VEREINFACHTE METHODEN ZUR BERECHNUNG DES KORRELATIONSKOEFFIZIENTEN BEI NORMALER KORRELATION

VON

ALFRED KÄRSNA

TARTU 1935

VEREINFACHTE METHODEN ZUR BERECHNUNG DES KORRELATIONSKOEFFIZIENTEN BEI NORMALER KORRELATION

VON

ALFRED KÄRSNA

TARTU 1935

Acta et Commentationes Universitatis Tartuensis (Dorpatensis) A XXIX. 2.

K. Mattiesens Buchdruckerei Ant.-Ges., Tartu 1935.

Die erste Aufgabe der Korrelationsrechnung bildet die Bestimmung des Korrelationskoeffizienten. Diese Aufgabe ist theoretisch einfach, ihre praktische Ausführung jedoch sehr zeitraubend und mit verschiedenen Schwierigkeiten verbunden.

Zur Berechnung des Korrelationskoeffizienten auf Grund von 1000 Wertepaaren sind, je nach der Größe der Zahlen, 3—6 Stunden erforderlich. Oft ist auch eine genaue Ermittelung des Korrelationsfaktors gar nicht notwendig und genügt für praktische Zwecke schon eine annähernde Schätzung der Güte der Korrelation. Deshalb wäre eine vereinfachte numerische und graphische Methode zur Berechnung des Korrelationskoeffizienten von großem Werte. Von den schon vorhandenen Methoden ist aber keine einzige vollständig befriedigend, denn im besten Falle wird die Dauer der Rechenarbeiten um das zehnfache verkürzt, d. h. bei 1000 Wertepaaren auf $^1/_2$ bis 1 Stunde reduziert, was in manchen Fällen immerhin zu viel ist. Außerdem verlangen einige Methoden auch eine große Geläufigkeit im mathematischen Rechnen (z. B. die Pearson'sche Methode des Punktezählens).

Die zwei folgenden Methoden, von denen die eine auf der Messung des Korrelationsfeldes, die andere auf dem Abzählen der Punkte beruht, sind von den erwähnten Mängeln frei.

Beide Methoden sind bei normaler Korrelation vollständig genau. Bei nicht normaler Korrelation ist der erhaltene Koeffizient mit einem gewissen Fehler behaftet, jedoch genügt uns in diesem Falle auch der nach der allgemein gebräuchlichen Formel berechnete Koeffizient nicht ganz, und es ist außer seiner Bestimmung noch die Berechnung der Parameter höheren Grades erforderlich. Die Regressionsgleichung nimmt alsdann eine ganz andere Gestalt an. Sollte das Bestimmen derselben erforderlich sein, so würde die Berechnung der Parameter höheren Grades viel Zeit in Anspruch nehmen, wobei die mühevolle Bestimmung des ersten Korrelationskoeffizienten schon nicht mehr in Betracht käme.

Falls die Regressionskurve eine Gerade darstellt, sind die beiden zu beschreibenden Methoden sehr brauchbar.

In beiden Fällen muß zuerst das Korrelationsfeld gezeichnet werden, um feststellen zu können, ob überhaupt eine Korrelation vorhanden ist.

Die sich auf Messung des Feldes stützende Methode erfordert zur Durchführung der Rechnung bei einer beliebigen Zahl von Punkten 2—3 Minuten, während die andere Methode soviel Zeit beansprucht, als für das Zählen der Punkte notwendig ist (bei 1000 Punkten z. B. 5 Minuten). Bei einer kleineren Zahl von Punkten (unter 30—50) ist es ratsam die zweite Methode zu gebrauchen, die in solchen Fällen genauere Resultate liefert. Bei 30—50 und mehr Punkten sind beide Methoden gleich gut. Bei einer größeren Zahl von Punkten (300—500) ist dagegen die erste Methode vorteilhafter, da sie weniger Zeit in Anspruch nimmt.

Wenn wir im Korrelationsfelde $(x;\ y)$ (Fig. 1) die Dichte auf der III. Achse abtragen (z), so erhalten wir die sogenannte Korrelationsfläche $z = f(x;\ y)$, wobei $z dx dy$ die Wahrscheinlichkeit bedeutet, daß der Punkt ins Gebiet ds fällt (x bis $x + dx$, y bis $y + dy$).

Im Falle einer normalen Verteilung ergeben alle vertikalen Schnitte $x =$ konst. oder $y =$ konst. die Gauss'sche Kurve, wobei das Zentrum derselben auf der Regressionsgeraden $y = mx$ liegt.

$m = r \cdot \frac{\sigma y}{\sigma x}$, wenn wir $x =$ konst., und

$m = \frac{1}{r} \cdot \frac{\sigma y}{\sigma x}$, wenn wir $y =$ konst. setzen

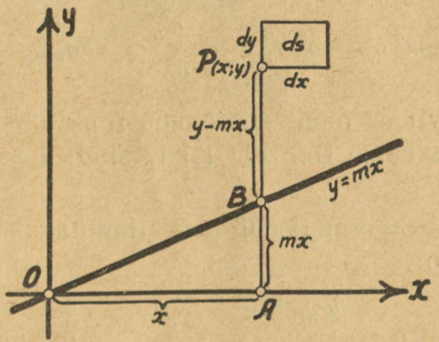

Fig. 1.

(r ist der Korrelationskoeffizient, σ_x, σ_y sind die mittleren Abweichungen der x- und y-Werte). Also:

$$z_A = z_O \cdot e^{-\frac{x^2}{2\sigma_x^2}}$$

$$z_A = z_B \cdot e^{-\frac{m^2 x^2}{2\sigma_y^2}}$$

$$z_P = z_B \cdot e^{-\frac{y^2 - 2mxy + m^2 x^2}{2\sigma_y^2}}$$

$$\frac{z_P}{z_A} = e^{-\frac{y^2 - 2mxy}{2\sigma_y^2}}$$

$$z_P = z_A \cdot \frac{z_P}{z_A} = z_O \cdot e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2 - 2mxy}{2\sigma_y^2}}$$

Betrachten wir die Linien gleicher Dichte, so ist $z_P =$ konst.

$$z_0 \cdot e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2 - 2mxy}{2\sigma_y^2}} = \underbrace{z_0 \cdot e^{-k}}_{\text{konst.}}.$$

Die Gleichung der genannten Linien ist

$$-\frac{x^2}{2\sigma_x^2} - \frac{y^2 - 2\,mxy}{2\sigma_y^2} = -k, \text{ oder}$$

$$\sigma_y^2 x^2 + \sigma_x^2 y^2 - 2\,m\sigma_x^2 xy = \underbrace{2\,k\,\sigma_x^2\,\sigma_y^2}_{\text{konst.}}.$$

Diese ergeben eine Kurve 2-ten Grades $ax^2 + by^2 + 2cxy = d$. Betrachten wir die Determinante

$$\Delta = \begin{vmatrix} a & c \\ c & b \end{vmatrix} = ab - c^2.$$

$$\Delta = \sigma_y^2\,\sigma_x^2 - m^2\,\sigma_x^4 = \sigma_y^2\,\sigma_x^2 - r^2\,\frac{\sigma_y^2}{\sigma_x^2}\,\sigma_x^4 = \sigma_y^2\,\sigma_x^2\,(1 - r^2) > 0,$$

da $r < 1$, falls wir es nicht mit einem funktionalen Zusammenhang zu tun haben. Da $\Delta > 1$ ist, sind die Linien gleicher Dichte Ellipsen.

Wenn das Steigungsmaß der Hauptachsen der Ellipse $\mu = \tan \alpha$ ist, so ist

$$\tan\,2\,\alpha = \frac{2\,m\sigma_x^2}{\sigma_x^2 - \sigma_y^2} = \frac{2\,rk}{1 - k^2}, \quad \text{wo } k = \frac{\sigma_y}{\sigma_x},$$

$$\text{oder } \frac{2\,\tan\alpha}{1 - \tan^2\alpha} = \frac{2\,rk}{1 - k^2}, \quad \text{wo } r = \frac{\mu\,(1 - k^2)}{k\,(1 - \mu^2)} \text{ ist.}$$

Die erhaltene Formel gibt die Möglichkeit nach dem gezeichneten Felde r zu berechnen. Für eine bestimmte Kurve der gleichen Dichte ist (Fig. 2) $\overline{OA} = c\sigma_y$ und $\overline{OB} = c\sigma_x$, wo c konstant ist, und $\frac{\overline{OA}}{\overline{OB}} = \frac{\sigma_y}{\sigma_x} = k$. Da dieses für jede Kurve gleicher Dichte gilt, so gilt es auch für die Kurve, außerhalb welcher keine Punkte mehr vorhanden sind.

Die Formel hat viele Nachteile:

1. μ und k sind praktisch annähernd gleich (wenn $\mu < 1$, so ist $\mu < k$; wenn $\mu > 1$, so ist $\mu > k$ und wenn $\mu = 1$, dann ist $\mu = k$).

2. μ und k unterscheiden sich nicht viel von 1, somit sind $1 - \mu^2$ und $1 - k^2$ kleine Zahlen mit großen relativen Fehlern.

3. Die graphische Ermittelung von k ist schwierig, denn es ist schwer auf der x-Achse und der y-Achse diejenigen Stellen zu finden, wo die Dichte gleich Null wird.

4. Ist $\mu = 1 = k$, so ist r unbestimmt.

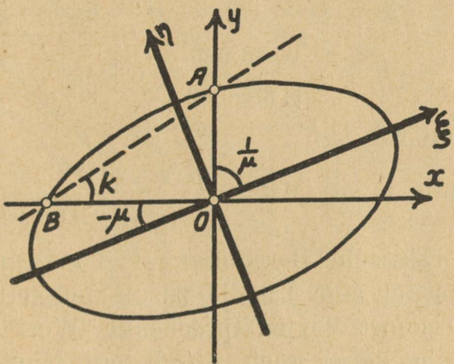

Fig. 2.

Da die Größe k mit den Achsen der Ellipse im Zusammenhang steht, so bringen wir deren Länge in die Gleichung hinein, da diese Achsen genauer bestimmbar sind.

Die Gleichung der Geraden OA ist $\eta = \frac{1}{\mu} \cdot \xi$ und diejenige der Geraden OB $\eta = -\mu\,\xi$. Indem wir das System dieser beiden Gleichungen und der Ellipsengleichung $b^2\xi^2 + a^2\eta^2 = a^2b^2$ auflösen, erhalten wir die Koordinaten der Punkte A und B:

$$\xi_A^2 = \frac{a^2 b^2 \mu^2}{a^2 + b^2 \mu^2},\ \eta_A^2 = \frac{a^2 b^2}{a^2 + b^2 \mu^2},\ \xi_B^2 = \frac{a^2 b^2}{a^2 \mu^2 + b^2} \text{ und } \eta_B^2 = \frac{a^2 b^2 \mu^2}{a^2\mu^2 + b^2};$$

$$\overline{OA}^2 = \xi_A^2 + \eta_A^2 \text{ und } \overline{OB}^2 = \xi_B^2 + \eta_B^2;$$

$$k^2 = \frac{\overline{OA}^2}{\overline{OB}^2} = \frac{a^2\mu^2 + b^2}{a^2 + b^2\mu^2};$$

wir setzen die erhaltenen Werte in den Ausdruck für r ein und erhalten

$$r = \frac{\mu\,(a^2 - b^2)}{\sqrt{(a^2 + b^2\mu^2)\,(a^2\mu^2 + b^2)}}.$$

Die Berechnungen erfolgen folgendermaßen: aus der Zeichnung bestimmen wir μ, a, b und führen die Berechnungen nach folgendem Schema durch:

$$\begin{array}{c|c|l}
\mu & \mu^2 & \\
a & a^2 & \mu^2 a^2 \\
b & b^2 & \mu^2 b^2 \\
\hline
\multicolumn{2}{c}{a^2 - b^2} &
\end{array}
\qquad \text{Beispiel:} \quad
\begin{array}{r|r|r}
\mu = 0\cdot 83 & 0\cdot 688 & \\
a = \ \ 17 & 289 & 199 \\
b = \ \ \ 8 & 64 & 44 \\
\hline
 & 225 &
\end{array}$$

$$r = \frac{0\cdot 83.225}{\sqrt{333.263}} = 0\cdot 63.$$

Bezeichnen wir

$$\frac{a}{b} \text{ mit } \nu, \text{ so ist } r = \frac{\mu\,(\nu^2 - 1)}{\sqrt{(\nu^2 + \mu^2)\,(\nu^2\mu^2 + 1)}}.$$

Für die praktische Berechnung von r kann man auf Grund der letzten Formel eine Tabelle zusammenstellen, so daß für jedes μ und ν sofort der entsprechende Wert von r abzulesen ist. Die Tabelle kann auch durch ein Nomogramm ersetzt werden.

Dividieren wir den Zähler und den Nenner durch μ^2 und bezeichnen wir den reziproken Wert von μ durch $\frac{1}{\mu} = \mu'$, so erhalten wir:

$$r = \frac{\frac{1}{\mu}(\nu^2 - 1)}{\sqrt{\left(\nu^2 + \frac{1}{\mu^2}\right)\left(\frac{\nu^2}{\mu^2} + 1\right)}} = \frac{\mu'^2\,(\nu^2 - 1)}{\sqrt{(\nu^2 + \mu'^2)\,(\nu^2\mu'^2 + 1)}};$$

somit ergibt jeder Wert von μ und sein reziproker Wert dasselbe r. Daher genügt es das Nomogramm nur für $\mu \leqslant 1$ zu zeichnen. Bei $\mu > 1$ kann man dann den reziproken Wert benutzen. Es ist vorteilhaft die μ-Achse logarithmisch zu wählen, denn in diesem Falle würden die μ- und auch die $\frac{1}{\mu}$-Skalen zusammenfallen und immer denselben relativen Fehler liefern. Man kann auch die ν-Achse logarithmisch wählen, denn die relative Veränderlichkeit von ν ist wesentlicher als die absolute. So ruft eine Änderung des ν von 1·1 bis 1·6 (um 0·5) eine größere Änderung des r hervor, als die Veränderung von 8·5 bis 9·0. Auch erhalten die Kurven in der logarithmischen xy-Ebene eine kleinere Krümmung, während sie im gewöhnlichen Koordinatensystem stark gebogen sind.

Solch ein Nomogramm stellt Fig. 3 dar. Längs der μ-Achse (horizontale Achse) sind zu den μ-Werten noch die *arctan* μ-Werte in Graden hinzugeschrieben, da der Winkel

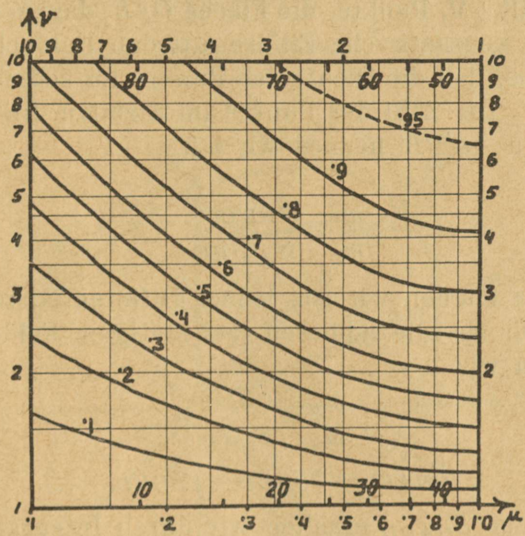

Fig. 3.

auch nach Graden gemessen werden kann. Die Werte $\frac{1}{\mu}$ sind am oberen Rande des Nomogramms eingetragen. Auf der Vertikalachse findet man die Werte von ν. Die Schnittpunkte der Parallelen zu den Achsen, durch die entsprechenden ν und μ gezogen, ergeben den Wert von r, den man nötigenfalls dem Augenmaß nach interpolieren kann.

Im folgenden betrachten wir die Methode des Abzählens der Punkte in den Quadranten. Zur Vereinfachung der Berechnungen wählen wir $\sigma_x = \sigma_y$, denn es ist bekannt, daß

1. der Korrelationskoeffizient von σ_x und σ_y nicht abhängt, und daß

2. die Zahl der Punkte von σ_x und σ_y nicht abhängt.

Bemerkung: Die Achsen sind durch die arithmetischen Mittelwerte von x und y gezogen und die Anzahl der Punkte entspricht den Bezeichnungen in Fig. 6. Die Fläche OBC (Fig. 4) enthält $\frac{1}{4} M_1$ Punkte, die Fläche OAB dagegen $\frac{1}{4} M_2$. Wir nehmen eine konzentrische Ellipse mit den Halbachsen $a + \Delta a$ und $b + \Delta b$ so, daß die Dichte der Punkte im schmalen Streifen konstant ist. Die Zahl der Punkte im Streifen $BB'CC'$ nennen wir m_1, die in $AA'BB'$ nennen wir m_2.

So ist

$$\frac{m_1}{m_2} = \frac{S_{BB'CC'}}{S_{AA'BB'}}.$$

Dasselbe können wir von jedem Differentialstreifen sagen. Und wenn wir sie alle addieren (bei stetigem Felde: integrieren), erhalten wir

$$\frac{M_1}{M_2} = \frac{S_{OBC}}{S_{OAB}}$$

$$S_{OAB} = S_{OABD} - S_{OBD}.$$

Die Fläche S_{OABD} erhalten wir durch Integration:

$$S_{OABD} = \frac{b}{a} \int\limits_0^{\xi_0} \sqrt{a^2 - \xi^2}\, d\xi =$$

$$= \frac{b}{a} \left\{ \xi \sqrt{a^2 - \xi^2} \Big|_0^{\xi_0} - \int\limits_0^{\xi_0} \sqrt{a^2 - \xi^2}\, d\xi + a^2 \arcsin \frac{\xi}{a} \Big|_0^{\xi_0} \right\}, \text{ oder}$$

$2\, S_{OABD} = \dfrac{a^2 b^2}{a^2 + b^2} + ab \arcsin \dfrac{b}{\sqrt{a^2 + b^2}}$, wenn wir statt ξ_0 seinen Wert $\xi_0 = \dfrac{ab}{\sqrt{a^2 + b^2}}$ einsetzen, welchen wir als die eine Koordinate des Schnittpunktes von OB ($\eta = \xi$) mit der Ellipse erhalten.

Bezeichnen wir $\arcsin \frac{b}{\sqrt{a^2+b^2}} = \arctan \frac{b}{a} = \Theta$, so ist $S_{OAB} = \frac{1}{2}\, ab\, \Theta$, denn das abzuziehende Glied ist $S_{OBD} = = \frac{a^2 b^2}{2(a^2+b^2)}$. $S_{OAC} = \frac{1}{4}\, \pi\, ab$ (ein Viertel der Fläche der Ellipse) und $S_{OBC} = \frac{1}{4}\,(\pi - 2\Theta)\, ab$, oder

$$\frac{M_1}{M_2} = \frac{\pi - 2\Theta}{2\Theta} = \varkappa \text{ (neue Bezeichnung).}$$

Daraus erhalten wir $\pi - 2\Theta = 2\Theta\,\varkappa$

$$\Theta = \frac{\pi}{2(1+\varkappa)} = \arctan \frac{b}{a} \text{ und } \frac{b}{a} = \tan \frac{\pi}{2(1+\varkappa)}.$$

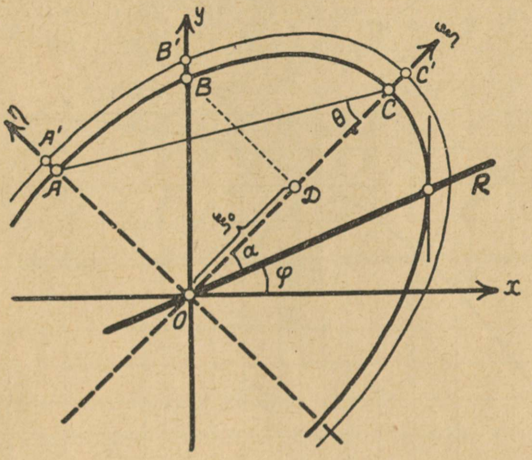

Fig. 4.

Wie Fig. 4 zeigt, ist die Regressionsgerade (R) der konjugierte Durchmesser der y-Achse, denn von ihr aus werden nach beiden Seiten in der Richtung der y-Achse die z-Werte kleiner.

Das Produkt der Steigungsmaße der konjugierten Durchmesser ist $= -\frac{b^2}{a^2}$; da das Steigungsmaß der y-Achse im (ξ, η)-Koordinatensystem gleich 1 ist, so ist $\tan \alpha = -\frac{b^2}{a^2}$; da $\alpha = = \varphi - \frac{\pi}{4}$, so ist $\tan\left(\varphi - \frac{\pi}{4}\right) = -\frac{b^2}{a^2}$, oder $\frac{\tan\varphi - 1}{1+\tan\varphi} = -\frac{b^2}{a^2}$,

$$\text{woraus } \tan\varphi = r = \frac{a^2 - b^2}{a^2 + b^2} = \frac{1 - \left(\frac{b}{a}\right)^2}{1 + \left(\frac{b}{a}\right)^2}.$$

Setzen wir hier den früher gefundenen Wert $\frac{b}{a}$ ein, so erhalten wir

$$r = \frac{1 - \tan^2 \frac{\pi}{2(1+\varkappa)}}{1 + \tan^2 \frac{\pi}{2(1+\varkappa)}}.$$

Beispiel: $M_1 = 162 \quad M_2 = 68 \qquad \varkappa = 2 \cdot 38$

$$\frac{\pi}{2(1+\varkappa)} = 25 \cdot 6^0 \quad \tan^2 25 \cdot 6^0 = 0 \cdot 231.$$

$$r = 0 \cdot 60$$

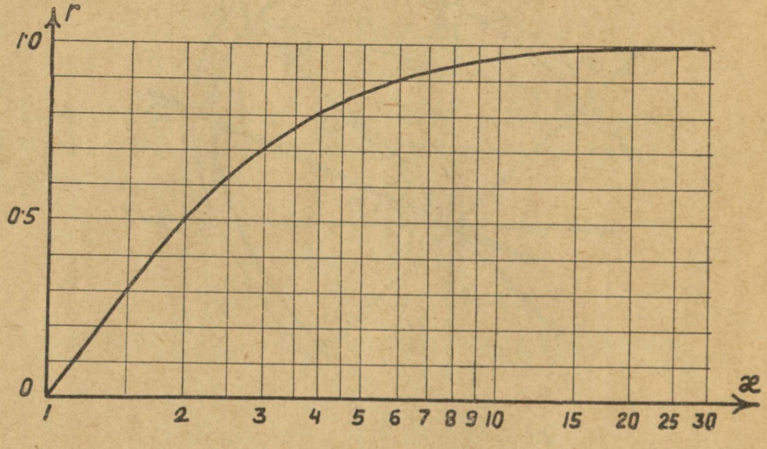

Fig. 5.

Mit Hilfe der Formel kann man eine Tabelle mit den einander zugeordneten Werten von $\varkappa$ und r zusammenstellen (Tab. 1). Man kann auch den Zusammenhang in der Form einer Kurve darstellen (Fig. 5).

Tab. 1.

$\varkappa$	1	1·1	1·2	1·5	2·0	2·5	3·0	3·5	4·0	5·0	6·0	8·0	10·0
r	0	0·08	0·14	0·31	0·50	0·62	0·71	0·77	0·81	0·87	0·90	0·94	0·96

Falls man die Werte von $\varkappa$ nicht berechnen will, kann man ein Nomogramm anfertigen, welches auf Grund von M_1

und M_2 die Ermittelung des r ermöglicht. Da $\varkappa = \frac{M_1}{M_2}$, so ist $\log \varkappa = \log M_1 - \log M_2$. Wählen wir ein logarithmisches Koordinatensystem, so erhalten wir für jedes $\varkappa$, also auch für jedes r, eine Gerade. Alle Geraden sind untereinander parallel. Das betreffende Nomogramm finden wir in Fig. 6.

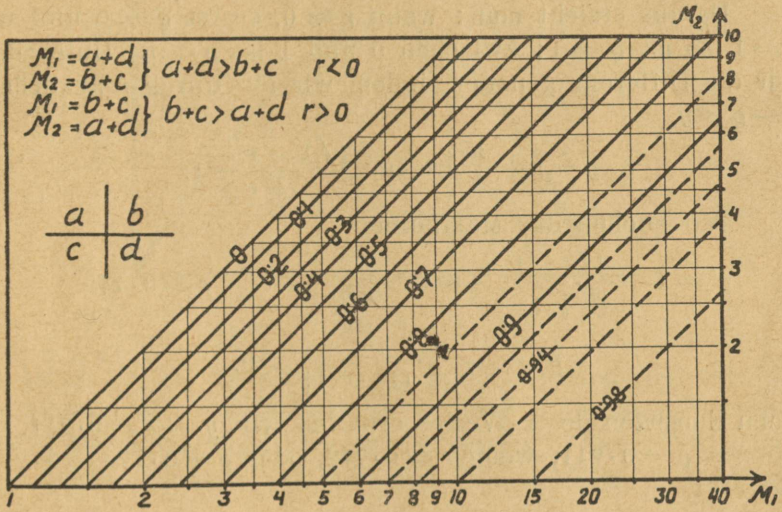

Fig. 6.

Somit wäre die Frage gelöst. Doch ist es oft praktisch, die letzten Gleichungen in einer anderen Form zu gebrauchen. Den Winkel $\frac{\pi}{2(1+\varkappa)}$ schreiben wir in der Form $\frac{2\pi M_2}{4(M_1+M_2)}$; wir addieren und subtrahieren πM_1 und erhalten $\frac{\pi[M_1+M_2-(M_1-M_2)]}{4(M_1+M_2)}$; nennen wir $\frac{M_1-M_2}{M_1+M_2} = \varrho$, so erhalten wir:

$$r = \frac{1 - \tan^2 \frac{\pi}{4}(1-\varrho)}{1 + \tan^2 \frac{\pi}{4}(1-\varrho)}.$$

Im vorigen Beispiel wäre:

$$\varrho = \frac{162 - 68}{162 + 68} = 0 \cdot 408 \qquad \frac{\pi}{4}(1 - 0 \cdot 408) = 25 \cdot 6^0.$$

$$r = 0 \cdot 60$$

Was die erforderliche Rechenarbeit anbelangt, so stehen beide Formeln auf gleicher Stufe. Trotzdem scheint der Ausdruck mit ϱ vorteilhafter zu sein, da $\varrho \approx r$. Lösen wir die Gleichung in Bezug auf ϱ auf, so erhalten wir:

$$\varrho = 1 - \frac{4}{\pi} \operatorname{arc\,tan} \sqrt{\frac{1-r}{1+r}}.$$

Daraus ersieht man: wenn $r = 0$, so ist $\varrho = 0$ und wenn $r = 1$, so ist $\varrho = 1$; zwischen 0 und 1 ist $\varrho < r$. Untersuchen wir die Differenz genauer, indem wir sie durch r ausdrücken. $r - \varrho = y$,

$$y = r + \frac{\pi}{4} \operatorname{arc\,tan} \sqrt{\frac{1-r}{1+r}} - 1.$$

Wir suchen das Maximum

$$y' = 1 - \frac{2}{\pi} \cdot \frac{1}{\sqrt{1-r^2}} = 0 \qquad r^2 = 1 - \frac{4}{\pi} \qquad r = 0 \cdot 772$$

$$y'' = -\frac{2}{\pi} \cdot \frac{r}{(1-r^2)^{\frac{3}{2}}} < 0.$$

Beim Einsetzen des r-Wertes erhalten wir $y_{max} = 0 \cdot 211$. Damit $r - \varrho = 0 \cdot 211$, wenn $r = 0 \cdot 772$, oder $\varrho = 0 \cdot 772 - 0 \cdot 211 = 0 \cdot 561$.

Zur Berechnung von r können wir folgende Formel benutzen: $r = \varrho + y$,

$$y = \frac{1 - \tan^2 \frac{\pi}{4}(1-\varrho)}{1 + \tan^2 \frac{\pi}{4}(1-\varrho)} - \varrho \quad \text{(durch } \varrho \text{ ausgedrückt).}$$

Zur Berechnung der y-Werte kann man sich einer Tabelle (Tab. 2) oder einer Kurve (Fig. 7) bedienen.

Tab. 2.

ϱ	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
y	0·06	0·11	0·15	0·19	0·21	0·21	0·19	0·15	0·09	0·00

Untersuchen wir noch den Sinn von ϱ im Korrelationsfelde.

Wir hatten $\frac{M_1}{M_2} = \frac{\pi - 2\Theta}{2\Theta}$, daraus ergibt sich:

$$\frac{M_1 - M_2}{M_1 + M_2} = \frac{\pi - 4\Theta}{\pi} = \varrho \text{ und } \pi - 4\Theta = \pi\varrho.$$

Wir addieren und subtrahieren $2\Theta\varrho$ und ordnen die Glieder, dann ist

$$\frac{1+\varrho}{1-\varrho}=\frac{\pi-2\Theta}{2\Theta}=\frac{M_1}{M_2}.$$

Der letzte Ausdruck sagt uns folgendes: befindet sich

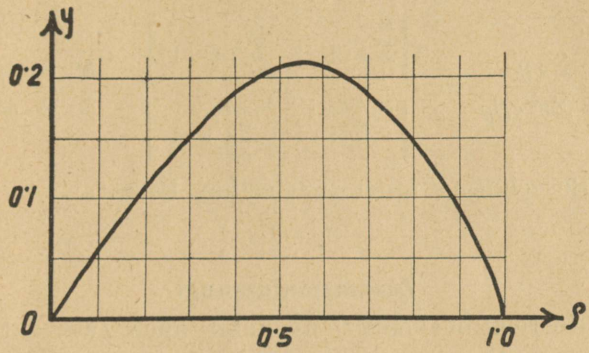

Fig. 7.

eine Masse M_1 in der Entfernung einer 1 oberhalb der x-Achse, eine andere M_2 in derselben Entfernung unterhalb derselben, so ist ϱ die Entfernung ihres gemeinsames Schwerpunktes von der x-Achse. Denken wir uns alle Punkte eines jeden Quad-

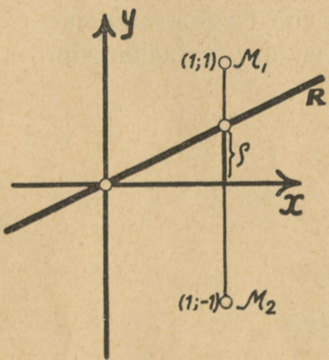

Fig. 8.

ranten des Feldes in einem Punkt konzentriert, dessen Koordinaten den absoluten Wert 1 haben (Fig. 8) (in diesem Falle ist $\sigma_x=\sigma_y=1$), so ergibt ϱ das Steigungsmaß der gesuchten Regressionsgeraden, also den Korrelationskoeffizienten.

Daraus ist ersichtlich, daß $\varrho<r$, denn im M_1-Felde hatten die Koordinaten der Punkte durchschnittlich größere absolute

Werte als im Felde M_2, und das Reduzieren der Entfernungen auf ein- und dieselbe Größe verkleinerte das Gewicht von M_1 und damit den Korrelationskoeffizienten.

Die auf die genannte Art definierten Größen ϱ können wir auch aus der gewöhnlichen Formel erhalten:

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}} = \frac{2\left[\sum_1^{M_1} 1 \cdot 1 + \sum_1^{M_2} 1 \cdot (-1)\right]}{2\left[\sum_1^{M_1+M_2} 1 \cdot 1 \cdot \sum_1^{M_1+M_2} 1 \cdot 1\right]} = \frac{M_1 - M_2}{M_1 + M_2} = \varrho.$$

Die Berechnung führt zum selben Resultat.

Zusammenfassung.

Der Korrelationskoeffizient kann nach zwei vereinfachten Methoden berechnet werden:

1) durch Messung des Korrelationsfeldes,

2) durch das Abzählen der Punkte in den Quadranten.

Die Berechnungen können durchgeführt werden:

1) mit Hilfe der Formeln — analytisch,

2) mit Hilfe von Tabellen und

3) mit Hilfe von Nomogrammen.

Die letzte Art ist die einfachste und ergibt am schnellsten das gesuchte Resultat.