23}\,r_{12}}{1 - r_{12}^2}$$

$$b_{32,1} = -\frac{b_{32} - b_{31}\,b_{12}}{1 - r_{12}^2} = -\frac{r_{23}}{b_{23}} \cdot \frac{r_{23} - r_{13}\,r_{12}}{1 - r_{12}^2}$$

Man erkennt ohne weiteres, daß die ursprünglichen Koeffizienten b_1, b_2, b_3 des homogenen Gleichungssystems wieder gefunden werden können: $b_{12,3}$ und $b_{13,2}$ enthalten beide den Ausdruck $1 - r_{23}^2$ im Nenner; ebenso $b_{21,3}$ und $b_{23,1}$ den entsprechenden $1 - r_{13}^2$, usw. Wir setzen also:

$$b_1 = 1 - r_{23}^2;\; b_2 = 1 - r_{13}^2;\; b_3 = 1 - r_{12}^2$$

und betrachten diese Ausdrücke als zu $[x_1\,x_1]$, $[x_2\,x_2]$, $[x_3\,x_3]$ zugehörig. Daraus ergibt sich folgendes Schema der ursprünglichen Koeffizienten:

$$\begin{array}{cccc}
 & b_1 & b_2 & b_3 \\
+ & (1 - r_{23}^2) & -\dfrac{r_{12}}{b_{21}}(r_{12} - r_{13}\,r_{23}) & -\dfrac{r_{13}}{b_{31}}(r_{13} - r_{12}\,r_{23}) \\
- & \dfrac{r_{12}}{b_{12}}(r_{12} - r_{13}\,r_{23}) & +\;(1 - r_{13}^2) & -\dfrac{r_{23}}{b_{32}}(r_{23} - r_{12}\,r_{13}) \\
- & \dfrac{r_{13}}{b_{13}}(r_{13} - r_{12}\,r_{23}) & -\dfrac{r_{23}}{b_{23}}(r_{23} - r_{13}\,r_{12}) & +\;(1 - r_{12}^2)
\end{array}$$

Es zeigt sich, daß die Bedeutung der Koeffizienten b_i ganz davon abhängig ist, in welcher der 3 Gleichungen sie stehen. Bezeichnen wir die letzteren der Reihe nach mit I bis III, dann müßte

$$b_{1,\mathrm{I}} = b_{1,\mathrm{II}} = b_{1,\mathrm{III}} \text{ sein, usw.}$$

Für ein aus Pütter, „Die Auswertung zahlenmäßiger Beobachtungen in der Biologie", pag. 48 u. f., entnommenes Beispiel ergeben sich aber folgende Werte:

	b_1	b_2	b_3	
I	0,721	0,492	4,02	
II	1,444	1,0	8,19	(absolute Werte)
III	0,0485	0,0336	0,285	

Diese Tatsache erklärt sich aus dem korrelativen Charakter der Beziehung. Immerhin bleibt die Proportionalität der Koeffizienten ($b_1 : b_2 : b_3$) einigermaßen gewahrt.

Die Korrelationskoeffizienten berechnen wir als geometrische Mittel aus den zugehörigen b-Werten.

$$r_{12,3} = \sqrt{b_{12,3}\, b_{21,3}} = \sqrt{\frac{b_{2,I}\, b_{1,II}}{b_{1,I}\, b_{2,II}}}$$

$$= \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)\,(1 - r_{23}^2)}}$$

$$r_{13,2} = \frac{r_{13} - r_{12}\, r_{23}}{\sqrt{1 - r_{12}^2)\,(1 - r_{23}^2)}} \qquad 24)$$

$$r_{23,1} = \frac{r_{23} - r_{12}\, r_{13}}{\sqrt{(1 - r_{12}^2)\,(1 - r_{13}^2)}}$$

Die Regressionskoeffizienten lassen sich bei Einführung neuer Streuungsmaße auf eine 13) ähnliche Form bringen. Setzen wir:

$$\sigma_{12} = \sigma_1 \sqrt{1 - r_{12}^2},\ \sigma_{13} = \sigma_1 \sqrt{1 - r_{13}^2}, \qquad 25)$$

usw.

dann ergibt sich mit Hilfe dieser, als bedingte bezeichneter Streuungen:

$$b_{12,3} = r_{12,3} \cdot \frac{\sigma_{13}}{\sigma_{23}} \qquad b_{13,2} = r_{13,2} \cdot \frac{\sigma_{12}}{\sigma_{32}}$$

$$b_{21,3} = r_{12,3} \cdot \frac{\sigma_{23}}{\sigma_{13}} \qquad b_{23,1} = r_{23,1} \cdot \frac{\sigma_{21}}{\sigma_{31}} \qquad 26)$$

usw.

Es mag an dieser Stelle erwähnt sein, daß die auf Yule zurückgehende Theorie (vgl. M. Exner: „Über die Korrelationsmethode") die Regressionskoeffizienten mit Streuungsmaßen höherer Ordnung bestimmt. So wird beispielsweise

$$b_{12,3} = r_{12,3} \cdot \frac{\sigma_{1,23}}{\sigma_{2,13}} \text{ gesetzt, wobei}$$

$$\sigma_{1,23} = \sigma_1 \sqrt{1 - r_{12}^2} \cdot \sqrt{1 - r_{13,2}^2}$$

$$\sigma_{2,13} = \sigma_2 \sqrt{1 - r_{12}^2} \cdot \sqrt{1 - r_{23,1}^2}$$

(vgl. Korrelationsrechnung mit 4 verbundenen Größen).

Es läßt sich aber leicht zeigen, daß der für $b_{12,3}$ angegebene Ausdruck sich in den einfacheren von 26) umformen läßt. Man findet:

$$\frac{\sigma_{1,23}}{\sigma_{2,13}} = \frac{\sigma_1 \sqrt{1 - r_{13,2}^2}}{\sigma_2 \sqrt{1 - r_{23,1}^2}}, \text{ und}$$

$$\frac{1 - r_{13,2}^2}{1 - r_{23,1}^2} = \frac{(1-r_{12}^2)(1-r_{13}^2)(1-r_{12}^2-r_{13}^2-r_{23}^2 \pm r_{12}^2 r_{23}^2 + 2r_{12} r_{13} r_{23})}{(1-r_{12}^2)(1-r_{23})(1-r_{12}^2-r_{13}^2-r_{23}^2 \pm r_{12}^2 r_{23}^2 + 2r_{12} r_{13} r_{23})}$$

$$= \frac{1 - r_{13}^2}{1 - r_{23}^2}$$

Die Beziehungsgleichungen können in eine 23) nachgebildete „Normalform" übergeführt werden, wenn man sie durch die bilateralen Streuungsmaße, die in den b-Werten im Zähler auftreten, dividiert. Wir erhalten auf diese Weise:

$$\begin{aligned}
\frac{x_1 - m_{x1}}{\sigma_{12} \cdot \sigma_{13}} &= r_{12,3} \frac{x_2 - m_{x2}}{\sigma_{12} \cdot \sigma_{23}} + r_{13,2} \frac{x_3 - m_{x3}}{\sigma_{13} \cdot \sigma_{32}} \\
\frac{x_2 - m_{x2}}{\sigma_{21} \cdot \sigma_{23}} &= r_{12,3} \frac{x_1 - m_{x1}}{\sigma_{21} \cdot \sigma_{13}} + r_{23,1} \frac{x_3 - m_{x3}}{\sigma_{23} \cdot \sigma_{31}} \qquad 27) \\
\frac{x_3 - m_{x3}}{\sigma_{31} \cdot \sigma_{32}} &= r_{13,2} \frac{x_1 - m_{x1}}{\sigma_{31} \cdot \sigma_{12}} + r_{23,1} \frac{x_2 - m_{x2}}{\sigma_{32} \cdot \sigma_{21}}
\end{aligned}$$

Aus diesem Ergebnis ist der Schluß zu ziehen, daß das Fehlen von Benennungen in der Normalform 23) nur dem Umstand zu verdanken ist, daß bei n zueinander in Beziehung gesetzten Größen die $x_i - m_{xi}$

durch n — 1 σ-Werte zu dividieren sind; bei zwei Größen also durch einen, bei dreien durch zwei, bei vieren durch drei, usw.

Das Vorzeichen der Koeffizienten 26) entspricht dem der Korrelationskoeffizienten selbst bei expliziter Darstellung der Gleichungen.

Es bleibt nun noch übrig, die Grenzen der partiellen Korrelationskoeffizienten zu bestimmen. Wohl ist anzunehmen, daß sie, wie die zweiseitigen Koeffizienten, Werte zwischen — 1 und + 1 mit Einschluß dieser Grenzen annehmen können; die in 24) gewonnenen Ausdrücke schließen aber die Möglichkeit größerer und kleinerer Werte nicht aus. Sie lassen lediglich erkennen, daß

1. positive, wie negative Werte möglich sind;
2. irgend ein $r_{ab,c} = 0$ wird für $r_{ab} = r_{ac}\, r_{bc}$;
3. alle partiellen Koeffizienten r zu Null werden für $r_{12} = r_{13} = r_{23} = 0$ (völlige Beziehungslosigkeit);
4. keiner dieser Koeffizienten zu Null wird, wenn nur einer der zweiseitigen (r_{ab}) gleich Null ist, wobei die Möglichkeit größerer Werte, als $|1|$ dahingestellt bleibt. Kleinere würden erhalten, wenn die Summe der Quadrate der von 0 verschiedenen r.. selbst kleiner als 1 wäre. Die Erfüllung dieser Bedingung ist zum mindesten wahrscheinlich, weil das Fehlen einer Beziehung zwischen zwei Größen kaum vereinbar ist mit der Existenz enger Zusammenhänge beider zu einer dritten Größe;
5. für $r_{ab} = 0$, $r_{ac} = 0$ auch $r_{ab,c}$ und $r_{ac,b}$ zu Null werden, während sich für $r_{bc,a}$ ein mit r_{bc} übereinstimmender Wert ergibt. —

Bezeichnen wir mit f den Ausdruck

$$\begin{aligned} & (1 - r_{23}^2)\,(1 - r_{13}^2)\,(1 - r_{12,3}^2) \\ = & (1 - r_{13}^2)\,(1 - r_{12}^2)\,(1 - r_{23,1}^2) \\ = & (1 - r_{12}^2)\,(1 - r_{23}^2)\,(1 - r_{13,2}^2) \\ = & \; 1 + 2r_{12}\, r_{13}\, r_{23} - r_{12}^2 - r_{13}^2 - r_{23}^2, \end{aligned} \qquad 28a)$$

dann ist leicht einzusehen, daß für

$f > 0$

$|r_{ab,c}| < 1$ wird.

Man kann also an Stelle der partiellen Koeffizienten r auch f untersuchen. Hat dieses Extremwerte, dann müssen seine partiellen Ableitungen erster Ordnung verschwinden. Nun ergibt sich etwa aus

$$\frac{\delta f}{\delta r_{12}} = 2r_{13}\, r_{23} - 2r_{12} = 0$$

$$r_{12} = r_{13}\, r_{23}$$

r_{12} ist also durch die beiden andern zweiseitigen Koeffizienten bestimmt, und zwar liegt in diesem Falle ein durch r_{12} bedingtes Maximum von f (f_m) vor:

$$f_m = 1 + r_{13}^2\, r_{23}^2 - r_{13}^2 - r_{23}^2$$
$$= (1 - r_{13}^2)\,(1 - r_{23}^2)$$

f_m kann also durch einen positiven, echten Bruch dargestellt werden. Bilden wir die Differenz $f_m - f$, dann finden wir

$$f_m - f = (r_{13}\, r_{23} - r_{12})^2 > 0$$

f kann also ebenfalls nur Werte zwischen 0 und $+1$ annehmen, woraus weiter sich ergibt, daß die partiellen Kkf. $r_{ab,c}$ nicht größer als $+1$ und nicht kleiner als -1 werden können.[1]

---

[1] Extremwerte von f.

Aus der Nullsetzung der partiellen Ableitungen erster Ordnung ergibt sich vorerst:

$$r_{12} = r_{13}\, r_{23}$$
$$r_{13} = r_{12}\, r_{23}$$
$$r_{23} = r_{12}\, r_{13}$$

Im Falle eines Extremwertes sind also entweder keine, oder zwei der Koeffizienten negativ. Multipliziert man mit r_{12} bzw. r_{13}, bzw. r_{23}, dann erhält man ferner:

$$r_{12}^2 = r_{13}^2 = r_{23}^2$$

Der in den Koeffizienten symmetrische Ausdruck f hat also nur einen Extremwert für

$$|r_{12}| = |r_{13}| = |r_{23}|.$$

Ersetzen wir daher die r_{ik} durch x, dann geht f über in

$$f = 1 \pm 2x^3 - 3x^2,$$

also in eine Funktion dritten Grades, die zwei Stellen haben muß, an denen f_x Null wird:

$$\frac{1}{6} \cdot \frac{df}{dx} = \pm x^2 - x = 0 \quad x = 0, \text{ bezw. } \pm 1$$

$$\frac{d^2 f}{dx^2} = 2x - 1$$

f hat also ein Maximum für $x = 0$ ($f_m = 1$) und ein Minimum für $x = 1$ ($f_m' = o$). Da $|x|$ nur Werte zwischen 0 und 1 haben kann, bleibt auch der Variabilitätsbereich für f innerhalb dieser Grenzen und daraus folgt anderseits

$$o \leqq |r_{ab,c}| \leqq 1$$

Aus $+1 \geqq f \geqq 0$ folgt ferner

$$1 + 2r_{12}\, r_{13}\, r_{23} \geqq r_{12}^2 + r_{13}^2 + r_{23}^2$$

Aus dieser Ungleichheit ist ohne weiteres zu erkennen, daß bei vorgeschriebenen r_{13} und r_{23} der dritte Koeffizient r_{12} nicht beliebig groß oder klein sein darf. Die Auflösung nach r_{12} (entsprechende Ausdrücke für r_{13} und r_{23} erhält man durch zyklische Vertauschung) ergibt:

$$r_{13}\, r_{23} - \sqrt{(1 - r_{13}^2)\,(1 - r_{23}^2)}$$

$$\leqq r_{12} \leqq r_{13}\, r_{23} + \sqrt{(1 - r_{13}^2)\,(1 - r_{23}^2)} \qquad 28)$$

r_{12} variiert also innerhalb eines Bereiches, dessen gesamte Breite durch

$$2b = 2 \left| \sqrt{(1 - r_{13}^2)\,(1 - r_{23}^2)} \right|$$

bestimmt ist, während füglich $r_{13}\, r_{23}$ als „Mittelwert" von r_{12} bezeichnet werden darf. Setzen wir

$$r_{12} = \cos \alpha$$

$$r_{13} = \cos \beta$$

$$r_{23} = \cos \gamma,$$

dann geht 28) über in

$$\cos(\beta + \gamma) \leqq \cos \alpha \leqq \cos(\beta - \gamma)$$

Den geometrischen Sinn dieser Beziehung kann man sich klar machen, wenn man durch die Spitze eines gleichschenkligen Dreiecks mit der Schenkellänge $s = 1$ und dem Basiswinkel β eine Gerade g so legt, daß sie die Basis unter dem Winkel γ schneidet. $\cos(\beta + \gamma)$ und $\cos(\beta - \gamma)$ ergeben sich dann als Projektionen der Schenkel auf der Geraden (Fig. 16).

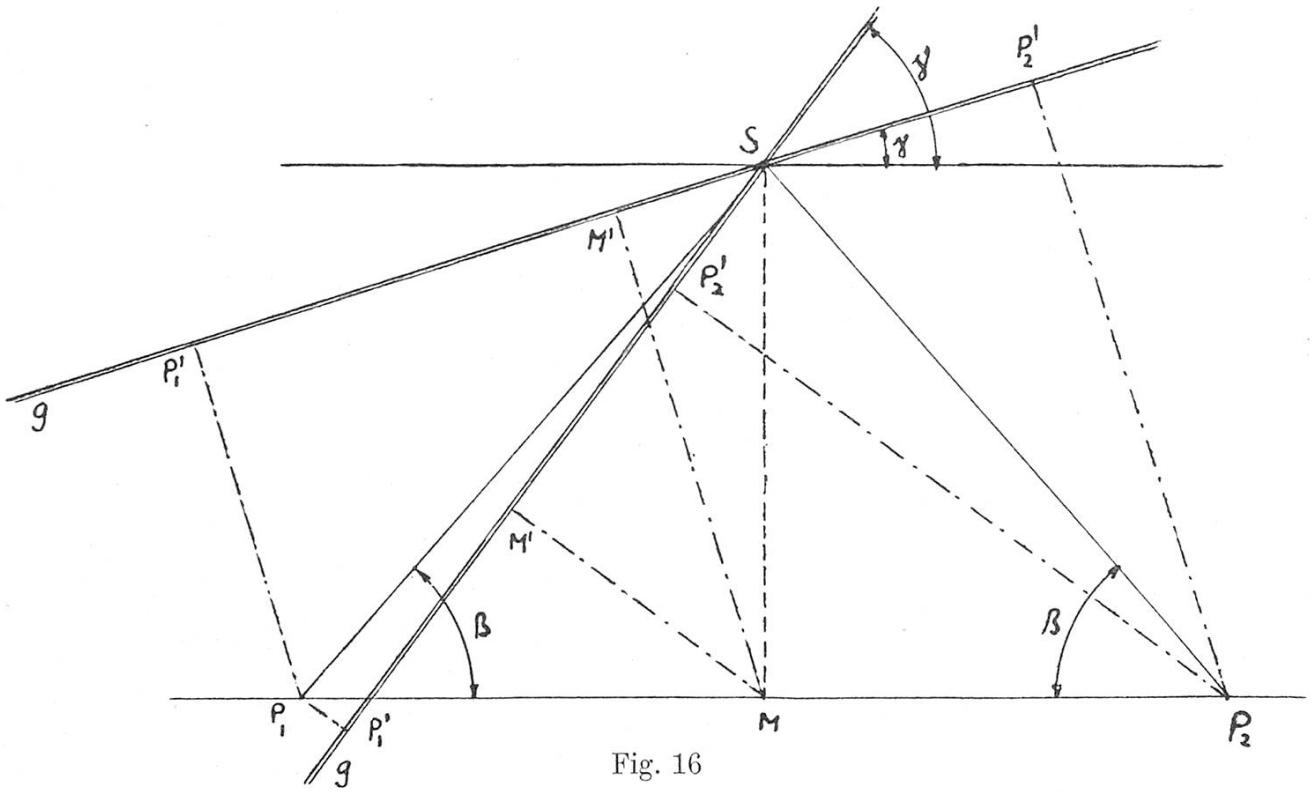

Fig. 16

$M' S = b = \sin\beta \sin\gamma$

$M' P'_1 = M' P'_2 = \cos\beta \cos\gamma$

$SP'_1 = \cos(\beta - \gamma)$

$SP'_2 = \cos(\beta + \gamma)$

Für $\gamma = 0$ ($r_{23} = 1$) wird

$\cos\alpha = \cos\beta$; $r_{12} = r_{13}$.

Besteht zwischen zweien von drei zueinander in Beziehung gesetzten Größen ein linear-funktioneller Zusammenhang, dann ist der korrelative beider zur dritten Größe durch denselben Koeffizienten r gekennzeichnet.

Für $\gamma = \beta$, d. h. bei Koeffizientengleichheit kann $r_{12} = 1$ sein.

Werden die Beziehungen zwischen einer Größe und zwei andern durch gleiche Korrelationskoeffizienten ausgedrückt, dann stehen diese andern Größen möglicherweise in funktionellem Zusammenhang.

Als untere Grenze für r_{12} findet man $2r_{13}^2 - 1$. Dieser Wert wird gleich Null, d. h. es besteht keine Beziehung zwischen x_1 und x_2, wenn

$$r_{13} = r_{23} = \pm 0{,}707$$

Bei einer verhältnismäßig engen Bindung zwischen x_1 und x_3 einerseits, x_2 und x_3 anderseits umfaßt der mögliche Variabilitätsbereich für r_{12} die volle Breite von Null bis Eins.

Für $\gamma = 90$ ($r_{23} = 0$) erhalten wir

$$-\sqrt{1 - r_{13}^2} \leqq r_{12} \leqq +\sqrt{1 - r_{13}^2}$$

$$r_{12}^2 + r_{13}^2 \leqq 1$$

Aus diesem letzten Ergebnis lassen sich folgende Vorzeichenregeln ableiten:

Ist die Summe der Quadrate zweier Koeffizienten r_{ik} größer als Eins, dann ist der dritte positiv oder negativ, je nachdem die beiden ersten gleiches, oder ungleiches Vorzeichen haben. Ist die Summe kleiner als Eins, dann können die negativen Vorzeichen in beliebiger Anzahl auftreten.

b) Vier verbundene Größen.

In gleicher Weise, wie bei drei verbundenen Größen, gewinnt man aus dem Gleichungssystem:

$$\begin{aligned}
&b_1 [x_1 x_1] + b_2 [x_1 x_2] + b_3 [x_1 x_3] + b_4 [x_1 x_4] = 0 \\
&b_1 [x_1 x_2] + b_2 [x_2 x_2] + \ldots \qquad\qquad\qquad\qquad = 0 \\
&b_1 [x_1 x_3] + \ldots \\
&\text{usw.}
\end{aligned}$$

das durch Division mit $[x_1 x_1]$, bzw. $[x_2 x_2]$, usw. und — um die Abhängigkeit der ersten Größe von den übrigen Größen auszudrücken — mit b_1 übergeht in:

$$\begin{aligned}
&1 + b_{12,..} b_{21} + b_{13,..} b_{31} + b_{14,..} b_{41} = 0 \\
&b_{12} + b_{12,..} \qquad + b_{13,..} b_{32} + b_{14,..} b_{42} = 0 \\
&b_{13} + b_{12,..} b_{23} + b_{13,..} \qquad + b_{14,..} b_{43} = 0 \\
&b_{14} + b_{12,..} b_{24} + b_{13,..} b_{34} + b_{14,..} \qquad = 0
\end{aligned}$$

die Regressionskoeffizienten

$$b_{12,34} = \frac{b_{12,4} - b_{13,4} b_{32,4}}{1 - r_{23,4}^2} \; (-)$$

$$b_{12,43} = \frac{b_{12,3} - b_{14,3} b_{42,3}}{1 - r_{24,3}^2} \; (-)$$

$$b_{13,24} = \frac{b_{13,4} - b_{12,4} b_{23,4}}{1 - r_{23,4}^2} \; (-)$$

$$b_{21,34} = \frac{b_{21,4} - b_{23,4} b_{31,4}}{1 - r_{13,4}^2} \; (-)$$

usw.

Aus ihnen erhalten wir unter Berücksichtigung der Möglichkeit, die Reihenfolge der hinter dem Komma aufgeführten Indizes zu ändern, also beispielsweise neben einem $b_{12,34}$ ein $b_{12,43}$ zu unterscheiden, insgesamt 12 Korrelationskoeffizienten. Sie ergeben sich wieder als geometrische Mittel zusammengehörender b-Werte. Da sich wegen

$$b_{21,3}\, b_{14,3}\, b_{42,3} = b_{12,3}\, b_{41,3}\, b_{24,3} = r_{12,3}\, r_{14,3}\, r_{24,3}$$

beispielsweise für $b_{12,34}$

$$\frac{r_{12,4}}{b_{21,4}} \cdot \frac{r_{12,4} - r_{13,4}\, r_{23,4}}{1 - r^2_{23,4}}$$

und für $b_{21,34}$

$$\frac{r_{12,4}}{b_{12,4}} \cdot \frac{r_{12,4} - r_{13,4}\, r_{23,4}}{1 - r^2_{13,4}}$$

setzen läßt, erhalten wir u. a.

$$\begin{aligned}
r_{12,34} &= \frac{r_{12,4} - r_{13,4}\, r_{23,4}}{\sqrt{(1 - r^2_{13,4})\,(1 - r^2_{23,4})}} \\
r_{12,43} &= \frac{r_{12,3} - r_{14,3}\, r_{24,3}}{\sqrt{(1 - r^2_{14,3})\,(1 - r^2_{24,3})}} \\
r_{13,24} &= \frac{r_{13,4} - r_{12,4}\, r_{23,4}}{\sqrt{(1 - r^2_{12,4})\,(1 - r^2_{23,4})}} \\
r_{14,23} &= \frac{r_{14,3} - r_{12,3}\, r_{24,3}}{\sqrt{(1 - r^2_{12,3})\,(1 - r^2_{24,3})}}
\end{aligned} \qquad 30)$$

usw.

Die vor dem Komma stehenden Indizes bezeichnen die Größen, zwischen denen die Abhängigkeit zu ermitteln ist. Hinter dem Komma sind die Konstanten angeführt. Wie dies bereits bei den Regressionskoeffizienten festgestellt werden konnte, wird zwischen der dem Komma unmittelbar folgenden und der nach dieser stehenden Größe ein Unterschied gemacht. Denkt man sich z. B. in $r_{12,34}$ den Index 4 weg, dann geht der vierseitige Koeffizient in den entsprechenden dreiseitigen $r_{12,3}$ über. Ähnlich wird aus $r_{12,43}$ $r_{12,4}$. Nimmt man also an, daß z. B. in $r_{12,3}$ der Einfluß der Größe x_3 ausgeschaltet wird, dann ist dies auch in $r_{12,34}$ der Fall, während im Gegensatz dazu die Größe x_4 schon in jenen dreiseitigen Koeffizienten eliminiert wird, auf die sich der vierseitige zurückführen läßt. Die Korrelationskoeffizienten

vierter Ordnung unterscheiden also zwischen schon ausgeschalteten und noch auszuschaltenden Größen und wir haben darum zu untersuchen, ob dies vielleicht nur in formeller Weise geschieht.

Setzen wir vorerst willkürlich z. B. $r_{12,34} = r_{12,43}$, dann ergibt sich nach 30) hieraus:

$$\frac{\sqrt{(1 - r^2_{14,3})(1 - r^2_{24,3})}}{\sqrt{(1 - r^2_{13,4})(1 - r^2_{23,4})}} = \frac{(r_{12,3} - r_{14,3}\, r_{24,3})}{(r_{12,4} - r_{13,4}\, r_{23,4})}$$

Nach 28a) ist nun

$$1 - r^2_{ab,c} = \frac{f_{abc}}{(1 - r^2_{ac})(1 - r^2_{bc})}$$

Mit Hilfe dieses Ausdrucks geht die linke Seite der Gleichung über in

$$\frac{\sqrt{(1 - r^2_{14})(1 - r^2_{24})}}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}} \; (= k)$$

Die rechte Seite wird nach einigen Umformungen:

$$k \cdot \frac{r_{12} - r_{13}\, r_{23} - r_{12}\, r^2_{34} - r_{14}\, r_{24} + r_{13}\, r_{34}\, r_{24} + r_{14}\, r_{23}\, r_{34}}{r_{12} - r_{13}\, r_{23} - r_{12}\, r^2_{34} - r_{14}\, r_{24} + r_{13}\, r_{34}\, r_{24} + r_{14}\, r_{23}\, r_{34}} = k$$

Da die Gleichung identisch erfüllt ist, machen wir die nicht unwichtige Feststellung, daß die Änderung der Reihenfolge der Konstanten ohne Einfluß auf die Größe eines Korrelationskoeffizienten bleibt. Es ist also

$$r_{12,34} = r_{12,43}$$

$$r_{13,24} = r_{13,42}$$

usw.

Jeder vierseitige r-Wert läßt sich also auf zwei Arten berechnen. Dasselbe trifft nun auch für die Regressionskoeffizienten zu.

Es ist

$$b_{12,34} = \frac{r_{12,4}}{b_{21,4}} \cdot \frac{r_{12,4} - r_{13,4}\, r_{23,4}}{1 - r^2_{23,4}}$$

$$= \frac{\sigma_{14}\, \sqrt{1 - r^2_{13,4}}}{\sigma_{24}\, \sqrt{1 - r^2_{23,4}}} \cdot r_{12,34}$$

$$= \frac{\sigma_{1,34}}{\sigma_{2,34}} \cdot r_{12,34}$$

wenn man als neues Maß der bedingten Streuung (vgl. pag. 77)

$$\sigma_{a,bc} = \sigma_a \cdot \sqrt{1 - r_{ac}^2} \cdot \sqrt{1 - r_{ab,c}^2} \qquad 31)$$

einführt. Nun gilt weiter

$$\sigma_{a,bc} = \sigma_a \cdot \frac{\sqrt{1 - r_{ac}^2} \cdot \sqrt{f_{abc}}}{\sqrt{1 - r_{ac}^2} \cdot \sqrt{1 - r_{bc}^2}} = \sigma_a \cdot \frac{\sqrt{f_{abc}}}{\sqrt{1 - r_{bc}^2}}$$

und

$$\sigma_{a,cb} = \sigma_a \cdot \sqrt{1 - r_{ab}^2} \cdot \sqrt{1 - r_{ac,b}^2}$$

$$= \sigma_a \cdot \frac{\sqrt{1 - r_{ab}^2} \cdot \sqrt{f_{abc}}}{\sqrt{1 - r_{ab}^2} \cdot \sqrt{1 - r_{cb}^2}} = \sigma_a \cdot \frac{\sqrt{f_{abc}}}{\sqrt{1 - r_{bc}^2}}$$

somit $\sigma_{a,bc} = \sigma_{a,cb}$;

wie bei den Korrelationskoeffizienten, so können also auch bei den Streuungsmaßen die hinter dem Komma stehenden Indizes (Konstante) vertauscht werden. Daraus folgt weiter:

$$b_{12,34} = \frac{\sigma_{1,34}}{\sigma_{2,34}} \cdot r_{12,34} = \frac{\sigma_{1,43}}{\sigma_{2,43}} \cdot r_{12,43} = b_{12,43}$$

und allgemein

$$b_{ab,cd} = b_{ab,dc} = \frac{\sigma_{a,cd}}{\sigma_{b,cd}} \cdot r_{ab,cd}$$

$$= \frac{\sigma_{a,dc}}{\sigma_{b,dc}} \cdot r_{ab,dc} \qquad 32)$$

Zusammenfassend läßt sich also sagen, daß bei den vierseitigen Koeffizienten r und b, die sich übrigens der Form nach nicht wesentlich von den entsprechenden dreiseitigen unterscheiden, als neue Eigenschaft die Vertauschbarkeit der hinter dem Komma aufgeführten Indizes, welche die Konstanten bezeichnen, festgestellt werden kann. Daraus ergibt sich die Möglichkeit, jeden vierseitigen Koeffizienten auf zwei Arten zu berechnen. Ebenso gilt dies von den neu eingeführten Streuungsmaßen $\sigma_{a,bc}$, die aus einem Korrelationskoeffizienten die zugehörigen b-Werte zu berechnen gestatten. Sie sind, wie wir neuerdings feststellen, niedrigerer Ordnung, als sie von der Theorie angegeben werden (vgl. pag. 77).

Über die Vorzeichen der Koeffizienten läßt sich folgendes aussagen:

Die Streuungsmaße haben positives und negatives Vorzeichen.

Die Korrelationskoeffizienten sind positiv oder negativ, je nachdem die im Zähler auftretende Differenz

$$r_{ab,d} - r_{ac,d}\, r_{bc,d}$$

positiv oder negativ ist.

Die Regressionskoeffizienten b haben gleiches oder ungleiches Vorzeichen, wie die Korrelationskoeffizienten, je nachdem die Beziehungsgleichung in expliziter oder impliziter Form dargestellt wird.

Die Grenzen der partiellen Korrelationskoeffizienten vierter Ordnung ergeben sich aus Ausdrücken f, die nach 28a) zu bilden sind. Wir haben im ganzen deren vier:

$$f_{123,4},\ f_{124,3},\ f_{134,2} \text{ und } f_{234,1}$$

zu unterscheiden. Der hinter dem Komma auftretende Index entspricht jener Größe, welche schon in den dreiseitigen Korrelationskoeffizienten ausgeschaltet, d. h. als Konstante behandelt wird. Die andere, in $r_{12,34}$ etwa die Größe x_3, wird den Größen gleichgesetzt, zwischen denen der Korrelationskoeffizient die Beziehung zu messen hat. Somit ist beispielsweise $f_{123,4}$ symmetrisch in $r_{12,4}$, $r_{13,4}$ und $r_{23,4}$ und da aus diesen dreiseitigen r-Werten drei vierseitige sich ergeben, sind diese auch $f_{123,4}$ zuzuordnen, was heißt, daß von den f-Werten aus gesehen insgesamt 12 Korrelationskoeffizienten vierter Ordnung zu unterscheiden sind. Davon sind aber, wie wir gesehen haben, je zwei einander gleich, so daß jeder der sechs verschiedenen zwischen zweien der f-Werte vermittelt. Allgemein ist

$$\begin{aligned} f_{abc,d} &= (1 - r^2_{ac,d})\,(1 - r^2_{bc,d})\,(1 - r^2_{ab,cd}) \\ &= (1 - r^2_{ab,d})\,(1 - r^2_{bc,d})\,(1 - r^2_{ac,bd}) \\ &= (1 - r^2_{ab,d})\,(1 - r^2_{ac,d})\,(1 - r^2_{bc,ad}) \\ &= 1 - r^2_{ab,d} - r^2_{ac,d} - r^2_{bc,d} + 2r_{ab,d}\, r_{ac,d}\, r_{bc,d} \end{aligned} \qquad 33)$$

Man sieht deutlich, daß $f_{abc,d}$ in gleicher Weise von den Korrelationskoeffizienten dritter Ordnung abhängig ist, wie f_{abc} von den r-Werten zweiter Ordnung. Da bereits nachgewiesen worden ist, daß die r-Werte dritter Ordnung innerhalb desselben Wertebereiches bleiben, wie jene zweiter Ordnung, wird auch $f_{abc,d}$, wie f_{abc}, nie kleiner als Null und nie größer als Eins sein. Daraus folgt aber:

$$0 \leqq |\, r_{ab,cd} \,| \leqq 1$$

Die Korrelationskoeffizienten vierter Ordnung haben also Werte zwischen — 1 und + 1 mit Einschluß dieser Grenzen.

Die neu gewonnenen Koeffizienten gestatten nun, die Beziehungsgleichung zwischen vier verbundenen Größen aufzustellen. Setzen wir der Einfachheit halber x_k für

$$x_k - m_{x_k}$$

dann wird

$$x_1 = \frac{\sigma_{1,34}}{\sigma_{2,34}} \cdot r_{12,34} \cdot x_2 + \frac{\sigma_{1,24}}{\sigma_{3,24}} \cdot r_{13,24} \cdot x_3 + \frac{\sigma_{1,23}}{\sigma_{4,23}} r_{14,23} \cdot x_4$$

Dividiert man die Gleichung durch die drei im Zähler auftretenden Streuungsmaße, dann geht sie in die „Normalform" über (vgl. 27) und 23)). — Zu obiger Gleichung gehören noch drei zusätzliche, in denen x_2, bzw. x_3, bzw. x_4 durch die übrigen x_k ausgedrückt werden.

Wir haben noch die Beziehung zwischen den f-Werten vierter Ordnung zu untersuchen. Es ist

$$\frac{f_{abc,d}}{f_{abd,c}} = \frac{(1 - r^2_{ac,d})(1 - r^2_{bc,d})}{(1 - r^2_{ad,c})(1 - r^2_{bd,c})} \frac{(1 - r^2_{ab,cd})}{(1 - r^2_{ab,cd})}$$

$$= \frac{(1 - r^2_{ac,d})(1 - r^2_{bc,d})}{(1 - r^2_{ad,c})(1 - r^2_{bd,c})} = \frac{(1 - r^2_{ac})(1 - r^2_{bc})}{(1 - r^2_{ad})(1 - r^2_{bd})}$$

Besteht zwischen x_c und x_d eine streng lineare Beziehung, dann geht dieser Quotient in

$$\frac{f_{abc}}{f_{abd}} = 1$$

über. Je lockerer dagegen der Zusammenhang zwischen x_c und x_d, also zwischen den Größen, die in den Korrelationskoeffizienten $r_{ab,cd}$ als Konstante behandelt werden, ist, um so stärker werden sich auch die beiden f-Werte vierter Ordnung voneinander unterscheiden. Ihr Quotient kann somit als Maß der Beziehung zwischen den Konstanten aufgefaßt werden.

c) n verbundene Größen.

Es ist leicht einzusehen, daß das Gleichungssystem

$$\sum_1^n b_i [x_1 x_i] = 0$$

$$\sum_1^n b_i [x_2 x_i] = 0$$

$$\vdots$$

$$\sum_1^n b_i [x_n x_i] = 0$$

Koeffizienten $b_{ik,\ldots}$ und aus diesen abgeleitete $r_{ik,\ldots}$, $\sigma_{i,\ldots}$, $f_{ikl,\ldots}$ liefern wird, die sich von den entsprechenden dritter und vierter Ordnung nicht in der Form, sondern nur in der Zahl der Konstanten unterscheiden werden. Das Ableitungsverfahren, das an zwei Beispielen ausführlich dargestellt worden ist, ergibt folgende allgemeine Formeln:

$$r_{ab,cd\ldots n} = \frac{r_{ab,d\cdots n} - r_{ac,d\cdots n}\, r_{bc,d\cdots n}}{\sqrt{(1 - r^2_{ac,d\ldots n})\,(1 - r^2_{bc,d\ldots n})}} \qquad 34)$$

$$\sigma^2_{a,bc\ldots lmn} = \sigma^2_a\,(1 - r^2_{an})\,(1 - r^2_{am,n})\,(1 - r^2_{al,mn})\ldots$$
$$\ldots(1 - r^2_{ab,c\ldots lmn}) \qquad 35)$$

$$b_{ab,c\ldots n} = \frac{\sigma_{a,c\ldots n}}{\sigma_{b,c\ldots n}} \cdot r_{ab,c\ldots n}$$
$$b_{ba,c\ldots n} = \frac{\sigma_{b,c\ldots n}}{\sigma_{a,c\ldots n}} \cdot r_{ab,c\ldots n} \qquad 36)$$

$$\begin{aligned} f_{abc,d\ldots n} &= (1 - r^2_{ac,d\ldots n})\,(1 - r^2_{bc,d\ldots n})\,(1 - r^2_{ab,cd\ldots n}) \\ &= (1 - r^2_{ab,d\ldots n})\,(1 - r^2_{bc,d\ldots n})\,(1 - r^2_{ac,bd\ldots n}) \\ &= (1 - r^2_{ab,d\ldots n})\,(1 - r^2_{ac,d\ldots n})\,(1 - r^2_{bc,ad\ldots n}) \\ &= 1 - r^2_{ab,d\ldots n} - r^2_{ac,d\ldots n} - r^2_{bc,d\ldots n} \\ &\quad + 2r_{ab,d\ldots n}\, r_{ac,d\ldots n}\, r_{bc,d\ldots n} \end{aligned} \qquad 37)$$

Sie bestätigen schon früher gemachte Feststellungen, soweit sich diese auf den Wertebereich der Korrelationskoeffizienten, auf deren Größenbeziehungen untereinander, auf das Auftreten von bestimmten Vorzeichen bei Korrelations- und Regressionskoeffizienten usw. beziehen. Man lese hierüber die Ausführungen pag. 79/83 und 87 nach.

Wir haben nun weiter noch zu untersuchen, ob auch bei Koeffizienten n-ter Ordnung die Möglichkeit der Permutierung der Konstanten besteht, wie sie bereits nachgewiesen worden ist für Koeffizienten vierter Ordnung. Dies ist schon darum notwendig, weil nach Exner („Über die Korrelationsmethode") von Yule Formeln angegeben werden, die von den oben angeführten etwas abweichen. So ist nach Yule

$$r_{12,34\ldots m} = \frac{r_{12,34\ldots(m-1} - r_{1m,34\ldots(m-1)}\, r_{2m,34\ldots(m-1)}}{\sqrt{(1 - r^2_{1m,34\ldots(m-1)})\,(1 - r^2_{2m,34\ldots(m-1)})}}$$

Ändern wir in diesem Korrelationskoeffizienten die Reihenfolge der Konstanten in

$$r_{12,m34\ldots(m-1)},$$

dann ergibt sich auch nach 34) der angeführte Ausdruck. Wenn also nachgewiesen werden kann, daß diese Änderung den Wert des Korrelationskoeffizienten nicht beeinflußt, dann ist damit auch gezeigt, daß nicht nur nach der Yuleschen Formel das r n-ter Ordnung berechnet werden kann, sondern daß es insgesamt (n — 2)! Möglichkeiten gibt, r zu ermitteln. — Yule gibt ferner folgendes allgemeine Maß der bedingten Streuung an:

$$\sigma^2_{1,23\ldots m} = \sigma^2_1(1 - r^2_{12})\,(1 - r^2_{13,2})\,(1 - r^2_{14,23})\ldots$$
$$\cdot\,(1 - r^2_{1m,23\ldots(m-1)})$$

Führen wir bei diesem Streuungskoeffizienten die Konstanten in der umgekehrten Reihenfolge auf, dann erhalten wir wiederum nach 35) die Yulesche Formel. Wir haben also zu zeigen, daß auch bei den Streuungsmaßen die Reihenfolge der hinter dem Komma aufgeführten Konstanten beliebig ist. — Endlich hat man nach Yule die Regressionskoeffizienten nach folgender allgemeiner Formel zu berechnen:

$$b_{12,3\ldots m} = r_{12,3\ldots m} \cdot \frac{\sigma_{1,23\ldots m}}{\sigma_{2,13\ldots m}}$$

Sie unterscheidet sich von 36) insofern, als sie die Regressionskoeffizienten b aus den zugehörigen r-Werten mit Größen σ höherer Ordnung bestimmt. Auf diesen Umstand ist bereits bei der Besprechung der Koeffizienten dritter und vierter Ordnung hingewiesen worden. Es ist also unsere Aufgabe, festzustellen, ob für jedes beliebige m der Quotient

$$\frac{\sigma_{1,23\ldots m}}{\sigma_{2,13\ldots m}} = \frac{\sigma_{1,3\ldots m}}{\sigma_{2,3\ldots m}}$$

ist. —

Wir fragen uns vorerst, ob sich der Wert eines Korrelationskoeffizienten nicht ändert, wenn die Konstanten permutiert werden. Es genügt dabei, zu untersuchen, ob etwa

$$r_{12,3k\ldots n} = r_{12,k3\ldots n}$$

sei, das heißt die Änderung der Reihenfolge der beiden dem Komma unmittelbar folgenden Indizes (Konstanten) keinen Einfluß auf die Größe von r habe. Denn nach 34) wird jedes r n-ter Ordnung auf

r-Werte (n — 1)-ter Ordnung zurückgeführt, bei denen die Möglichkeit einer solchen Inversion der Indizes auch besteht. Geht man bis auf die Korrelationskoeffizienten vierter Ordnung zurück und zieht man alle Inversions-, bzw. Transpositionsmöglichkeiten in Betracht, dann lassen sich auch für $r_{12,\ldots n}$ alle (n — 2)! Permutationen der Indizes hinter dem Komma erhalten. Es ist nun

$$r_{12,3k..n} = \frac{r_{12,k..n} - r_{13,k..n}\, r_{23,k..n}}{\sqrt{(1 - r^2_{13,k..n})\,(1 - r^2_{23,k..n})}}$$

und

$$r_{12,k3..n} = \frac{r_{12,3..n} - r_{1k,3..n}\, r_{2k,3..n}}{\sqrt{(1 - r^2_{1k,3..n})\,(1 - r^2_{2k,3..n})}}$$

Aus der Gleichheit dieser beiden Korrelationskoeffizienten wäre nun zu schließen, daß

$$\frac{\sqrt{(1 - r^2_{13,k..n})\,(1 - r^2_{23,k..n})}}{\sqrt{(1 - r^2_{1k,3..n})\,(1 - r^2_{2k,3..n})}} = \frac{r_{12,k..n} - r_{13,k..n}\, r_{23,k..n}}{r_{12,3..n} - r_{1k,3..n}\, r_{2k,3..n}}$$

sei. Formen wir diese „Gleichung" in der pag. 60 dargestellten Weise um, dann geht die linke Seite über in

$$\frac{\sqrt{f_{13k,..}\, f_{23k,..}\,(1 - r^2_{13,..})\,(1 - r^2_{3k,..})\,(1 - r^2_{23,..})\,(1 - r^2_{3k,..})}}{\sqrt{(1 - r^2_{1k,..})\,(1 - r^2_{3k,..})\,(1 - r^2_{2k,..})\,(1 - r^2_{3k,..})\, f_{13k,..}\, f_{23k,..}}}$$

$$= \frac{\sqrt{(1 - r^2_{13,..})\,(1 - r^2_{23,..})}}{\sqrt{(1 - r^2_{1k,..})\,(1 - r^2_{2k,..})}} = k$$

Für die rechte Seite ergibt sich

$$k \cdot \frac{\begin{array}{l} r_{12,..} - r_{1k,..}\, r_{2k,..} - r_{12,..}\, r^2_{3k,..} \pm r_{1k,..}\, r_{2k,..}\, r^2_{3k,..} \\ \quad - r_{13,..}\, r_{23,..} + r_{1k,..}\, r_{23,..}\, r_{3k,..} + r_{13,..}\, r_{2k,..}\, r_{3k,..} \end{array}}{\begin{array}{l} r_{12,..} - r_{13,..}\, r_{23,..} - r_{12,..}\, r^2_{3k,..} \pm r_{13,..}\, r_{23,..}\, r^2_{3k,..} \\ \quad - r_{1k,..}\, r_{2k,..} - r_{1k,..}\, r_{23,..}\, r_{3k,..} + r_{13,..}\, r_{2k,..}\, r_{3k,..} \end{array}}$$

$$= k$$

Aus diesem Ergebnis ist zu schließen, daß tatsächlich

$$r_{12,3k..} = r_{12k3..}$$

ist. Es besteht also für alle partiellen Korrelationskoeffizienten höherer als dritter Ordnung die Möglichkeit, die Indizes hinter dem

Komma zu permutieren. Jeder Indexpermutation entspricht eine bestimmte Anordnung der Konstanten, d. h. jener Größen, die bei der Messung der Beziehung zwischen irgend zwei x-Werten ausgeschaltet werden — und damit auch einem bestimmten Modus der Koeffizientenberechnung. Denn wenn auch grundsätzlich alle $r_{ab\ldots n}$ identisch gleich sind, wird doch sowohl in 34), als auch in der von Yule angegebenen Formel zwischen auszuschaltenden und ausgeschalteten Größen ein Unterschied gemacht und damit die Reihenfolge der Konstanten als wesentlich hingestellt. Haben wir die Beziehungen zwischen n verbundenen Größen zu ermitteln, dann läßt sich somit jeder der $\binom{n}{2}$ Korrelationskoeffizienten n-ter Ordnung auf $(n-2)!$ Arten berechnen. Ein möglicher Berechnungsmodus wird in der Yuleschen Formel, ein anderer in 34) angegeben. Beide sind gleichwertig.

Wir haben in zweiter Linie nun festzustellen, ob bei den Streuungsmaßen σ die Reihenfolge der Indizes hinter dem Komma geändert werden kann. Das ist dann der Fall, wenn u. a. aus zwei benachbarten Indizes eine Inversion sich bilden läßt, denn durch Vertauschungen benachbarter Elemente erhält man jede beliebige Permutation. Es ist nach 35)

$$\sigma_{a,\ldots hk\ldots n} = \sigma_{a\ldots} \sqrt{1 - r^2_{ak,\ldots}} \cdot \sqrt{1 - r^2_{ah,k\ldots}} \cdots$$

$$= \frac{\sigma_{a\ldots} \sqrt{1 - r^2_{ak,\ldots}} \cdot \sqrt{f_{ahk,\ldots}} \cdots}{\sqrt{1 - r^2_{ak,\ldots}} \cdot \sqrt{1 - r^2_{hk,\ldots}}}$$

und

$$\sigma_{a,\ldots kh\ldots n} = \sigma_{a\ldots} \sqrt{1 - r^2_{ah,\ldots}} \cdot \sqrt{1 - r^2_{ak,h\ldots}} \cdots$$

$$= \frac{\sigma_{a\ldots} \sqrt{1 - r^2_{ah,\ldots}} \cdot \sqrt{f_{ahk,\ldots}} \cdots}{\sqrt{1 - r^2_{ah,\ldots}} \cdot \sqrt{1 - r^2_{hk,\ldots}}},$$

folglich

$$\sigma_{a,\ldots hk\ldots n} = \sigma_{a,\ldots kh\ldots n}$$

Auch bei den Streuungsmassen ist also die Reihenfolge der hinter dem Komma aufgeführten Konstanten beliebig, woraus wieder auf verschiedene Berechnungsmöglichkeiten und die Gleichwertigkeit von 35) mit der Formel von Yule geschlossen werden kann.

Nach Yule sind endlich — auch Exner und Pütter verwenden übereinstimmend die nämlichen Formeln — die Koeffizienten b mit Streuungsmaßen höherer Ordnung, als sie in 36) angegeben sind, zu berechnen. Streuungsmaße höherer Ordnung sind aber gleichbedeutend

mit vermehrter Rechenarbeit und darum kommt der Untersuchung, ob etwa

$$\frac{\sigma_{1,23\ldots n}}{\sigma_{2,13\ldots n}} \text{ auf } \frac{\sigma_{1,3\ldots n}}{\sigma_{2,3\ldots n}}$$

zurückgeführt werden könne, auch eine gewisse praktische Bedeutung zu. Nach 35) ist nun

$$\frac{\sigma_{1,23\ldots n}}{\sigma_{2,13\ldots n}} = \frac{\sigma_{1\ldots} \sqrt{1-r^2_{13,\ldots n}}\ \sqrt{1-r^2_{12,3\ldots n}}}{\sigma_{2\ldots} \sqrt{1-r^2_{23,\ldots n}}\ \sqrt{1-r^2_{12,3\ldots n}}}$$

Im Zähler und im Nenner fällt der letzte Faktor weg, so daß der Quotient übergeht in

$$\frac{\sigma_{1,3\ldots n}}{\sigma_{2,3\ldots n}}$$

Geben wir die Streuungsmaße n-ter Ordnung nach Yule an, dann erhalten wir

$$\frac{\sigma_{1,23\cdot\cdot n}}{\sigma_{2,13\cdot\cdot n}} = \frac{\sigma_1 \sqrt{1-r^2_{12}}\ \sqrt{1-r^2_{13,2}}\ \sqrt{1-r^2_{14,23}}\ldots}{\sigma_2 \sqrt{1-r^2_{12}}\ \sqrt{1-r^2_{23,1}}\ \sqrt{1-r^2_{24,13}}\ldots}$$

Die beiden zweiten Faktoren in Zähler und Nenner fallen durch Kürzung weg. Je die dritten ... n-ten Faktoren gehören demselben f-Werte an und können deshalb auf Faktoren niedrigerer Ordnung zurückgeführt werden. So ist zum Beispiel

$$\frac{\sqrt{1-r^2_{13,2}}}{\sqrt{1-r^2_{23,1}}} = \frac{\sqrt{1-r^2_{13}}}{\sqrt{1-r^2_{23}}},$$

$$\frac{\sqrt{1-r^2_{14,23}}}{\sqrt{1-r^2_{24,13}}} = \frac{\sqrt{1-r^2_{14,3}}}{\sqrt{1-r^2_{24,3}}}$$

usw.

Somit geht der Quotient über in

$$\frac{\sigma_1 \sqrt{1-r^2_{13}}\ \sqrt{1-r^2_{14,3}}\ldots}{\sigma_2 \sqrt{1-r^2_{23}}\ \sqrt{1-r^2_{24,3}}\ldots} = \frac{\sigma_{1,3\ldots n}}{\sigma_{2,3\ldots n}}$$

Die Formel 36) besteht also zu Recht und dürfte der Yuleschen vorzuziehen sein.

Zusammenfassend stellen wir also fest:

Alle für die korrelativen Beziehungen zwischen n Größen maßgebenden Koeffizienten r, σ und b haben gemeinsam, daß sie formal, aber nicht grundsätzlich zwischen den Konstanten, bzw. den Indizes, die für diese stehen, einen Unterschied machen. Das hat zur Folge, daß es eine vom Grad der Koeffizienten abhängige Zahl von Möglichkeiten gibt, diese zu berechnen. Als zusätzliches Ergebnis darf ferner vermerkt werden, daß die Berechnung der Regressionskoeffizienten aus den zugehörigen r-Werten mit Hilfe von Streuungsmaßen niedriger Ordnung, als Yule sie angibt, erfolgen kann.

Einige Eigenschaften der partiellen Korrelationskoeffizienten

Wir haben den zweiseitigen Korrelationskoeffizienten als Maß für die zwischen zwei Größen bestehende Beziehung kennen gelernt. Je enger die Beziehung ist, um so mehr nähert sich der absolute Wert von r der Grenze Eins. Umgekehrt läßt ein Wert $r = 0$ auf das völlige Fehlen einer Beziehung schließen. Sind mehrere Größen untereinander verbunden, dann mißt der partielle Korrelationskoeffizient die Beziehung zwischen je zweien, indem er die übrigen = konst. setzt und damit ihren Einfluß ausschaltet. Nun ist zu berücksichtigen, daß sich die Annahme der Existenz mehrseitiger Beziehungen auf die Voraussetzung primär zweiseitiger Zusammenhänge stützt. Das Warum? ist leicht zu beantworten. Es besteht nämlich niemals die Möglichkeit, auf empirischem Wege Veränderungen einer Größe als bedingt, oder bedingend (darüber entscheidet unsere Vorstellung von der Art der Beziehung) festzustellen, wenn nicht mit dieser Größe auch andere Größen sich verändern. Diesem Umstand trägt auch das Bildungsgesetz der maßgebenden Koeffizienten Rechnung: Ein r-Wert n-ter Ordnung wird auf r-Werte (n — 1)-ter und damit schließlich auf solche zweiter Ordnung zurückgeführt. — Nehmen wir nun an, es sei bei drei vorgegebenen Größen x_1, x_2 und x_3 einerseits die Beziehung zwischen x_1 und x_2, anderseits jene zwischen x_2 und x_3 als ausgesprochen eng festgestellt worden. Nun wird offenbar der Korrelationskoeffizient 3. Ordnung $r_{12,3}$, der die Beziehung zwischen x_1 und x_2 unter Ausschaltung des Einflusses von x_3 mißt, keinen sehr hohen Wert haben können, da für x_3 = konst. x_2 den Bewegungen von x_1 kaum in dem Maße folgt, wie es die enge Verbindung zwischen x_1 und x_2 erwarten ließe. Je größer die Zahl der miteinander in Zusammenhang stehenden Größen ist, um so kleiner müßte also der Einfluß, ausgedrückt durch die partiellen Korrelationskoeffizienten, sein, den

jede einzelne auf eine bestimmte „bedingte" Größe hat. Nehmen wir an, es werden bei n verbundenen Größen alle zweiseitigen Beziehungen durch den gleichen Korrelationskoeffizienten gemessen. Dann werden auch alle partiellen r-Werte gleich groß und wir erhalten — der Einfachheit halber sei für

$$r_{ab,\ldots n} \longrightarrow r_{(n)}$$

gesetzt —

$$r_{(2)} = r \qquad \text{und für} \qquad r = 1:$$

$$r_{(3)} = \frac{r}{1+r} \qquad = \frac{1}{2}$$

$$r_{(4)} = \frac{r}{1+2r} \qquad = \frac{1}{3}$$

$$r_{(5)} = \frac{r}{1+3r} \qquad = \frac{1}{4}$$

$$\vdots$$

$$r_{(n)} = \frac{r}{1+(n-2)\,r} \qquad = \frac{1}{n-1}$$

Wir stellen also fest, daß für den Fall gleicher zwei- und mehrseitiger Beziehungen der Wert der Korrelationskoeffizienten um so kleiner ausfällt, je höher ihre Ordnungszahl bzw. die Zahl der verbundenen Größen ist. Ist insbesondere der zweiseitige Koeffizient r gleich 1 und läßt somit auf linear-funktionelle Beziehungen zwischen je zwei Größen schließen, dann wird der zahlenmäßig durch 1 ausgedrückte Gesamteinfluß gleichmäßig unter alle n — 1 „unabhängige" Größen aufgeteilt. Daß das so sein muß, kann leicht gezeigt werden. Es sei

$$x_1 = a_{12}\,x_2 = a_{13}\,x_3 = \ldots\ldots = a_{1n}\,x_n$$

und entsprechend

$$x_2 = a_{21}\,x_1 = a_{23}\,x_3 = \ldots\ldots = a_{2n}\,x_n$$

usw.

wobei die Proportionalitätskonstanten a_{ik} positiv oder negativ sein können. Nun ist zum Beispiel

$$\frac{x_1}{n-1} = \frac{a_{12}\,x_2}{n-1} = \frac{a_{13}\,x_3}{n-1} = \ldots\ldots = \frac{a_{1n}\,x_n}{n-1}$$

so daß sich ergibt:

$$x_1 = \frac{1}{n-1}(a_{12}\,x_2 + a_{13}\,x_3 + \ldots + a_{1n}\,x_n),$$

oder, wenn man die a_{1k} durch $\frac{\sigma_1}{\sigma_2}$ ersetzt:

$$x_1 = \frac{1}{n-1}\left(\frac{\sigma_1}{\sigma_2}x_2 + \frac{\sigma_1}{\sigma_3}x_3 + \ldots + \frac{\sigma_1}{\sigma_n}x_n\right)$$

Ebenso erhalten wir:

$$x_2 = \frac{1}{n-1}\left(\frac{\sigma_2}{\sigma_1}x_1 + \frac{\sigma_2}{\sigma_3}x_3 + \ldots + \frac{\sigma_2}{\sigma_n}x_n\right) \qquad 38)$$

$$\vdots$$

$$x_n = \frac{1}{n-1}\left(\frac{\sigma_n}{\sigma_1}x_1 + \frac{\sigma_n}{\sigma_2}x_2 + \ldots + \frac{\sigma_n}{\sigma_{n-1}}x_{n-1}\right)$$

Es ist klar, daß eine Gleichung allein den zwischen je zwei Größen bestehenden funktionellen Beziehungen nicht gerecht wird. Ist beispielsweise x_k als Funktion aller übrigen x_i dargestellt, dann ist irgendein x_m mit einem x_n nur über x_k und nicht direkt verbunden. Alle Größen außer x_k werden somit unzutreffenderweise als unabhängige Veränderliche gekennzeichnet. Das n Gleichungen umfassende System 38) dagegen entspricht den Voraussetzungen, von denen ausgegangen wurde: Es ist nur identisch erfüllt, wenn gesetzt werden kann:

$$x_k = \frac{\sigma_k}{\sigma_1}x_1 = \frac{\sigma_k}{\sigma_2}x_2 = \ldots = \frac{\sigma_k}{\sigma_n}x_n$$

Der Vergleich von 38) mit dem System der Regressionsgleichungen:

$$x_1 = r_{12,\cdot\cdot n}\,\frac{\sigma_{1,3\cdot\cdot n}}{\sigma_{2,3\cdot\cdot n}}\cdot x_2 + r_{13,\cdot\cdot n}\,\frac{\sigma_{1,2\cdot\cdot n}}{\sigma_{3,2\cdot\cdot n}}\cdot x_3 + \ldots$$

$$x_2 = r_{12,\cdot\cdot n}\,\frac{\sigma_{2,3\cdot\cdot n}}{\sigma_{1,3\cdot\cdot n}}\cdot x_1 + r_{23,\cdot\cdot n}\,\frac{\sigma_{2,1\cdot\cdot n}}{\sigma_{3,1\cdot\cdot n}}\cdot x_3 + \ldots \qquad 39)$$

$$x_3 = r_{13,\cdot\cdot n}\,\frac{\sigma_{3,2\cdot\cdot n}}{\sigma_{1,2\cdot\cdot n}}\cdot x_1 + r_{23,\cdot\cdot n}\,\frac{\sigma_{2,3\cdot\cdot n}}{\sigma_{2,1\cdot\cdot n}}\cdot x_2 + \ldots$$

usw.

zeigt nun deutlich, daß der Faktor $\frac{1}{n-1}$ für die partiellen Korrelationskoeffizienten steht; sind diese unter sich gleich, dann erhält man als allgemeine Beziehungsgleichung

$$x_k = \frac{\sigma_k \cdot r}{1 + (n-2)\,r} \sum_1^n \frac{x_i}{\sigma_i}, \qquad 40)$$

die in eine der Gleichungen von 38) übergeht, wenn man $r = 1$ setzt. Sind alle x_i unter sich gleich und gleich x_h, dann wird weiter

$$x_k = \frac{(n-1)\,r}{1 + (n-2)\,r} \cdot \frac{\sigma_k}{\sigma_h} \cdot x_h \qquad 41)$$

In diesem Ausdruck ist n, die Zahl der verbundenen Größen in N ($= n - 1$), die Zahl der Fälle, in denen zwischen den zwei Größen x_k und x_h derselbe korrelative Zusammenhang festgestellt werden kann, überzuführen. Dann ergibt sich:

$$x_k = \frac{Nr}{1 + (N-1)\,r} \cdot \frac{\sigma_k}{\sigma_h} \cdot x_h \qquad 42)$$

Für $N = 1$ wird

$$x_k = r \cdot \frac{\sigma_k}{\sigma_h} \cdot x_h;$$

42) geht also in 6) über.

$$w_N = \frac{Nr}{1 + (N-1)\,r} \qquad 43)$$

entspricht somit dem Korrelationskoeffizienten r, wenn die Messung des Zusammenhangs zwischen x_k und x_h N-mal denselben Wert ergeben hat. Da nun w_N mit wachsendem N zunimmt und dabei für jeden von Null verschiedenen Wert von r asymptotisch der Grenze 1 sich nähert, ziehen wir den Schluß,

daß die Wahrscheinlichkeit eines funktionellen Zusammenhangs zwischen x_k und x_h mit der Zahl der Fälle, in denen sich dieselbe korrelative Beziehung zwischen beiden Größen ergab, zunimmt.

Maß dieser Wahrscheinlichkeit ist w_N. Ergeben sich in N Bestimmungen verschiedene Werte von r, dann ist entsprechend zu setzen:

$$w_N = \frac{\Sigma\, r}{1 + (N-1)\dfrac{\Sigma\, r}{N}} \qquad 44)$$

Auch in diesem Fall nimmt w_N mit N zu.

Nach dieser Feststellung vom Wesen und der Bedeutung der zweiseitigen Beziehungen, die sich aus der Betrachtung der partiellen

Korrelationskoeffizienten ergeben hat, kehren wir zu den letztern zurück. Die r-Werte n-ter Ordnung werden, wie wir gesehen haben, um so kleiner, je enger die zweiseitigen Beziehungen zwischen n Größen sind. Deren mehrfache Messung wird in diesem Falle wichtiger sein, als die Durchführung der partiellen Korrelationsrechnung. Zu untersuchen bleibt nun aber, unter was für Umständen sich hohe Werte für die partiellen Korrelationskoeffizienten ergeben, bzw. was solche hohe Werte zu bedeuten haben.

Bereits wurde gezeigt, daß die zwischen n Größen bestehenden korrelativen Beziehungen nicht in einer Gleichung allein dargestellt werden können. Die Koeffizienten r n-ter Ordnung sind ja partielle, d. h. sie messen nach geltender Auffassung den Zusammenhang zwischen zwei Größen unter Ausschaltung weiterer und nehmen infolgedessen auch primäre zweiseitige Beziehungen an. Ändert sich also eine Größe, dann müssen grundsätzlich alle mit ihr verbundenen Größen sich auch ändern und ist ein Teil derselben gleich konstans oder gleich Null gesetzt, dann ist dies auch bei den übrigen der Fall. Eine Gleichung, die erwa x_k durch die übrigen x_i ausdrückt, kann unmöglich all diesen gegenseitigen Bindungen gerecht werden; sie stellt vielmehr die x_i als unabhängige Veränderliche dar, d. h. billigt ihnen die Möglichkeit zu, sich innerhalb beliebiger Grenzen unabhängig voneinander zu verändern. Diese Möglichkeit ist aber auszuschließen. Vom Standpunkt der partiellen Korrelationsrechnung hat man unter linearer Beziehung einen Zusammenhang zu verstehen, der entsprechend 38) alle zwischen n Größen bestehenden Bindungen in einem System von n Regressionsgleichungen [vgl. 39)] darstellt. In diesem geben die partiellen Korrelationskoeffizienten an, wie die Größen x_i, durch die etwa x_k ausgedrückt wird, an dem auf x_k ausgeübten Gesamteinfluß partizipieren. Dieser ist bei streng linearem Zusammenhang gleich 1 zu setzen.

Wenn es üblich ist, zu erklären, der partielle Korrelationskoeffizient messe die Beziehung zwischen zwei Größen bei Konstanthaltung (Pütter), oder Nullsetzung (Exner) aller übrigen, so ist diese Erklärung dahingehend zu ergänzen, daß die auf die verschiedenen x_i entfallenden Teilbeziehungen oder Teileinflüsse als additive Komponenten aufzufassen sind, von denen jede nur auf Kosten aller übrigen groß sein kann. Es ist dem partiellen Korrelationskoeffizienten als solchem ein Höchstwert zuzubilligen, aber Höchstwerte in Mehrzahl schließen sich gegenseitig aus. Das ergibt sich auch aus folgendem:

Es wurde bereits festgestellt (vgl. pag. 82), daß bei linear-funktionellem Zusammenhang zwischen zwei Größen die Beziehungen

beider zu einer dritten Größe durch den selben Korrelationskoeffizienten gekennzeichnet sind. Für

$$r_{12} = \pm 1 \quad \text{ist also}$$

$$r_{1k} = \pm r_{2k}$$

Daraus folgt weiter:

$$r_{12,k} = \pm \frac{r_{12} - r_{1k}\, r_{2k}}{\sqrt{1 - r_{1k}^2}\, \sqrt{1 - r_{2k}^2}} = \pm 1$$

$$r_{1k,2} = \pm \frac{r_{1k} - r_{12}\, r_{2k}}{\sqrt{1 - r_{12}^2}\, \sqrt{1 - r_{3k}^2}} = 0$$

und ebenso $r_{2k,1} = 0$; ferner

$$r_{1k,i} = \pm \frac{r_{1k} - r_{1i}\, r_{ki}}{\sqrt{1 - r_{1i}^2}\, \sqrt{1 - r_{ki}^2}}$$

$$r_{2k,i} = \pm \frac{r_{2k} - r_{2i}\, r_{ki}}{\sqrt{1 - r_{2i}^2}\, \sqrt{1 - r_{ki}^2}},$$

also $$r_{1k,i} = \pm r_{2k,i}$$

Wir können also allgemein schließen:

Ist $$r_{hk} = \pm 1,$$

dann wird auch

$$r_{hk,\ldots n} = \pm 1 \qquad 45)$$

und

$$r_{hi,\ldots n} = \pm r_{ki,\ldots n}$$

$$r_{hi,k\ldots n} = r_{ki,h\ldots n} = 0 \qquad 46)$$

Eine streng lineare Beziehung zwischen zwei Größen bleibt somit auch dann streng linear, wenn man Einflüsse weiterer Größen auf die ersten beiden berücksichtigt.

Hat man bei der Untersuchung der partiellen korrelativen Beziehungen zwischen mehreren Größen festgestellt, daß zwischen zweien von ihnen der Zusammenhang ein streng linearer ist, dann wird die Verbindung beider mit derselben dritten Größe durch den absolut gleichen Koeffizienten r gemessen. Dabei ist vorausgesetzt, daß die an der Verbindung nicht beteiligte Größe nicht gleich konstant gesetzt wird; ist dies dagegen der Fall, dann erhält der Koeffizient r den Wert Null.

Es darf an dieser Stelle abschließend noch darauf hingewiesen werden, daß das in 45) und 46) dargestellte Verhalten der partiellen Korrelationskoeffizienten durchaus den Vorstellungen entspricht, die wir uns vom Wesen und von den Eigenschaften mehrseitiger Beziehungen machen. Das gilt auch für den eingangs dieser Betrachtungen untersuchten Spezialfall gleicher Beziehungen zweiter Ordnung.

Korrelationsverhältnis und totaler Korrelationskoeffizient

In der partiellen Korrelationsrechnung treten neben den Korrelationskoeffizienten als der Form und teilweise auch dem Wesen nach neue Maße die der bedingten Streuung auf. Auf die bedingte Streuung haben wir im folgenden noch einzugehen, da sich aus ihr weitere Maße ergeben, die für die Beurteilung eines korrelativen Zusammenhangs von einiger Bedeutung sind.

Besteht zwischen zwei Größen, etwa zwischen y und x, eine einigermaßen lockere, d. h. korrelative Bindung, dann entspricht einem bestimmten $x = X_i$ nicht ein bestimmtes $y = Y_i$, wie bei funktionellen Zusammenhängen, sondern eine Gruppe von y_i-Werten, die im besten Fall eine Häufungsstelle Y_i gemäß dem Gaußschen Verteilungsgesetz besitzen. Nehmen wir als wahrscheinlichsten Wert Y_i wiederum jenen an, für den

$$\sum^{i} (y_i - Y_i)^2$$

zu einem Minimum wird, dann ist

$$Y_i = \frac{1}{n_i} \sum^{i} y_i$$

und

$$\sigma^{2}{}^{i}_{y} = \frac{1}{n_i} \sum (y_i - Y_i)^2$$

Durch $x = X_i$ ist also nicht nur ein Mittelwert Y_i, sondern auch eine Streuung σ^{i}_{y} bestimmt, die man füglich als bedingte bezeichnen könnte. Man zieht es aber vor, die quadratischen Abweichungen über alle i im ganzen Bereich zu summieren und das aus dieser Summe sich ergebende Mittel Quadrat der bedingten Streuung zu nennen. Ist die Zahl der y-Werte gleich N, dann ergibt sich

$$\sigma^{2}{}'_{y} = \frac{1}{N} \sum^{i} \sum^{i} (y_i - Y_i)^2$$

$$= \frac{1}{N} \sum^{i} p_i (y_i - Y_i)^2 \qquad 47)$$

Diese Art der Berechnung der bedingten Streuung gestaltet sich nun etwas mühsam. Es ist aber möglich, σ_y' auf die allgemeine Streuung σ_y zurückzuführen. Bezeichnet man — vgl. Fig. 17 — mit m_y das Mittel aller y-Werte im ganzen Bereich von X_a bis X_b, dann wird

$$\sum_i (y_i - m_y)^2 = \sum_i p_i ((y_i - Y_i) + (Y_i - m_y))^2$$

$$= \sum_i p_i (y_i - Y_i)^2 + \sum_i p_i (Y_i - m_y)^2$$

woraus sich durch Division mit N weiter ergibt:

$$\sigma_y^2 = \sigma_y'^2 + \sigma_{Y_i}^2 \qquad 48)$$

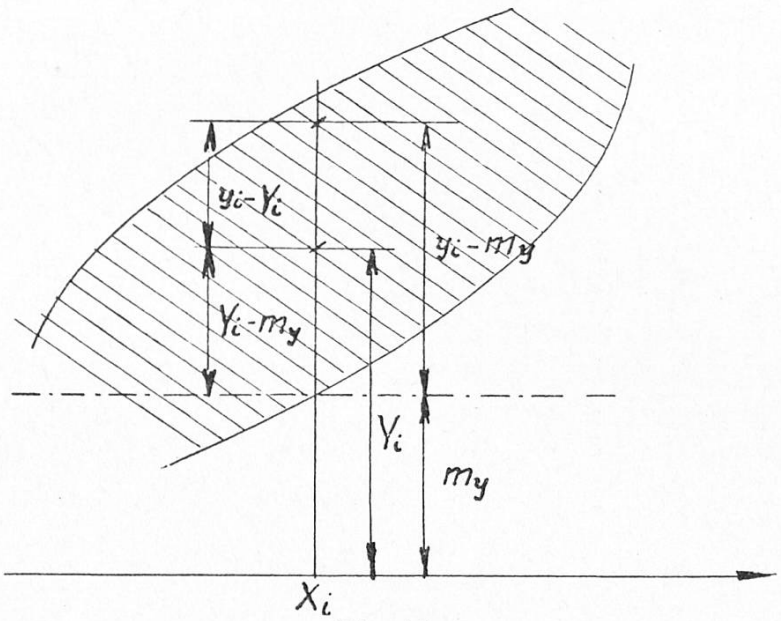

Fig. 17

Die bedingte Streuung ist also stets kleiner, als die allgemeine. Wäre $\sigma_y' = \sigma_y$, dann müßten alle Y_i mit m_y übereinstimmen. Das ist aber nur dann möglich, wenn zwischen y und x keine Beziehung besteht. Anderseits müssen für $\sigma_y' = 0$ die y_i mit den Y_i zusammenfallen; jedem X_i würde dann nur noch ein $y = Y_i$ entsprechen, was nur bei funktionellen Zusammenhängen zwischen y und x möglich ist. Ein Vergleich der bedingten Streuung mit der allgemeinen gibt uns also die Möglichkeit, den Grad der Verbundenheit zweier Veränderlicher zu bestimmen. Dividieren wir 48) mit σ_y^2, dann erhalten wir

$$\rho_y^2 = \left(\frac{\sigma_{Yi}}{\sigma_y}\right)^2 = 1 - \left(\frac{\sigma_y'}{\sigma_y}\right)^2 \qquad 49)$$

und entsprechend

$$\rho_x^2 = \left(\frac{\sigma_{Xi}}{\sigma_x}\right)^2 = 1 - \left(\frac{\sigma_x'}{\sigma_x}\right)^2$$

ρ^2 wird nach Pearson als Korrelationsverhältnis bezeichnet. Wie 49) zeigt, ist es nur positiver Werte von Null bis Eins fähig, unterscheidet sich also in dieser Beziehung vom Korrelationskoeffizienten. Dagegen ist es wie dieser in der Lage, festzustellen, wie eng die Bindung zwischen y und x ist. Aus dem obern Grenzwert $+1$ ergibt sich ein funktioneller Zusammenhang; 0 läßt auf das Fehlen einer korrelativen Bindung schließen. Während aber der Korrelationskoeffizient nur hohe Werte annehmen kann, wenn zwischen zwei Veränderlichen eine lineare Beziehung besteht, sagt das Korrelationsverhältnis nichts über die Art des Abhängigkeitsgesetzes aus. Darin liegt auch sein Vorzug vor dem Korrelationskoeffizienten: ρ^2 ist tatsächlich ein Abhängigkeitsmaß, während r nur dann diese Bedeutung zukommt, wenn man die Beziehung mit Sicherheit als lineare ansprechen kann.

Der Zusammenhang zwischen r und ρ^2 ergibt sich aus folgender Überlegung:

Hat man einer Reihe von x-Werten eine Reihe von y-Werten so zugeordnet, daß irgendeinem $x = X_i$ ein oder mehrere Werte y_i zugehören, dann wird der Korrelationskoeffizient r die Aufstellung folgender Regressionsgleichung erlauben:

$$Y - m_y = r \cdot \frac{\sigma_y}{\sigma_x} \cdot (X_i - m_x)$$

Geben wir X_i vor, dann wird sich zu diesem Wert ein einziges Y berechnen lassen, das die Gleichung identisch erfüllt. Setzen wir dagegen in der Gleichung eines der Wertepaare der Korrelationstabelle ein, dann wird sie bis auf einen Fehler v richtig sein. Bekanntlich wird nun r so bestimmt, daß die Summe aller Fehlerquadrate zu einem Minimum wird. Diese Summe ist nach Früherem

$$\frac{1}{N}\sum^{i}((y_i - m_y) - r \cdot \frac{\sigma_y}{\sigma_x} \cdot (X_i - m_x))^2 = \sigma_y^2 \cdot (1 - r^2)$$

Da
$$r \cdot \frac{\sigma_y}{\sigma_y} \cdot (X_i - m_x) = Y - m_y,$$

folgt:

$$\begin{aligned} & \sum^{i}((y_i - m_y) - (Y - m_y))^2 \\ = & \sum^{i}(y_i - Y)^2 \\ = & \sum^{i}((y_i - Y_i) + (Y_i - Y))^2 = N \cdot \sigma_y^2 \cdot (1 - r^2) \end{aligned}$$

Nach Division mit N wird somit

$$\sigma^{2'}_{y} = \sigma^2_y \cdot (1 - r^2) - \sigma^2_Y \qquad 50)$$

wenn $\sigma^2_Y = \frac{1}{N}(Y_i - Y)^2$ gesetzt wird.

Nun ist nach 49)

$$\sigma^{2'}_{y} = \sigma^2_y \cdot (1 - \rho^2_y),$$

woraus weiter folgt:

$$\sigma^2_y \cdot (1 - \rho^2_y) = \sigma^2_y \cdot (1 - r^2) - \sigma^2_Y$$

und $\sigma^2_y (\rho^2_y - r^2) = \sigma^2_Y,$

also $\rho^2_y = r^2 + (\frac{\sigma_Y}{\sigma_y})^2$

und ebenso 51)

$$\rho^2_x = r^2 + (\frac{\sigma_X}{\sigma_X})^2$$

Dieses Ergebnis besagt folgendes:

Ist das Abhängigkeitsgesetz kein lineares, dann wird das Korrelationsverhältnis ρ^2 größer, als der quadrierte Korrelationskoeffizient. Der Betrag, um den die Beziehung von einer linearen abweicht, wird zahlenmäßig durch

$$A^2_x = (\frac{\sigma_X}{\sigma_y})^2, \text{ bzw. } A^2_y = (\frac{\sigma_Y}{\sigma_y})^2 \qquad 51\,a)$$

ausgedrückt. Bei linearen Beziehungen zwischen y und x müssen die Streuungsmittel Y_i und X_i auf den den Regressionsgleichungen entsprechenden Geraden liegen, d. h. σ_X und σ_Y werden $= 0$ sein. In diesem Fall erhält man nach 50)

$$\sigma^{2'}_{y} = \sigma^2_y (1 - r^2) \qquad 52)$$

und $\sigma^{2'}_{x} = \sigma^2_x (1 - r^2)$

die bedingten Streuungen sind also gleich dem Summenmittel aller Fehlerquadrate.

Ist N, die Zahl der y-Werte, so klein, daß jedem X_i nur ein einziges Y_i $(= y_i)$ entspricht, dann wird die bedingte Streuung zu Null und damit $\rho^2 = 1$ und $r^2 = 1$ nach 49) und 52). Es wird in diesem Fall ein

funktioneller Zusammenhang zwischen y und x vorgetäuscht. Wenn man die bedingte Streuung und das Korrelationsverhältnis berechnen will, wird man stets dafür zu sorgen haben, daß den Werten jeder Veränderlichen mehrere Werte der andern entsprechen. Das wird gegebenenfalls durch eine passende Klasseneinteilung erreicht.

Die nachstehende Figur stellt die Beziehung zwischen den Streuungsmaßen, ρ und r und dem mittleren Fehlerquadrat, das der Einfachheit halber mit V^2 bezeichnet wird, in anschaulicher Weise dar.

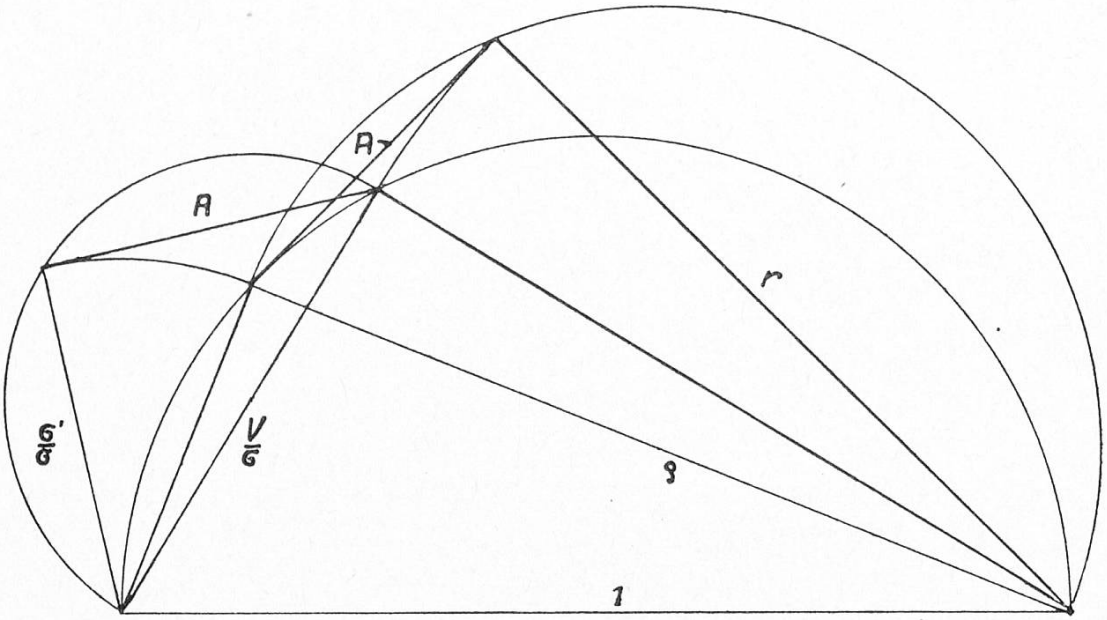

Fig. 18

Im nachfolgenden Beispiel soll die Anwendung der abgeleiteten Formeln dargestellt werden. Das Zahlenmaterial ist H. Rautmann: „Untersuchungen über die Norm" entnommen. Gegenstand der Untersuchung ist die Korrelation zwischen Körpergewicht und Körpergröße von Männern im Alter von 20 bis 25 Jahren. Die Tabelle enthält in der Horizontalen die um 49 kg verminderten Körpergewichte; in der Vertikalen sind die Körpergrößen in den 151,5 cm übersteigenden Beträgen aufgetragen. Intervallbreiten: Zeilen je 5 kg, Kolonnen je 4 cm.

Körper-größe cm	Körpergewicht kg 2,5 – 7,5	7,5 – 12,5	12,5 – 17,5	17,5 – 22,5	22,5 – 27,5	27,5 – 32,5	32,5 – 37,5	37,5 – 42,5	Summe	
32—36				3	1	2	2		8	8
28—32			3	11	12	3	2	1	32	7
24—28		1	9	37	18	5	1		71	6
20—24	1	7	44	55	21	8	3		139	5
16—20	4	40	72	42	10	1			169	4
12—16	13	51	51	17	4				136	3
8—12	23	31	12	3	2				71	2
4— 8	9	10	3						22	1
Summe	50	140	194	168	68	19	8	1	648	
	1	2	3	4	5	6	7	8		

1. Bestimmung der Mittelwerte jeder Zeile und jeder Kolonne. — Bei gleichbleibender Intervallbreite und von Eins verschiedenen Häufigkeitswerten ist die Anwendung eines besonderen Summierungsverfahrens von Vorteil: Wird

$$\begin{array}{ll} a & n_1 \\ a + d & n_2 \\ a + 2d & n_3 \\ a + 3d & n_4 \\ \ldots\ldots\ldots\ldots & \\ a + (n-1)\,d & n_n \end{array}$$

mit a der Mittelwert des ersten Intervalls und mit d die Intervallbreite bezeichnet und ist n_k der dem Mittelwert $a + (k-1)d$ zugehörende Häufigkeitswert, dann ist

$$an_1 + (a + d)n_2 + (a + 2d)n_3 + \ldots + (a + (n-1)\,d)n_n$$

$$= a \sum_1^n n + d\,(\sum_2^n n + \sum_3^n n + \ldots + \sum_{n-1}^n n) \qquad 53)$$

Die Anwendung dieser Formel ist einfach und bedarf keiner besonderen Erläuterungen. Die Tabelle ergibt folgende Mittelwerte:

X_i:	8,6	10,1	13,1	15,5	19,5	21,4	23,9	26,9
Y_i:	11,2	14,2	17,7	21,6	23,2	25,4	27,5	30,0

2. Die Berechnung des Korrelationskoeffizienten. — Man erhält vorerst als Mittel aller x-, bzw. aller y-Werte:

$$m_x = \frac{1}{648}(22.8{,}6 + 71.10{,}1 + \ldots) = 16{,}2$$

$$m_y = \frac{1}{648}(50.11{,}2 + 140.14{,}2 + \ldots) = 18{,}4$$

Durch diese Mittel ist in der Tabelle ein Achsenkreuz festgelegt, das durch jene Kolonne und jene Zeile gebildet wird, welche m_x und m_y enthalten. Es sind dies die 3. Kolonne und die 4. Zeile. Deren Mitten: $M_x = 15$ und $M_y = 18$ werden als angenäherte Mittel [vgl. 16) bis 18)] verwendet. Man findet weiter:

$$\Sigma\Delta x^2 = 27\,600 - 963{,}12 = 26\,636{,}88 \qquad \sigma_x^2 = 41{,}097$$

$$\Sigma\Delta y^2 = 23\,312 - 104{,}32 = 23\,207{,}68 \qquad \sigma_y^2 = 35{,}814$$

$$\Sigma\Delta x\Delta y = 16\,580 - 316{,}98 = 16\,263{,}02$$

$$r = 0{,}653 \pm 0{,}022 \qquad \sigma = 6{,}410 \qquad \sigma_y = 5{,}985$$

$$b_{21} = 0{,}610 \qquad b_{12} = 0{,}699$$

Zwischen Körpergewicht und Körpergröße besteht somit eine deutlich positive Korrelation. Nimmt das Körpergewicht um 1 kg zu, dann wird der Körper durchschnittlich um 0,61 cm länger. Umgekehrt entspricht einer Längenzunahme des Körpers um 1 cm eine Gewichtszunahme von 0,699 kg. Der Fehler des Korrelations- und der Regressionskoeffizienten macht rund 3,4 % aus. — Wir haben nun zu untersuchen, ob und in welchem Maße die Beziehung von einer linearen abweicht.

3. Berechnung der bedingten Streuung und des Korrelationsverhältnisses. — Wir benützen dazu ein erstes Mal die Formeln 48) und 49), ein anderes Mal 50) und 51) und haben auf diese Weise die Möglichkeit, zu prüfen, welche sich bei der praktischen Durchführung der Rechnung als vorteilhafter erweisen.

Tabelle a.

X_i	Y_i	m_y	Δ	n	$n \cdot \Delta^2$
5	11,2	18,4	— 7,2	50	2592,00
10	14,2	do.	— 4,2	140	2469,60
15	17,7	—	— 0,7	194	95,06
20	21,6	—	+ 3,2	168	1720,32
25	23,2	—	+ 4,8	68	1566,72
30	25,4	—	+ 7,0	19	931,00
35	27,5	—	+ 9,1	8	662,48
40	30,0	—	+11,6	1	134,56
				648	10171,74

$$\sigma^2_{Y_i} = 10\,171{,}74 : 648 = 15{,}697$$

$$\sigma^2_{y'} = \sigma^2_y - \sigma^2_{Y_i} = 35{,}814 - 15{,}697 = 20{,}117$$

$$\rho^2_y = 15{,}697 : 35{,}814 = 0{,}4383$$

$$\rho^2_y - r^2 = 0{,}0118$$

Benützt man zur Berechnung der bedingten Streuung und des Korrelationsverhältnisses die Formeln 50) und 51), dann hat man vorerst mit Hilfe der Regressionsgleichung

$$Y - 18{,}4 = 0{,}610\,(X_i - 16{,}2),$$

oder

$$Y = 0{,}61 \cdot X_i + 8{,}52$$

die Werte Y für $X_i = 5, 10, 15, \ldots$ zu bestimmen (Ber. von $\sigma^2_{x'}$ und ρ^2_x siehe nachher). Hierauf werden die Differenzen $\Delta = Y_i - Y$ ermittelt und dann gleich verfahren wie vorhin.

Tabelle b.

X_i	Y_i	Y	Δ	n	$n \cdot \Delta^2$
5	11,2	11,6	— 0,4	50	8,00
10	14,2	14,6	— 0,4	140	22,40
15	17,7	17,7	0	194	0
20	21,6	20,7	+ 0,9	168	136,08
25	23,2	23,8	— 0,6	68	24,48
30	25,4	26,8	— 1,2	19	27,36
35	27,5	29,9	— 2,3	8	42,32
40	30,0	32,9	— 2,9	1	8,41
				648	269,05

$$\sigma_Y^2 = 269{,}05 : 648 = 0{,}415$$

$$\sigma^2_{\cdot y} = 35{,}814 \cdot 0{,}5735 - 0{,}415 = 20{,}124$$

$$\rho_y^2 = r^2 + 0{,}0116 = 0{,}4381$$

Die Unterschiede, die sich bei der Berechnung von $\sigma^2_{\cdot y}$ und ρ_y^2 nach beiden Verfahren ergeben, sind auf die abgerundeten Werte von den nicht vorgegebenen X_i und Y_i, von m_x und m_y und endlich der Y aus der Regressionsgleichung zurückzuführen. Sie sind so geringfügig, daß sie unser Urteil über die Art der korrelativen Beziehung, das wir auf Grund der erzielten Ergebnisse fällen können, nicht zu beeinflussen oder zu beeinträchtigen in der Lage sind.

Da das zweite Verfahren eher weniger Rechenarbeit erfordert, dürfte es dem ersten vorzuziehen sein.

Es soll nun der Vollständigkeit halber noch $\sigma^2_{\cdot x}$ und ρ_x^2 berechnet werden.

Tabelle c.

Y_i	X_i	X	Δ	n	$n \cdot \Delta^2$
6	8,6	7,5	+ 1,1	22	26,62
10	10,1	10,3	— 0,2	71	2,84
14	13,1	13,1	0	136	0
18	15,5	15,9	— 0,4	169	27,04
22	19,5	18,7	+ 0,8	139	88,96
26	21,4	21,5	— 0,1	71	0,71
30	23,9	24,3	— 0,4	32	5,12
34	26,9	27,1	— 0,2	8	0,32
				648	151,61

$$\sigma_x^2 = 0{,}234 \quad \sigma^2_{\cdot x} = 23{,}337$$

$$\rho_x^2 = r^2 + 0{,}0057 = 0{,}4322$$

Wir stellten bereits fest, daß ρ_y^2 kaum von r^2 abweicht. Daraus ist auf eine lineare Beziehung zwischen Körpergewicht und Körpergröße zu schließen, was wiederum notwendig macht, daß auch ρ_x^2 nur wenig von r^2 sich unterscheidet. Die Rechnung bestätigt unsere Vermutung: Die Abweichungen der ρ^2 bleiben innerhalb der für den Korrelationskoeffizienten angegebenen Fehlergrenzen. So ist

ρ_y^2 nur um 2,7 %
ρ_x^2 nur um 1,3 % größer als r^2

Das Ergebnis ist darum auch wichtig, weil es zeigt, daß eine Übereinstimmung zwischen Korrelationsverhältnis und Korrelationskoeffizient auch bei einer beliebig lockeren linearen Beziehung möglich ist. Für jeden Wert von r wird also

$$\left(\frac{\sigma_Y}{\sigma_y}\right)^2, \text{ bzw. } \left(\frac{\sigma_X}{\sigma_x}\right)^2$$

die Abweichung der Beziehung von einer linearen zahlenmäßig ausdrücken. Die vollständige Untersuchung eines korrelativen Zusammenhangs wird also neben der Berechnung des Korrelationskoeffizienten und der b-Werte auch die Ermittlung des Korrelationsverhältnisses zu umfassen haben. Erst wenn wir die Gewißheit haben, daß eine Beziehung linear ist, werden wir r als Maß für die Enge einer Bindung gelten lassen können und diese Gewißheit verschafft uns eben nur das Korrelationsverhältnis.

Der Umstand, daß bei linearen Beziehungen ρ^2 und r^2 miteinander übereinstimmen, hat dazu veranlaßt, ein dem Korrelationsverhältnis entsprechendes Maß für lineare Zusammenhänge zwischen mehreren Veränderlichen zu schaffen. So gibt

$$R^2 = 1 - \frac{\sigma^{2,}}{\sigma^2} \qquad 53)$$

(vgl. 49) an, wie stark eine Größe von andern Größen abhängig ist. R wird als totaler Korrelationskoeffizient bezeichnet. Zwischen der bedingten Streuung σ' und der allgemeinen besteht nun nach 52) folgende Beziehung:

$$\sigma^{2,}_y = \sigma^2_y (1 - r^2_{xy})$$

$$\sigma^{2,}_x = \sigma^2_x (1 - r^2_{xy})$$

ist also y mit x linear verbunden, dann findet die Bedingtheit der Streuung in

$$1 - r^2_{xy}$$

ihren zahlenmäßigen Ausdruck. σ' ist bei völliger Beziehungslosigkeit gleich σ und hat einen untern Grenzwert Null, wenn der Zusammenhang zwischen y und x ein streng linearer ist. Steht nun ein x_1 mit einer Reihe von Werten x_2, x_3 ... x_n in Verbindung, dann ist die Streuung bedingt sowohl durch die Beziehung zwischen x_1 und x_2, als auch durch jene zwischen x_1 und x_3, wenn x_2 = konstans, als auch

durch jene zwischen x_1 und x_4, wenn x_2 und x_4 konstant usw., so daß wir bei n verbundenen Größen entsprechend 35) als bedingte Streuung erhalten

$$\sigma' = \sigma_{1,2..n} = \sigma_1 \sqrt{(1 - r_{12}^2)(1 - r_{13,2}^2)(1 - r_{14,23}^2)\ldots}$$

Für zwei verbundene Größen ist darum

$$R_{1,2}^2 = 1 - (1 - r_{12}^2) = r_{12}^2$$

für drei verbundene Größen:

$$R_{1,23}^2 = 1 - (1 - r_{12}^2)(1 - r_{13,2}^2)$$

$$= 1 - (1 - r_{13}^2)(1 - r_{12,3}^2)$$

$$R_{2,13}^2 = 1 - (1 - r_{12}^2)(1 - r_{23,1}^2)$$

$$R_{3,12}^2 = 1 - (1 - r_{13}^2)(1 - r_{23,1}^2)$$

und entsprechend für n verbundene Größen

$$R_{1,23..n}^2 = 1 - \frac{1}{\sigma_1^2} \cdot \sigma_{1,23..n}^2 \qquad 53\,a)$$

Wie für das Korrelationsverhältnis ergibt sich auch für den quadrierten totalen Korrelationskoeffizienten einen von 0 bis + 1 reichenden Variabilitätsbereich. Während die Möglichkeit $R^2 = 1$ nicht auszuschließen ist, da eine einzige strenglineare Beziehung zwischen der „abhängig Veränderlichen" und einem beliebigen x_k genügt, um die Abweichung von 1 Null werden zu lassen, wird kaum je der Fall eintreten, daß gleichzeitig alle r-Werte und damit auch der totale Korrelationskoeffizient gleich Null sind. Es ist ja nicht Aufgabe der partiellen Korrelationsrechnung, Beziehungen zu suchen oder nachzuweisen, sondern lediglich auf dem Wege der Erfahrung festgestellte Zusammenhänge zu messen bzw. den auf jede einzelne von n Größen entfallenden Anteil am Gesamteinfluß zu ermitteln, den alle an einer als „abhängig veränderlich" ausgesonderten Größe nehmen. Wie der totale Korrelationskoeffizient zeigt, ist dieser Gesamteinfluß, absolut genommen, höchstens gleich 1 zu setzen. Es sei in diesem Zusammenhang an die Ausführungen über die Eigenschaften der partiellen Korrelationskoeffizienten erinnert.

Die Normalkorrelation

Es wurde seinerzeit die korrelative Beziehung in Parallele zum funktionellen Zusammenhang gesetzt. Während bei letzterem — er möge durch $y = f(x)$ ausgedrückt sein — zu jedem Wert von x ein bestimmter Wert von y gehört, ist bei einer Korrelation jedem x eine Gruppe von y-Werten zuzuordnen und umgekehrt. Das Auftreten dieser y- (x-) Werte unterliegt dabei einer Bindung derart, daß zu einem bestimmten $x = X_i$ ($y = Y_i$) ein bestimmtes Streuungsmittel m_y^i (m_x^i) zugehört, um das sich die y_i (x_i) gemäß dem Gaußschen Verteilungsgesetz anordnen. Eine Normalkorrelation ist nun dadurch gekennzeichnet, daß die x- und die y-Werte in ihrer Gesamtheit dem Gaußschen Verteilungsgesetz unterliegen. Die Werte eines x-Kollektivs sind also zwischen den Grenzen a und b nicht gleichmäßig verteilt, sondern sie drängen sich so um einen Mittelwert m_x zusammen, daß zwei Drittel aller x-Werte zwischen $m_x + \sigma_x$ und $m_x - \sigma_x$ liegen. Genau dasselbe ist vom Kollektiv der y-Werte zu sagen. Es ist nun ohne weiteres klar, daß diese Normalkorrelation nie dann vorliegt, wenn man eine Größe als bedingt durch eine andere aufzufassen hat; eine „unabhängig" Veränderliche gibt es bei einem vorgeschriebenen Verteilungsgesetz nicht. Dieses läßt auf ein übergeordnetes Prinzip schließen, dem auch die Anordnung der y-Werte untersteht. Die Normalkorrelation gehört in die Gruppe der symptomatischen Zusammenhänge. Als einigermaßen normalkorrelative Beziehungen seien Beispiele aus Weber: Variation- und Erblichkeits-Statistik angeführt: Korrelation zwischen Körpergröße und Körpergewicht (vgl. das zuletzt durchgerechnete Beispiel), zwischen proportionellem Brustumfang und Körpergröße, Körpergröße und Beckenbreite, Blutdruck und Körpergröße usw. Um nur eines herauszugreifen: Es ist weder die Körpergröße durch die Beckenbreite, noch die Beckenbreite durch die Körpergröße bedingt; es geht nicht an, die eine oder die andere der beiden Veränderlichen als „abhängig" bzw. „unabhängig" zu kennzeichnen; die Relation zwischen beiden ist vielmehr Ausdruck eines übergeordneten Formgesetzes. — Ähnlich wie bei zweiseitigen Beziehungen, so wird man auch bei mehrseitigen dann von Normalkorrelationen zu sprechen haben, wenn alle einander zugeordneten Kollektive dem Gaußschen Verteilungsgesetz unterstehen.

Nun ist zu sagen, daß mit der Kennzeichnung durch „normal" weder ein Werturteil über die eine oder andere Art von Beziehung gefällt werden soll, noch drückt man damit aus, daß die Anwendung der Korrelationsrechnung nicht bei allen Arten nicht-funktioneller

Zusammenhänge zulässig sei. Sie erscheint aber insofern als gerechtfertigt, als in den Normalkorrelationen für die einander zugeordneten Kollektive wahre Mittelwerte zu definieren sind und es darum gegeben erscheint, die Abweichungen von diesen Mittelwerten miteinander in Beziehung zu setzen. Wie aber leicht gezeigt werden kann, hat diese „Verlegung des Koordinatensystems" keinen Einfluß auf die Form der maßgebenden Koeffizienten.

Hat man festgestellt, daß zwischen den zwei Größen y und x eine lineare Beziehung wahrscheinlich ist, dann wird man für die Reihe von n empirisch ermittelten Wertepaaren y_i, x_i die Koeffizienten der Gleichung

$$a \cdot x_i + b \cdot y_i + c = v_i$$

so bestimmen, daß die Summe aller Fehlerquadrate ein Minimum wird. Quadriert man diese Gleichung und summiert über alle i von 1 bis n, dann erhält man

$$a^2 [xx] + 2ab [xy] + b^2 [yy] + 2ac [x] + 2bc [y] + nc^2 = [vv]$$

Differenziert man diese Gleichung partiell nach a, b und c, dann ergibt sich:

$$a [xx] + b [xy] + c [x] = 0$$

$$a [xy] + b [yy] + c [y] = 0$$

$$a [x] + b [y] + nc = 0$$

Die letzte Gleichung gesagt, daß die Konstante c so zu bestimmen ist, daß $[v] = 0$ wird. Setzt man den Wert für c in den ersten beiden Gleichungen ein, dann erhält man weiter:

$$a [xx] - a \frac{[x]^2}{n} + b [xy] - b \frac{[x] [y]}{n} = 0$$

$$a [xy] - a \frac{[x] [y]}{n} + b [yy] - b \frac{[y]^2}{n} = 0$$

Nach 17) und 18) gehen aber diese Gleichungen über in

$$a [\Delta x \Delta x] + b [\Delta x \Delta y] = 0, \qquad \frac{a}{b} = - r \frac{\sigma_y}{\sigma_x}$$

also

$$a [\Delta x \Delta y] + b [\Delta y \Delta y] = 0, \qquad \frac{b}{a} = - r \frac{\sigma_x}{\sigma_y}$$

Man erkennt, daß sich für die Regressionskoeffizienten die schon bekannten Werte ergeben, da man nach Einsetzung des Wertes für c wieder das homogene Gleichungssystem erhält. Wie eng die Bindung zwischen y und x ist, zeigt ebenfalls derselbe Korrelationskoeffizient r; als mittleres Fehlerquadrat erhält man

$$\frac{1}{n}[vv] = \sigma_y^2 (1 - r^2),$$

$$\frac{1}{n}[vv] = \sigma_x^2 (1 - r^2),$$

(vgl. pag. 102), je nachdem man y oder x als abhängige Veränderliche auffaßt. Endlich ergibt sich als Beziehungsgleichung:

$$a \cdot x_i - a \cdot m_x + b \cdot y_i - b \cdot m_y = v_i$$

und hieraus

$$y - m_y = r \frac{\sigma_y}{\sigma_x} (x - m_x), \quad \text{bzw.}$$

$$x - m_x = r \frac{\sigma_x}{\sigma_y} (y - m)_y$$

Wir stellen also fest, daß die Verlegung des Koordinatensystems durch den Punkt (m_x, m_y) die zwangsläufige Folge der Bedingung

$$[vv] = \text{Min.}$$

ist und nichts mit der für die Normalkorrelation geforderten Werteverteilung in den einander zugeordneten Kollektiven zu tun hat. Vom Standpunkt der Korrelationsrechnung aus hat man nicht zu fordern, daß die x-Werte einem bestimmten Verteilungsgesetz unterliegen und die Häufigkeit der y-Werte kann ebenfalls für jedes x beliebig sein. Dagegen ist, wie bereits erwähnt wurde, anzunehmen, daß die zu einem bestimmten $x = X_i$ gehörenden y_i-Werte um deren Mittelwert sich gemäß dem Gaußschen Verteilungsgesetz anordnen werden und umgekehrt. Man wird darum auch die Beziehungen zwischen diesen Mittelwerten m_x^i und m_y^i, die der Einfachheit halber mit X_i und Y_i bezeichnet seien, untersuchen, so daß die Beziehungsgleichung übergeht in

$$Y - m_y = r \frac{\sigma_y}{\sigma_x} (X - m_x)$$

$$X - m_x = r \frac{\sigma_x}{\sigma_y} (Y - m_y)$$

Über die Gewichte der X_i und der Y_i ist das Notwendige bereits pag. 57—60 gesagt worden.

Alle Ausführungen, die hier über zwei verbundene Größen gemacht wurden, gelten nun ohne Einschränkung auch im Falle einer mehrseitigen Verbundenheit. Besteht zum Beispiel zwischen drei Größen folgende lineare Beziehung:

$$a \cdot x_{1i} + b_{2i} \cdot c \cdot x_{3i} + d = v_i$$

und ermittelt man die Koeffizienten a, b, c und d so, daß wiederum [vv] = Min. wird, dann ergibt sich durch partielle Differentiation von [vv] nach allen 4 Konstanten vorerst:

$$a\,[x_1\,x_1] + b\,[x_1\,x_2] + c\,[x_1\,x_3] + d\,[x_1] = 0$$
$$a\,[x_1\,x_2] + b\,[x_2\,x_2] + c\,[x_2\,x_3] + d\,[x_2] = 0$$
$$a\,[x_1\,x_3] + b\,[x_2\,x_3] + c\,[x_3\,x_3] + d\,[x_3] = 0$$
$$a\,[x_1] + b\,[x_2] + c\,[x_3] + n \cdot d = 0$$

Aus der letzten Gleichung entnehmen wir

$$d = -\frac{1}{n}(a\,[x_1] + b\,[x_2] + c\,[x_3])$$

Setzen wir diesen Wert von d in den drei ersten Gleichungen ein, dann erhalten wir in

$$a\,[\Delta x_1\,\Delta x_1] + b\,[\Delta x_1\,\Delta x_2] + c\,[\Delta x_1\,\Delta x_3] = 0$$
$$a\,[\Delta x_1\,\Delta x_2] + b\,[\Delta x_2\,\Delta x_2] + c\,[\Delta x_2\,\Delta x_3] = 0$$
$$a\,[\Delta x_1\,\Delta x_3] + b\,[\Delta x_2\,\Delta x_3] + c\,[\Delta x_3\,\Delta x_3] = 0$$

wieder das bekannte homogene Gleichungssystem, das uns die partiellen b-, r- und σ-Werte liefert. Es erübrigt sich daher, das Verfahren bei 4 (n) Größen zu wiederholen. Auch in der partiellen Korrelationsrechnung ist die besondere Verteilung der Werte in den einander zugeordneten Kollektiven ohne Bedeutung und damit eine Vorzugsstellung der Normalkorrelation abzulehnen.

Abschließend können wir sagen, daß es im weiten Bereich möglicher korrelativer Beziehungen eine spezielle Gruppe gibt, die das Ergebnis bzw. der sinnenfällige Ausdruck eines übergeordneten Formungsprinzipes ist. Es ist eine Gruppe von symptomatischen Korrelationen, die wegen der Anordnung der Werte nach dem Gaußschen Verteilungsgesetz in den Kollektiven als Normalkorrelationen bezeichnet werden.

Die praktische Durchführung der partiellen Korrelationsrechnung

Gegenstand der Untersuchung sind die Beziehungen zwischen

a. dem schweizerischen Großhandelsindex,

b. der Kurs-Diskrepanz, bezogen auf das englische Pfund,

c. der schweizerischen Ausfuhr nach Großbritannien.

Über den Begriff der Kurs-Diskrepanz sei noch folgendes vorausgeschickt:

Wird der Wechselkurs zwischen zwei Staaten mit verschiedenen Währungseinheiten durch k ausgedrückt so, daß eine Geldeinheit des einen gleich k Geldeinheiten des andern Staates ist, dann stellen sich die Preise einer bestimmten Ware p_i und p_i' in beiden Währungsgebieten in einer Währungseinheit ausgedrückt folgendermaßen dar:

$$p_i \text{ ./. } kp_i', \quad \text{oder} \quad \frac{1}{k}p_i \text{ ./. } p_i'$$

Für den Käufer dieser Ware ist es nun gleichgültig, woher er sie bezieht, sobald

$$p_i = kp_i'$$

gesetzt werden kann. Man spricht in diesem Fall von Kaufkraft-Parität der beiden Währungen in bezug auf p_i. Nun ist aber für die Wirtschaft eines Landes nicht die Preisbeziehung für eine Ware wichtig, sondern jene für den Preisdurchschnitt aller Waren. Bei allgemeiner Kaufkraft-Parität müßte also entsprechend die Beziehung gelten:

$$\frac{[m_i p_i]}{[m_i]} = k \frac{[m_i p_i']}{[m_i]}$$

Ersetzen wir endlich die Durchschnittswerte durch auf ein bestimmtes Basisjahr bezogene Relativwerte (Indizes), dann erhalten wir endlich:

$$I = ck \cdot I' \quad \text{bei Kaufkraft-Parität.}$$

c ist dabei eine von der Wahl des Basisjahres abhängige Konstante. — Es braucht wohl nicht besonders betont zu werden, daß eine solche Parität einen idealen Gleichgewichtszustand darstellt, der nur bei vollständiger Freizügigkeit im Handelsverkehr über die Grenzen an-

nähernd erreicht werden kann. Da aber der Index über die Änderung der allgemeinen Preislage zu orientieren hat, kann der in der oben angeführten Beziehung auftretende Proportionalitätsfaktor c auch die Veränderung einer allfälligen Inkongruenz zwischen Parität und den tatsächlichen Preisverhältnissen zugunsten der einen oder der andern Währung aufzeigen: Ein zunehmender c-Wert deutet eine Verbilligung der Fremdpreise und damit eine Erschwerung der eigenen Ausfuhr an usw. Lassen wir c unverändert, dann können wir leicht den Soll-Kurs oder den Soll-Index berechnen, der die Handelsbeziehungen zwischen zwei Staaten den gleichen Bedingungen unterstellen würde, wie im Basisjahr. Man vergleiche in diesem Zusammenhang die Ausführungen im Mémorandum sur les Monnaies (Société des Nations) 1924 Genf, pag. 40/41, und die instruktiven Tableaux 0 und VIII.

Im vorliegenden Beispiel wurde der Kurs für das Bezugsjahr 1929 gleich 100 gesetzt; in diesem Fall ist der Quotient $\frac{I_s}{I}$ direkt ein Maß des Sollkurses. Dieser ist für 1929 ebenfalls auf 100 gebracht und damit dem wirklichen Kurs vergleichbar gemacht worden. Als Kursdiskrepanz $D_k = k' - k$ wird die Differenz zwischen Sollkurs und wirklichem Kurs bezeichnet. Endlich ist in der nachfolgenden Tabelle die schweizerische Ausfuhr nach Großbritannien (A) für 1929 = 100 gesetzt. I_s = schweizerischer Großhandelsindex; I = Index für Großbritannien (1929 = 100). Das Zahlenmaterial wurde dem Statistischen Jahrbuch der Schweiz entnommen.

Tabelle a.

Jahr	I_s	I	k'	k	A	D_k
1929	100	100	100	100	100	0
30	89	87	102	99,5	91,5	2,5
31	78	77	101	93	82	8
32	68	74,5	91	72	30	19
33	64,5	74,5	87	68	31	19
34	64	77	83	62	29	21
35	64	77,5	83	60	27	23
36	68	82,5	83	66*	34	17

* interpol.

Dieser Tabelle entnehmen wir, wie bereits erwähnt, die Werte I_s, D_k und A (x_1, x_2, x_3) zur Durchführung der Korrelationsrechnung.

Tabelle b.

n	Δx_1 +	Δx_1 —	Δx_2 +	Δx_2 —	Δx_3 +	Δx_3 —	Δx_1^2	Δx_2^2	Δx_3^2
1	20			10	30		400	100	900
2	9			7,5	21,5		81	56	462
3		2		2	12		4	4	144
4		12	9			40	144	81	1600
5		15,5	9			39	240	81	1521
6		16	11			41	256	121	1681
7		16	13			43	256	169	1849
8		12	7			36	144	49	1296
	29	73,5	49	19,5	63,5	199	1525	661	9453
		—44,5	+29,5			—135,5			

$$m_{x_1} = 80 - \frac{44,5}{8} = 74,44 \qquad [\Delta x_1^2] = 1525 - \frac{1980}{8} = 1277,5$$

$$m_{x_2} = 10 + \frac{29,5}{8} = 13,69 \qquad [\Delta x_2^2] = 661 - \frac{871}{8} = 552,1$$

$$m_{x_3} = 70 - \frac{135,5}{8} = 53,06 \qquad [\Delta x_3^2] = 9453 - \frac{18370}{8} = 7157$$

n	$\Delta x_1 \Delta x_2$ +	$\Delta x_1 \Delta x_2$ —	$\Delta x_1 \Delta x_3$ +	$\Delta x_1 \Delta x_3$ —	$\Delta x_2 \Delta x_3$ +	$\Delta x_2 \Delta x_3$ —
1		200	600			300
2		67,5	193,5			161,3
3	4			24		24
4		108	480			360
5		139,5	605			351
6		176	656			451
7		208	688			559
8		84	432			252
	4	983,0	3654,5	24	—	2458,3
		—979,0	+3630,5			—2458,3

$$[\Delta x_1 \Delta x_2] = -\ 979 + \frac{1313}{8} = -\ 815$$

$$[\Delta x_1 \Delta x_3] = +\ 3631 - \frac{6030}{8} = +\ 2877$$

$$[\Delta x_2 \Delta x_3] = -\ 2458 + \frac{3997}{8} = -\ 1958$$

$$r_{12} = -\ 0{,}968 \pm 0{,}022$$

$$r_{12} = +\ 0{,}952 \pm 0{,}033$$

$$r_{23} = -\ 0{,}984 \pm 0{,}011$$

$$\sigma_1 = \pm\ 12{,}64 \qquad \sigma_{12} = \pm\ 3{,}17 \qquad \sigma_{13} = \pm\ 3{,}87$$

$$\sigma_2 = \pm\ 8{,}32 \qquad \sigma_{21} = \pm\ 2{,}09 \qquad \sigma_{23} = \pm\ 1{,}48$$

$$\sigma_3 = \pm\ 29{,}91 \qquad \sigma_{31} = \pm\ 9{,}16 \qquad \sigma_{32} = \pm\ 5{,}33$$

Aus den Rechnungsergebnissen ist zu schließen, daß zwischen Durchschnittspreisstand, Kursdiskrepanz und Ausfuhr enge zweiseitige Zusammenhänge bestehen. Sinkt der Durchschnittspreisstand, dann nimmt die Kursdiskrepanz zu, während die Ausfuhr ebenfalls eine Abnahme zeigt. Am linearen Charakter der Beziehungen ist kaum zu zweifeln, da es fast unmöglich ist, daß die entsprechenden Korrelationsverhältnisse außerhalb der Fehlergrenzen der r-Werte bleiben. Da einem Wert einer Größe nur je ein Wert der andern beiden Größen entspricht, können die ρ^2 nicht berechnet werden.

Wir finden weiter:

$$r_{12,3} = \frac{-\ 0{,}968 + 0{,}952 \cdot 0{,}984}{0{,}306 \cdot 0{,}178} = -\ 0{,}569 \pm 0{,}240$$

$$r_{13,2} = \frac{+\ 0{,}952 - 0{,}968 \cdot 0{,}984}{0{,}251 \cdot 0{,}178} = 0$$

$$r_{23,1} = \frac{-\ 0{,}984 + 0{,}968 \cdot 0{,}952}{0{,}251 \cdot 0{,}306} = -\ 0{,}807 \pm 0{,}124$$

Dem Bild, das wir uns von dem Zusammenwirken der drei Größen I_s, D_k und A auf Grund der angestellten Überlegungen zu machen haben, werden die partiellen Korrelationskoeffizienten viel eher gerecht, als die zweiseitigen. Es durfte bei den hohen Werten dieser letzteren erwartet werden, daß die Koeffizienten r dritter Ordnung kleiner ausfallen würden. Nicht anzunehmen war dagegen, daß die

enge positive Bindung zwischen Preisstand und Ausfuhr zum Beispiel vollständig verschwinden könnte bei Konstanthaltung der Kursdiskrepanz. Wir sehen, daß die partielle Korrelationsrechnung die Möglichkeit bietet, den Sondereinfluß jeder einzelnen Größe auf jede einzelne von vielen zu bestimmen. Die obigen Ergebnisse lassen folgende Schlüsse zu:

Der Warenverkehr über die Grenze erfährt keine Hemmungen oder wenigstens keine Veränderung, wenn die Kursdiskrepanz 0 ist, bzw. konstant bleibt; wenn also die Preisspiegel diesseits und jenseits der Grenze gleichsinnige proportionale Bewegungen aufwärts oder abwärts ausführen;

wenn bei festem Inlandspreisstand einem sinkenden (steigenden) Kurs ein steigender (sinkender) Auslandspreisstand parallel geht;

wenn bei festem Auslandspreisstand Kurs und Inlandspreisstand gleichsinnige Bewegungen zeigen.

Wenn die Warenausfuhr konstant ist, dann zeigen Inlandspreisstand und Kursdiskrepanz eine schwache negative Korrelation; jede Preissenkung muß sich also in einem Druck auf den Kurs auswirken; jede Preiserhöhung hebt auch den Kurs.

Endlich zeigt der hohe Wert von $r_{23,1}$, daß die Abnahme der Warenausfuhr zur Hauptsache auf die steigende Kursdiskrepanz zurückzuführen ist. Dabei ist gleichgültig, auf welche Weise die Veränderungen von D_k zustande kommen. Bleibt I_s nicht fest, dann ist die Beziehung zwischen D_k und A nahezu eine streng lineare ($r_{23} = -0{,}984$). Bleibt der Inlandspreisstand unverändert, dann muß zwangsläufig die Bindung zwischen Kursdiskrepanz und Ausfuhr lockerer werden, weil ja auch I_s mit D_k in Beziehung steht. Im vorliegenden Fall wird also der Zusammenhang wohl der sein, daß primär der sinkende Inlandspreisstand die Ausbildung einer Kursdiskrepanz zur Folge hatte und diese hierauf eine Schrumpfung der Ausfuhr veranlaßte.

Aus den partiellen Korrelationskoeffizienten ergeben sich nun folgende b-Werte:

$$b_{12,3} = -0{,}569\frac{3{,}87}{1{,}48} = -1{,}49 \pm 0{,}63$$

$$b_{21,3} = -0{,}218 \pm 0{,}092$$

$$b_{13,2} = b_{31,2} = 0$$

$$b_{23,1} = -0{,}184 \pm 0{,}028$$

$$b_{32,1} = -3{,}54 \pm 0{,}54$$

Die Beziehungsgleichungen lauten somit:

$$x_1 - m_{x_1} = -1{,}49\,(x_2 - m_{x_2})$$

$$x_2 - m_{x_2} = -0{,}218\,(x_1 - m_{x_1}) - 0{,}184\,(x_3 - m_{x_3})$$

$$x_3 - m_{x_3} = -3{,}54\,(x_2 - m_{x_2})$$

oder kürzer

$$\Delta x_1 = -1{,}49\,\Delta x_2$$

$$\Delta x_2 = -0{,}218\,\Delta x_1 - 0{,}184\,\Delta x_3$$

$$\Delta x_3 = -3{,}54\,\Delta x_2$$

Diese Beziehungsgleichungen zeigen mit aller Deutlichkeit, daß eine Verbindung zwischen Inlandspreisstand und Ausfuhr nur über die Kursdiskrepanz herzustellen ist. Direkte Zusammenhänge bestehen zwischen I_s und D_k; ferner zwischen A und D_k. Diese müssen sich auch aus den zweiseitigen r-Werten ergeben. Man findet:

$$\Delta x_1 = -0{,}968\frac{12{,}64}{8{,}32}\,\Delta x_2 = -1{,}47\,\Delta x_2$$

$$\Delta x_3 = -0{,}984\frac{29{,}91}{8{,}32}\,\Delta x_2 = -3{,}54\,\Delta x_2$$

Die zweiseitigen Regressionskoeffizienten unterscheiden sich kaum von den partiellen.

Da nach der 3. Gleichung A durch D_k bedingt ist, eliminieren wir in der 2. Gleichung Δx_3. Wir erhalten auf diese Weise:

$$\Delta x_2 = -0{,}626\,\Delta x_1$$

Mit Hilfe von r_{12} erhielte man

$$\Delta x_2 = -0{,}637\,\Delta x_1$$

Der Unterschied der beiden Regressionskoeffizienten beträgt 0,011; er ist also kleiner als der Fehler von $b_{21,3}$, so daß man die partielle Beziehungsgleichung, welche angibt, wie D_k durch I_s und A bedingt ist, durch die sachlich richtigere zweiseitige, die D_k nur durch I_s ausdrückt, ersetzen kann. Das Ergebnis wäre also folgendes:

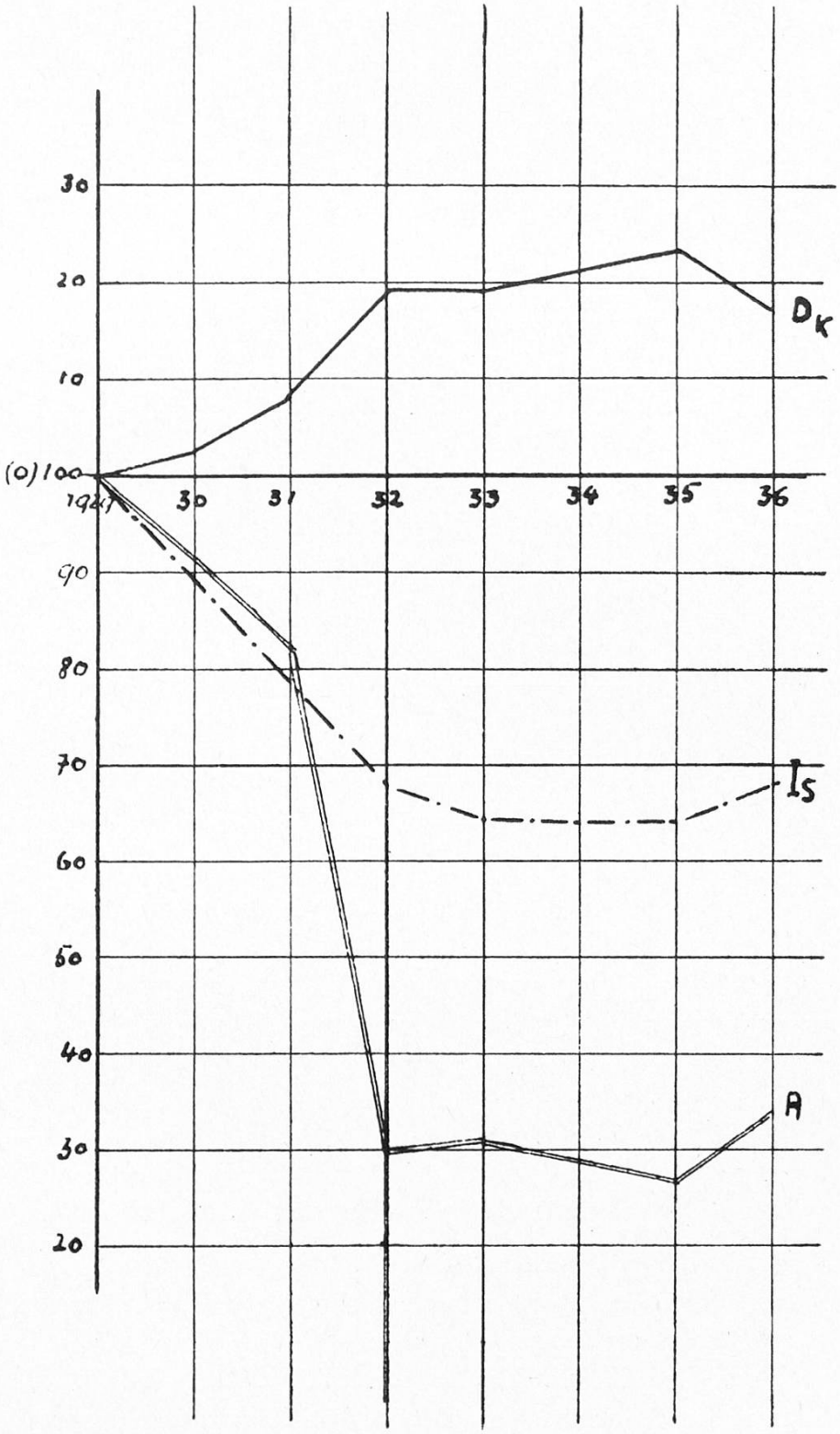

Fig. 19

Es ist die Kursdiskrepanz durch den Inlandspreisstand und die Warenausfuhr durch die Kursdiskrepanz bedingt und es entspricht

einer Zunahme der Diskrepanz um eine Einheit eine Abnahme des Durchschnittspreisstandes um 1,49 Einheiten;

einer Zunahme des Durchschnittspreisstandes um eine Einheit eine Abnahme der Diskrepanz um 0,626 Einheiten;

einer Zunahme der Diskrepanz um eine Einheit eine Abnahme der Ausfuhr um 3,54 Einheiten.

Es bleibt noch übrig, die partiellen Streuungen 3. Ordnung und daraus die totalen Korrelationskoeffizienten abzuleiten. Man erhält:

$$\sigma_{1,23} = \pm 12{,}64 \cdot 0{,}251 = \pm 3{,}17$$

$$\sigma_{2,13} = \pm\ 8{,}32 \cdot 0{,}146 = \pm 1{,}215$$

$$\sigma_{3,12} = \pm 29{,}91 \cdot 0{,}178 = \pm 5{,}33$$

und hieraus

$$R_{1,23} = \sqrt{1 - 0{,}251^2} = 0{,}968$$

$$R_{2,13} = \sqrt{1 - 0{,}146^2} = 0{,}9893$$

$$R_{3,12} = \sqrt{1 - 0{,}178^2} = 0{,}984$$

Da Is nur an D_k und A nur an D_k gebunden ist, entsprechen die totalen Korrelationskoeffizienten $R_{1,23}$ und $R_{3,12}$ den Werten r_{12} und r_{23}. Der hohe Wert des Koeffizienten $R_{2,13}$ zeigt an, daß die Kursdiskrepanz praktisch fast völlig durch Preisbewegung und Ausfuhr, und zwar, wie sich aus den Beziehungsgleichungen ergibt, hauptsächlich durch erstere bedingt ist.

Verzeichnis der verwendeten Literatur

Felix M. Exner: Über die Korrelationsmethode Fischer. Jena 1913.

Aug. Pütter, Dr.: Die Auswertung zahlenmäßiger Beobachtungen in der Biologie. Gruyter, Berlin 1929.

F. Bauer, Dr.: Korrelationsrechnung. Math.-Phys. Bibl., Bd. 75.

Erna Weber, Dr.: Variations- und Erblichkeits-Statistik. Lehmann, München 1935.

Bureau International du Travail: La Crise de Chômage 1920—23. Genf 1924.

Société des Nations: Mémorandum sur les Monnaies 1913—23. Genf 1924.

Eidgenössisches Statistisches Amt: Statistisches Jahrbuch der Schweiz.