neste; die überwiegende Mehrzahl dieser Kokons war gleichgerichtet, senkrecht, und sie standen wie Bienenzellen nebeneinander; zweigabwärts im Hauptneste eine kleine Gruppe von 6 Kokons und zwischen dieser und der Hauptgruppe zwei mehr isolierte Kokons; endlich zweigaufwärts eine Gruppe von 5 Kokons, die wagrecht im Neste lagen. Die abgestreiften Raupenhäute lagen als schwarze geschrumpfte Körper großenteils außerhalb der Kokons an deren Hinterende, wo sie aus einer kleinen Öffnung herausragten. — Die ersten Schmetterlinge schlüpften zahlreich in der Nacht vom 18. zum 19. Juni. Sie zeigten keine Neigung zur Geselligkeit.

Weitere Versuche zur Lösung schwebender Fragen konnten in diesem Jahre nicht durchgeführt werden. Da ich nicht weiß, wann ich diese Untersuchungen werde fortführen können, gebe ich die bisher gewonnenen Ergebnisse bekannt in der Hoffnung, daß sie zu weiteren Forschungen auf dem so arg vernachlässigten Gebiete der Tiersoziologie anregen mögen.

Berlin-Charlottenburg im November 1920.

Regressionsgleichungen numerischer Merkmale nach Pearsons verallgemeinerter Korrelationstheorie.

Von **Georg Duncker.**

Mit 2 Figuren.

Während die lineare Regression eines numerischen Merkmals auf ein zweites, zu dem es in Korrelation steht, biologisch wohl bekannt ist, und ihre Gleichung in biostatistischen Untersuchungen vielfach angewendet wird, ist dies bezüglich anderer Regressionsformen nicht der Fall, obwohl Karl Pearson bereits 1905 eine unschwer anwendbare Methode zu ihrer Behandlung angegeben hat. Den für Biologen bestimmten neueren Darstellungen der Korrelationslehre in deutscher Sprache (z. B. Goldschmidt 1911, Betz 1911, Johannsen 1913, Exner 1913, Lang 1914, Collier 1921) ist nichts darüber zu entnehmen. Im nachstehenden soll deshalb diese Lücke ausgefüllt werden; in mathematischer Hinsicht wird nur die Kenntnis des binomischen Lehrsatzes vorausgesetzt.

1. Vorbegriffe.

Numerische Merkmale sind solche, deren Varianten in Zahlen ausgedrückt werden können. Ihre statistische Untersuchung ergibt als empirisches Resultat die Variationsreihen derselben, von der allgemeinen Form

Varianten:	V_1	V_2	V_3	. . .
Frequenzen:	f_1	f_2	f_3	. . .

wobei die Gesamtzahl der untersuchten Fälle

$$n = \Sigma\,(f).$$

Eine Variationsreihe kann durch ihre Bestimmungswerte, nämlich das arithmetische Mittel $A = \frac{1}{n} \Sigma (V)$,

die Hauptabweichung $s = \sqrt{\frac{1}{n} \Sigma (V - A)^2}$,

die Momentquotienten $\beta_\nu = \dfrac{\frac{1}{n} \Sigma (V - A)^\nu}{s^\nu}$,

beschrieben werden. Dann sind A und s in derselben Einheit, wie die Varianten V des Merkmals, benannte, die Momentquotienten β_ν dagegen unbenannte Werte und von den letzteren speziell

$$\beta_o = 1, \ \beta_1 = O, \ \beta_2 = 1$$

Konstante. Momentquotienten gerader Ordnung ($\beta_{2\nu}$) ergeben stets positive, mit steigendem ν wachsende Größen, während diejenigen ungerader Ordnung ($\beta_{2\nu \pm 1}$) unabhängig voneinander positiv, negativ oder gleich Null sein können.

Untersucht man zwei numerische Merkmale bei n Individuen, so stellt jedes Individuum eine Variantenkombination dieser beiden Merkmale dar. Dann entspricht den Variationsreihen der Einzelmerkmale das Kombinationsschema des Merkmalpaares, wie etwa das Folgende:

Kombinationsschema der Stachel- und der Weichstrahlzahlen in der Rückenflosse von *Acerina cernua* L. (Kaulbarsch).

I. Stachelzahlen.

II. Weichstrahlzahlen.

V	11	12	13	14	15	16	Σ_{II}
15			9	12	1		22
14	1	2	55	171	16		245
13			102	711	160	6	979
12			21	308	230	9	568
11			2	30	42	4	78
10				1	4		5
9					1	1	2
Σ_I	1	2	189	1233	454	20	$1899 = n$

Die Varianten der beiden Merkmale (I: 11—16, II: 9—15) sind am oberen und am linken Rand des Schemas notiert. Innerhalb dieser Umrahmung finden sich die Frequenzen der einzelnen Variantenkombinationen, $f_{\mu,\nu}$, deren vertikale und horizontale Summen (Σ_I und Σ_{II}) gleich den einzelnen Frequenzen der totalen Variationsreihen der beiden Merkmale, $f_{\mu,0}$ und $f_{0,\nu}$, sind. Daher ergeben diese Frequenzen die Beziehung

(1) $$\Sigma(f_{\mu,\nu}) = \Sigma(f_{\mu,0}) = \Sigma(f_{0,\nu}) = n.$$

Zwischen den beiden Merkmalen besteht Korrelation, solange

(2) $$\Sigma\left(f_{\mu,\nu} - \frac{f_{\mu,0}\, f_{0,\nu}}{n}\right)^2 > 0$$

oder, sofern

$$\varphi = f : n,$$

solange

(3) $$\Sigma(\varphi_{\mu,\nu} - \varphi_{\mu,0}\, \varphi_{0,\nu})^2 > 0.$$

Die einzelnen Kombinationsfrequenzen der Spalten und der Zeilen des Kombinationsschemas bilden mit den neben bezw. über ihnen notierten Varianten die zugeordneten Variationsreihen des einen Merkmals, welche durch die über, bezw. vor ihr angegebene Variante des andern bedingt sind; so bedingt die Stachelzahl 15 mit der Gesamtfrequenz 454 die folgende zugeordnete Variationsreihe der Weichstrahlzahlen

9	10	11	12	13	14	15
1	4	42	230	160	16	1.

Die zugeordneten Variationsreihen können durch gleichartige Bestimmungswerte, wie die totalen, beschrieben werden; die Symbole der zugeordneten Bestimmungswerte seien von denen der totalen durch einen Strich unterschieden. Demnach sind die Bestimmungswerte der totalen und der zugeordneten Variationsreihen für das

erste Merkmal		zweite Merkmal	
total	zugeordnet	total	zugeordnet
A_I	A'_I	A_{II}	A'_{II}
s_I	s'_I	s_{II}	s'_{II}
$\beta_{\nu,0}$	$\beta'_{\nu,0}$	$\beta_{0,\nu}$	$\beta'_{0,\nu}$.

Bei fehlender Korrelation sind die homologen Bestimmungswerte aller zugeordneten Variationsreihen eines Merkmals denen seiner totalen gleich, bei vorhandener von diesen und untereinander verschieden. Z. B. sind die Bestimmungswerte der totalen und der der Stachelzahl 15 zugeordneten Variationsreihe der Weichstrahlzahlen

$$A_{II} = 12.759,\quad s_{II} = 0.788,\quad \beta_{03} = -0.166,\quad \beta_{04} = 3.772$$
$$A'_{II} = 12.313,\quad s'_{II} = 0.750,\quad \beta'_{03} = -0.207,\quad \beta'_{04} = 4.026$$

Ferner ist die mittlere zugeordnete Hauptabweichung, definiert durch

$$[s'_I] = \sqrt{\frac{1}{n}\Sigma(f_{II}\, s'^2_I)}$$

bezw.

$$[s'_{II}] = \sqrt{\frac{1}{n}\Sigma(f_I\, s'^2_{II})},$$

d. h. die Wurzel aus dem Mittel der Quadrate aller zugeordneten Hauptabweichungen eines Merkmals, um so kleiner im Vergleich zur totalen Hauptabweichung desselben, je intensiver die Korrelation des Merkmalpaares ist, und wird bei vollkommener Korrelation zu Null. Ein einfaches Maß der Korrelationsintensität ist daher die Korrelationsquote jedes der beiden Merkmale

$$\tau^2 = \frac{s^2 - [s']^2}{s^2} \tag{4}$$

mit den Grenzwerten Null bei fehlender und Eins bei vollkommener Korrelation und dem wahrscheinlichen Fehler

$$E(\tau) = \lambda \frac{1 - \tau^2}{\sqrt{n}}$$

($\lambda = 0.67449$). Nun ist τ^2_I nicht notwendig stets gleich τ^2_{II}. Für unser obiges Beispiel aber ist

$$\tau_I^2 = \tau_{II}^2 = 0.14100.$$

Außer den bereits angeführten kommt zur Beschreibung des Kombinationsschemas zweier numerischer Merkmale noch eine weitere Gruppe von Bestimmungswerten in Betracht, die unter dem Namen Produkt-Momentquotienten zusammengefaßt seien. Es sind dies die Produktmittel der zur μ-ten bezw. ν-ten Potenz erhobenen relativen, d. h. in der entsprechenden totalen Hauptabweichung ausgedrückten, Abweichungen der individuell kombinierten Varianten jedes der beiden Merkmale von ihrem totalen Mittel, mithin

$$\beta_{\mu,\nu} = \frac{\frac{1}{n} \Sigma \left[(V_I - A_I)^\mu (V_{II} - A_{II})^\nu \right]}{s^\mu{}_I \, s^\nu{}_{II}}$$

oder kürzer, sofern $x = V - A$,

$$\beta_{\mu,\nu} = \frac{\frac{1}{n} \Sigma (x_I{}^\mu x_{II}{}^\nu)}{s_I{}^\mu \, s_{II}{}^\nu}.$$

Sie sind unbenannte voneinander unabhängige Werte und entsprechen den Momentquotienten der isoliert betrachteten Variationsreihen. Der erste derselben

$$\beta_{11} = \frac{\frac{1}{n} \Sigma (x_I \, x_{II})}{s_I \, s_{II}}$$

ist der allgemein bekannte, in der Literatur meistens mit r oder mit ϱ bezeichnete Korrelationskoeffizient mit dem wahrscheinlichen Fehler

$$E(\beta_{11}) = \lambda \frac{1 - \beta^2{}_{11}}{\sqrt{n}}.$$

Die praktische Berechnung sämtlicher in Betracht kommender Bestimmungswerte habe ich in einer demnächst in den Wissensch. Meeresunters. (Helgoland) erscheinenden Arbeit (Die Korrelation zwischen Länge und Gewicht bei Fischen) ausführlich dargestellt.

Regression ist das Größenverhältnis der Abweichungen (x') der zugeordneten Mittel des einen Merkmals von seinem totalen Mittel zu den Abweichungen (x) der sie bedingenden Varianten vom totalen Mittel des anderen. Ist dies Verhältnis konstant, so ist die Regression linear, und es gilt für sie die Regressionsgleichung ersten Grades

(5) $$x'_{II} = \varkappa_{10}\, x_I \text{ bezw. } x'_I = \varkappa_{01}\, x_{II}$$

sowie

(6) $$\frac{x'_{II}}{s_{II}} = \gamma^{(1)}{}_1 \frac{x_I}{s_I} \quad \text{bezw.} \quad \frac{x'_I}{s_I} = \gamma^{(1)}{}_1 \frac{x_{II}}{s_{II}}.$$

Hier ist $x'_{II} = A'_{II} - A_{II}$, $x_I = V_I - A_I$ usw., sowie

$$\varkappa_{10} = \gamma^{(1)}{}_1 \frac{s_{II}}{s_I}, \; \varkappa_{01} = \gamma^{(1)}{}_1 \frac{s_I}{s_{II}}$$

und

$$\gamma^{(1)}{}_1 = \beta_{11}.$$

Aus (5) folgt laut Definition

(7) $$\begin{cases} A'_{II} = A_{II} - \varkappa_{10}\, A_I + \varkappa_{10}\, V_I \\ A'_I = A_I - \varkappa_{01}\, A_{II} + \varkappa_{01}\, V_{II}. \end{cases}$$

Die benannte Werte ergebenden Gleichungen (5) bezw. (7) seien als physische, die unbenannte Werte ergebenden Gleichungen (6) als absolute lineare Regressionsgleichungen des zweiten Merkmals auf das erste, bezw. des ersten auf das zweite bezeichnet.

Lineare Regression kommt in der Natur zwar außerordentlich häufig, jedoch keineswegs ausschließlich vor. Nicht lineare Regression besteht z. B. notwendig zwischen der Totallänge und dem Volumen von Organismen, da erstere eine lineare, letzteres eine dreidimensionale Größe ist; ferner zwischen der Totallänge und solchen sonstigen linearen Dimensionen, deren relative Größe mit der Totallänge abändert. Aber auch bei Zählungen gleichartiger Organe verschiedener Systeme, bei Erblichkeitsbeziehungen u. a. m. ist nicht lineare Regression beobachtet worden.

Die graphische Darstellung der Werte A' des zugeordneten Merkmals als Ordinaten zu den Abszissen V des bedingenden ergibt einen Linienzug, die Regressionslinie des zugeordneten auf das bedingende Merkmal. Diese ist bei linearer Regression eine Gerade, bei nicht linearer eine irgendwie gekrümmte Kurve, die Regressionskurve.

2. Regressionsgleichungen zweiten und höheren Grades.

Pearsons verallgemeinerte Korrelationstheorie (1905) ermöglicht die Wiedergabe linearer und nicht linearer Regressionskurven durch ein System von Gleichungen verschiedenen Grades. Seinem Verfahren liegt die Annahme zugrunde, daß der Verlauf der Regressionskurve durch Mac Laurins Reihe

(8) $$y = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \gamma_3 x^3 + \gamma_4 x^4 + \cdots$$

dargestellt werden könne, vorausgesetzt, daß die Koeffizienten γ_r derselben sich für höhere Werte von ν rasch dem Betrag Null nähern. In der Form von (6) wird (8) zu

$$(9) \qquad \frac{x'_{II}}{s_{II}} = \gamma_0 + \gamma_{10}\frac{x_I}{s_I} + \gamma_{20}\frac{x_I^2}{s_I^2} + \gamma_{30}\frac{x_I^3}{s_I^3} + \gamma_{40}\frac{x_I^4}{s_I^4} + \ldots$$

Denkt man sich (9) für sämtliche n beobachteten Einzelwerte von x_I, also n-mal, niedergeschrieben und die homologen Glieder dieser n Gleichungen summiert, so ergibt sich nach Division mit n als Mittelgleichung derselben

$$(10) \qquad \frac{1}{n}\Sigma\left(\frac{x'_{II}}{s_{II}}\right) = \gamma_0 + \gamma_{10}\frac{1}{n}\Sigma\left(\frac{x_I}{s_I}\right) + \gamma_{20}\frac{1}{n}\Sigma\left(\frac{x_I^2}{s_I^2}\right) + \gamma_{30}\frac{1}{n}\Sigma\left(\frac{x_I^3}{s_I^3}\right) + \ldots$$

Da nun, wie eingangs definiert,

$$\frac{1}{n}\Sigma\left(\frac{x^\nu}{s^\nu}\right) = \beta_\nu$$

und daher $\beta_0 = 1$, $\beta_1 = 0$ und $\beta_2 = 1$, so wird (10) zu

$$(11) \qquad 0 = \gamma_0 + \gamma_{20} + \gamma_{30}\beta_{30} + \gamma_{40}\beta_{40} + \ldots,$$

so daß

$$(12) \qquad \gamma_0 = -\gamma_{20} - \gamma_{30}\beta_{30} - \gamma_{40}\beta_{40} - \ldots$$

Multipliziert man jetzt die Einzelglieder von (9) mit $\frac{x_I}{s_I}$ und bildet hierauf die Mittelgleichung

$$(13) \qquad \frac{1}{n}\Sigma\left(\frac{x_I x'_{II}}{s_I\, s_{II}}\right) = \gamma_0\frac{1}{n}\Sigma\left(\frac{x_I}{s_I}\right) + \gamma_{10}\frac{1}{n}\Sigma\left(\frac{x_I^2}{s_I^2}\right) + \gamma_{20}\frac{1}{n}\Sigma\left(\frac{x_I^3}{s_I^3}\right) + \gamma_{30}\frac{1}{n}\Sigma\left(\frac{x_I^4}{s_I^4}\right) + \ldots,$$

so folgt, da

$$\frac{1}{n}\Sigma\left(\frac{x_I x'_{II}}{s_I s_{II}}\right) = \frac{1}{n}\Sigma\left(\frac{x_I x_{II}}{s_I s_{II}}\right) = \beta_{11},$$

aus dieser

$$(14) \qquad \beta_{11} = \gamma_{10} + \gamma_{20}\beta_{30} + \gamma_{30}\beta_{40} + \gamma_{40}\beta_{50} + \ldots,$$

so daß

$$(15) \qquad \gamma_{10} = \beta_{11} - \gamma_{20}\beta_{30} - \gamma_{30}\beta_{40} - \gamma_{40}\beta_{50} - \ldots$$

In analoger Weise erhält man durch Multiplikation von (9) mit $\frac{x_I^2}{s_I^2}$, $\frac{x_I^3}{s_I^3}$ usw. und nach Bildung der entsprechenden Mittelgleichungen

$$(16) \qquad \begin{cases} \beta_{21} = \gamma_0 + \gamma_{10}\beta_{30} + \gamma_{20}\beta_{40} + \gamma_{30}\beta_{50} + \gamma_{40}\beta_{60} + \ldots \\ \beta_{31} = \gamma_0\beta_{30} + \gamma_{10}\beta_{40} + \gamma_{20}\beta_{50} + \gamma_{30}\beta_{60} + \gamma_{40}\beta_{70} + \ldots \\ \beta_{41} = \gamma_0\beta_{40} + \gamma_{10}\beta_{50} + \gamma_{20}\beta_{60} + \gamma_{30}\beta_{70} + \gamma_{40}\beta_{80} + \ldots \end{cases}$$

und so fort. Nach Substitution der in (12) und (15) gefundenen Werte von γ_0 und γ_{10} nehmen die letzteren Gleichungen die Form

$$(17)\quad\begin{cases}\beta_{21}=\beta_{11}\beta_{30}+\gamma_{20}(\beta_{40}-\beta^2{}_{30}-1)+\gamma_{30}(\beta_{50}-\beta_{30}\beta_{40}-\beta_{30})+\\ \qquad\gamma_{40}(\beta_{60}-\beta_{30}\beta_{50}-\beta_{40})+\cdots\\ \beta_{31}=\beta_{11}\beta_{40}+\gamma_{20}(\beta_{50}-\beta_{30}\beta_{40}-\beta_{30})+\gamma_{30}(\beta_{60}-\beta^2{}_{40}-\beta^2{}_{30})+\\ \qquad\gamma_{40}(\beta_{70}-\beta_{40}\beta_{50}-\beta_{30}\beta_{40})+\cdots\\ \beta_{41}=\beta_{11}\beta_{50}+\gamma_{20}(\beta_{60}-\beta_{30}\beta_{50}-\beta_{40})+\gamma_{30}(\beta_{70}-\beta_{40}\beta_{50}-\beta_{30}\beta_{40})+\\ \qquad\gamma_{40}(\beta_{80}-\beta^2{}_{50}-\beta^2{}_{40}+\cdots\end{cases}$$

usw. an. Setzt man jetzt die Differenz

$$\beta_{20}\,\beta_{\mu,1}-\beta_{11}\,\beta_{\mu+1,\,0}=\delta_{\mu,\,1}$$

und die in Klammern befindlichen Faktoren der Gleichungen (17) in ihrer allgemeinen Form

$$\beta_2\beta_{\iota+\varkappa-2}-\beta_\iota\beta_\varkappa-\beta_2\beta_{\iota-1}\beta_{\varkappa-1}=\varepsilon^{(\iota)}{}_\varkappa,$$

so ist zunächst

$$\delta_{11}=0$$

und $\varepsilon^{(\iota)}{}_\varkappa$ eine Größe $\overline{\iota+\varkappa}$-ter Ordnung. Ferner ist

$$\varepsilon^{(\iota)}{}_\varkappa=\varepsilon^{(\varkappa)}{}_\iota,$$

$$\varepsilon^{(\iota)}{}_1=\varepsilon^{(\iota)}{}_2=\varepsilon^{(\varkappa)}{}_1=\varepsilon^{(\varkappa)}{}_2=\varepsilon^{(1)}{}_\iota=\varepsilon^{(2)}{}_\iota=\varepsilon^{(1)}{}_\varkappa=\varepsilon^{(2)}{}_\varkappa=0.$$

Bei Anwendung der Abkürzungen $\delta_{\mu,1}$ und $\varepsilon^{(\iota)}{}_\varkappa$ erhält man aus (17) die übersichtlichen Beziehungen

$$(18)\quad\begin{cases}\delta_{21}=\gamma_{20}\,\varepsilon^{(3)}{}_{30}+\gamma_{30}\,\varepsilon^{(4)}{}_{30}+\gamma_{40}\,\varepsilon^{(5)}{}_{30}+\gamma_{50}\,\varepsilon^{(6)}{}_{30}+\cdots\\ \delta_{31}=\gamma_{20}\,\varepsilon^{(3)}{}_{40}+\gamma_{30}\,\varepsilon^{(4)}{}_{40}+\gamma_{40}\,\varepsilon^{(5)}{}_{40}+\gamma_{50}\,\varepsilon^{(6)}{}_{40}+\cdots\\ \delta_{41}=\gamma_{20}\,\varepsilon^{(3)}{}_{50}+\gamma_{30}\,\varepsilon^{(4)}{}_{50}+\gamma_{40}\,\varepsilon^{(5)}{}_{50}+\gamma_{50}\,\varepsilon^{(6)}{}_{50}+\cdots\\ \delta_{51}=\gamma_{20}\,\varepsilon^{(3)}{}_{60}+\gamma_{30}\,\varepsilon^{(4)}{}_{60}+\gamma_{40}\,\varepsilon^{(5)}{}_{60}+\gamma_{50}\,\varepsilon^{(6)}{}_{60}+\cdots\end{cases}$$

oder allgemein

$$(18\,\mathrm{a})\quad\begin{cases}\delta_{\mu,1}=\gamma_{20}\,\varepsilon^{(3)}{}_{\mu+1,0}+\gamma_{30}\,\varepsilon^{(4)}{}_{\mu+1,0}+\gamma_{40}\,\varepsilon^{(5)}{}_{\mu+1,0}+\\ \qquad\gamma_{50}\,\varepsilon^{(6)}{}_{\mu+1,0}+\cdots\\ \delta_{1,\nu}=\gamma_{02}\,\varepsilon^{(3)}{}_{0,\nu+1}+\gamma_{03}\,\varepsilon^{(4)}{}_{0,\nu+1}+\gamma_{04}\,\varepsilon^{(5)}{}_{0,\nu+1}+\\ \qquad\gamma_{05}\,\varepsilon^{(6)}{}_{0,\nu+1}+\cdots\end{cases}$$

Auf Grund dieser Beziehungen stehen also zur Auswertung der $\overline{\nu-1}$ Unbekannten γ_2 bis γ_ν stets $\overline{\nu-1}$ Gleichungen zur Verfügung, da die Werte $\varepsilon^{(\iota)}{}_\varkappa$ den Momentquotienten des bedingenden Merkmals, die Werte $\delta_{\mu,1}$ bezw. $\delta_{1,\nu}$ diesen und den Produkt-Momentquotienten des Merkmalpaares zu entnehmen sind.

Bei symmetrischer Variation des bedingenden Merkmals werden dessen Momentquotienten ungerader Ordnung $\beta_{2\nu+1}$ sämtlich zu Null. Daraus folgt, daß auch alle diejenigen Werte $\varepsilon^{(\iota)}{}_\varkappa$, bei denen die Summe $\overline{\iota+\varkappa}$ eine ungerade Zahl ergibt, in diesem Fall zu Null werden, und daß hier

$$\delta_{2\mu,1}=\beta_{2\mu,1}\ \text{und}\ \delta_{1,2\nu}=\beta_{1,2\nu}.$$

Multipliziert man ferner (9) mit $\frac{x'_{II}}{s_{II}}$, so erhält man die Mittelgleichung

(19) $$\frac{1}{n}\Sigma\left(\frac{x'_{II}{}^2}{s_{II}{}^2}\right)=\gamma_0\,\frac{1}{n}\Sigma\left(\frac{x'_{II}}{s_{II}}\right)+\gamma_{10}\frac{1}{n}\Sigma\left(\frac{x_I\,x'_{II}}{s_I\,s_{II}}\right)+$$
$$\gamma_{20}\frac{1}{n}\Sigma\left(\frac{x_I{}^2\,x'_{II}}{s_I{}^2\,s_{II}}\right)+\gamma_{30}\frac{1}{n}\Sigma\left(\frac{x_I{}^3\,x'_{II}}{s_I{}^3\,s_{II}}\right)+\ldots$$

Hier ist, wie leicht zu beweisen,

$$\Sigma\,(x'_{II})=0,\ \Sigma\,(x_I{}^\nu\,x'_{II})=\Sigma\,(x_I{}^\nu\,x_{II})$$

und

$$\frac{1}{n}\Sigma\left(\frac{x'_{II}{}^2}{s_{II}{}^2}\right)=\frac{s^2{}_{II}-[s'_{II}]^2}{s_{II}{}^2}$$
$$=\tau_{II}{}^2,$$

mithin

(20 a) $$\tau_{II}{}^2=\gamma_{10}\,\beta_{11}+\gamma_{20}\,\beta_{21}+\gamma_{30}\,\beta_{31}+\gamma_{40}\,\beta_{41}+\ldots$$

oder, nach Substitution des rechtsseitigen Ausdrucks von (15) für γ_{10},

(20 b) $$\tau_{II}{}^2=\beta_{11}{}^2+\gamma_{20}\,\delta_{21}+\gamma_{30}\,\delta_{31}+\gamma_{40}\,\delta_{41}+\ldots$$

Setzt man endlich die stets positive Differenz

$$\tau_{II}{}^2-\beta_{11}{}^2=\delta_{02}{}^2,$$

so erhält man aus (20 b) den Wert

(21) $$\delta_{02}{}^2=\gamma_{20}\,\delta_{21}+\gamma_{30}\,\delta_{31}+\gamma_{40}\,\delta_{41}+\ldots$$

mit dem wahrscheinlichen Fehler (cf. Blakeman 1905 p. 339 Gleichung XXVII)

$$E\,(\delta_{02}{}^2)=2\,\lambda\,\delta_{02}\sqrt{\frac{1-2\,\delta_{02}{}^2\,(1-\beta_{11}{}^2)+\delta_{02}{}^4}{n}}$$
$$E\,(\delta_{02})=\lambda\sqrt{\frac{1-2\,\delta_{02}{}^2\,(1-\beta_{11}{}^2)+\delta_{02}{}^4}{n}}.$$

Die Gleichungen (9) bis (21) gelten für die Regression des zweiten Merkmals auf das erste; die umgekehrte Beziehung ergibt sich überall, wie in (18 a), durch Vertauschung der Indizes der gefundenen Größen, so z. B. in

(21) $$\delta_{20}{}^2=\gamma_{02}\,\delta_{12}+\gamma_{03}\,\delta_{13}+\gamma_{04}\,\delta_{14}+\ldots$$

Bei linearer Regression (Regression ersten Grades) ist

$$\gamma_2=\gamma_3=\ldots=0,$$

daher nach (12)

$$\gamma^{(1)}{}_0=0$$

und nach (15)

$$\gamma^{(1)}{}_1=\beta_{11}.$$

Man erhält also aus (9) die absolute lineare Regressionsgleichung

(6) $$\frac{x'_{II}}{s_{II}}=\beta_{11}\,\frac{x_I}{s_I}.$$

Bei Regression zweiten Grades, der sogen. quadratischen Regression, ist

$$\gamma_3=\gamma_4=\ldots=0;$$

somit folgt aus (12)

$$\gamma^{(2)}{}_0=-\gamma^{(2)}{}_2,$$

aus (15)

$$\gamma^{(2)}{}_1 = \beta_{11} - \gamma^{(2)}{}_2\, \beta_3$$

und aus (18)

$$\gamma^{(2)}{}_{20} = \delta_{21} : \varepsilon^{(3)}{}_{30}.$$

Aus (9) ergibt sich daher die absolute quadratische Regressionsgleichung

$$(22) \qquad \frac{x'_{II}}{s_{II}} = -\gamma^{(2)}{}_{20} + (\beta_{11} - \gamma^{(2)}{}_{20}\,\beta_{30})\frac{x_I}{s_I} + \gamma^{(2)}{}_{20}\,\frac{x_I{}^2}{s_I{}^2}.$$

Bei Regression dritten Grades, d. h. bei kubischer Regression ist

$$\gamma_4 = \gamma_5 = \ldots = 0,$$

daher nach (12)

$$\gamma^{(3)}{}_0 = -\gamma^{(3)}{}_2 - \gamma^{(3)}{}_3\,\beta_3,$$

nach (15)

$$\gamma^{(3)}{}_1 = \beta_{11} - \gamma^{(3)}{}_2\,\beta_3 - \gamma^{(3)}{}_3\,\beta_4$$

und nach (18)

$$\gamma^{(3)}{}_{20} = \frac{\delta_{21} - \gamma^{(3)}{}_{30}\,\varepsilon^{(3)}{}_{40}}{\varepsilon^{(3)}{}_{30}}$$

$$= \gamma^{(2)}{}_{20} - \frac{\gamma^{(3)}{}_{30}\,\varepsilon^{(3)}{}_{40}}{\varepsilon^{(3)}{}_{30}}$$

$$\gamma^{(3)}{}_{30} = \frac{\delta_{31}\,\varepsilon^{(3)}{}_{30} - \delta_{21}\,\varepsilon^{(3)}{}_{40}}{\varepsilon^{(3)}{}_{30}\,\varepsilon^{(4)}{}_{40} - (\varepsilon^{(3)}{}_{40})^2}.$$

Dann ist

$$\gamma^{(3)}{}_{20} = \frac{\delta_{21}\,\varepsilon^{(4)}{}_{40} - \delta_{31}\,\varepsilon^{(3)}{}_{40}}{\varepsilon^{(3)}{}_{30}\,\varepsilon^{(4)}{}_{40} - (\varepsilon^{(3)}{}_{40})^2}.$$

So erhält man aus (9) die absolute kubische Regressionsgleichung

$$(23) \qquad \frac{x'_{II}}{s_{II}} = -\gamma^{(3)}{}_{20} - \gamma^{(3)}{}_{30}\,\beta_{30} + (\beta_{11} - \gamma^{(3)}{}_{20}\,\beta_{30} - \gamma^{(3)}{}_{30}\,\beta_{40})\frac{x_I}{s_I} +$$

$$\gamma^{(3)}{}_{20}\,\frac{x_I{}^2}{s_I{}^2} + \gamma^{(3)}{}_{30}\,\frac{x_I{}^3}{s_I{}^3}.$$

In entsprechender Weise ergeben sich die absoluten Regressionsgleichungen vierten und höheren Grades, z. B.

$$\gamma^{(4)}{}_0 = -\gamma^{(4)}{}_2 - \gamma^{(4)}{}_3\,\beta_3 - \gamma^{(4)}{}_4\,\beta_4$$

$$\gamma^{(4)}{}_1 = \beta_{11} - \gamma^{(4)}{}_2\,\beta_3 - \gamma^{(4)}{}_3\,\beta_4 - \gamma^{(4)}{}_4\,\beta_5$$

$$\gamma^{(4)}{}_2 = \gamma^{(2)}{}_2 - \frac{\gamma^{(4)}{}_3\,\varepsilon^{(3)}{}_4 + \gamma^{(4)}{}_4\,\varepsilon^{(3)}{}_5}{\varepsilon^{(3)}{}_3}$$

$$\gamma^{(4)}{}_3 = \gamma^{(3)}{}_3 - \frac{\gamma^{(4)}{}_4\,(\varepsilon^{(3)}{}_3\,\varepsilon^{(4)}{}_5 - \varepsilon^{(3)}{}_4\,\varepsilon^{(3)}{}_5)}{\varepsilon^{(3)}{}_3\,\varepsilon^{(4)}{}_4 - (\varepsilon^{(3)}{}_4)^2}$$

und

$$\gamma^{(4)}{}_{40} = \frac{\delta_{41}\,[\varepsilon^{(3)}{}_{30}\,\varepsilon^{(4)}{}_{40} - (\varepsilon^{(3)}{}_{40})^2] - \delta_{31}\,[\varepsilon^{(3)}{}_{30}\,\varepsilon^{(4)}{}_{50} - \varepsilon^{(3)}{}_{40}\,\varepsilon^{(3)}{}_{50}] - \delta_{21}\,[\varepsilon^{(3)}{}_{50}\,\varepsilon^{(4)}{}_{40} - \varepsilon^{(3)}{}_{40}\,\varepsilon^{(4)}{}_{50}]}{\varepsilon^{(5)}{}_{50}\,[\varepsilon^{(3)}{}_{30}\,\varepsilon^{(4)}{}_{40} - (\varepsilon^{(3)}{}_{40})^2] - \varepsilon^{(4)}{}_{50}[\varepsilon^{(3)}{}_{30}\,\varepsilon^{(4)}{}_{50} - \varepsilon^{(3)}{}_{40}\,\varepsilon^{(3)}{}_{50}] - \varepsilon^{(3)}{}_{50}\,[\varepsilon^{(3)}{}_{50}\,\varepsilon^{(4)}{}_{40} - \varepsilon^{(3)}{}_{40}\,\varepsilon^{(4)}{}_{50}]}.$$

Doch wird man Regressionsgleichungen höheren als dritten Grades nur ausnahmsweise anwenden, da die zur Auswertung ihrer Koeffizienten erforderlichen Momentquotienten siebenter und höherer Ordnung bereits mit sehr großen wahrscheinlichen Fehlern behaftet sind.

Die physischen Regressionsgleichungen erhält man aus den absoluten, wenn man x'_{II} durch $A'_{II} - A_{II}$, x_I durch $V_I - A_I$ ersetzt und die Gleichungen nach A'_{II} hin auflöst. Dann ergibt sich allgemein

$$(24)\quad A'_{II} = A_{II} + s_{II}\left(\gamma_0 - \gamma_{10}\frac{A_I}{s_I} + \gamma_{20}\frac{A_I^2}{s_I^2} - \gamma_{30}\frac{A_I^5}{s_I^3} + \gamma_{40}\frac{A_I^4}{s_I^4} \mp \ldots\right)$$
$$+ \frac{s_{II}}{s_I}\left(\gamma_{10} - 2\gamma_{20}\frac{A_I}{s_I} + 3\gamma_{30}\frac{A_I^2}{s_I^2} - 4\gamma_{40}\frac{A_I^3}{s} + \ldots\right)V_I$$
$$+ \frac{s_{II}}{s_I^2}\left(\gamma_{20} - 3\gamma_{30}\frac{A_I}{s_I} + 6\gamma_{40}\frac{A_I^2}{s_I^2} \mp \ldots\right)V_I^2$$
$$+ \frac{s_{II}}{s_I^3}\left(\gamma_{30} - 4\gamma_{40}\frac{A_I}{s_I} \pm \ldots\right)V_I^3$$
$$+ \frac{s_{II}}{s_I^4}\left(\gamma_{40} \mp \ldots\right)V_I^4$$
$$+ \ldots$$

Die Resultate der physischen Regressionsgleichungen aber lassen sich auch direkt aus denen der absoluten entnehmen, da ja

$$A' = A + x'.$$

Der etwaige Mangel an Übereinstimmung zwischen Beobachtung und Berechnung wird durch die mittlere quadratische Differenz der zugeordneten (beobachteten und berechneten) Mittelwerte, bezw. ihrer relativen Abweichungen gemessen. Die Einzeldifferenzen der letzteren, $\varDelta$, können positiv, negativ oder gleich Null sein und ergänzen sich zur Summe Null. Die mittlere quadratische Differenz der absoluten Regressionsgleichung ν-ten Grades des zweiten Merkmals auf das erste ist dann

$$[\varDelta_{II}]_\nu = \pm\sqrt{\frac{1}{n}\Sigma\,(f_I\,\varDelta_{II}^2)}.$$

Diese Größe läßt sich entweder aus den Einzeldifferenzen $\varDelta_{II}$ oder, mit Hilfe von τ_{II}^2, direkt bestimmen. Da nämlich

$$(25)\quad \frac{x'_{II}}{s_{II}} - \left(\gamma_0 + \gamma_{10}\frac{x_I}{s_I} + \gamma_{20}\frac{x_I^2}{s_I^2} + \gamma_{30}\frac{x_I^3}{s_I^3} + \ldots\right) - \varDelta_{II} = 0,$$

so ist

$$(26)\quad \left[\frac{x'_{II}}{s_{II}} - \left(\gamma_0 + \gamma_{10}\frac{x_I}{s_I} + \gamma_{20}\frac{x_I^2}{s_I^2} + \gamma_{30}\frac{x_I^3}{s_I^3} + \ldots\right)\right]^2 = \varDelta_{II}^2.$$

Nach Entwicklung des linksseitigen Ausdrucks von (26), nach Elimination von γ_0 und γ_{10} sowie nach Bildung der Mittelgleichung erhält man

$$(27)\quad [\varDelta_{II}]^2 = \tau_{II}^2 - \beta_{11}^2 - 2\left(\gamma_{20}\,\delta_{21} + \gamma_{30}\,\delta_{31} + \ldots\right) + \gamma_{20}^2\,\varepsilon^{(3)}_{30} + 2\gamma_{20}\gamma_{30}\,\varepsilon^{(3)}_{40} + 2\gamma_{20}\gamma_{40}\,\varepsilon^{(3)}_{50} + \ldots + \gamma_{30}^2\,\varepsilon^{(4)}_{40} + 2\gamma_{30}\gamma_{40}\,\varepsilon^{(4)}_{50} + \ldots + \gamma_{40}^2\,\varepsilon^{(5)}_{50} + \ldots$$

oder nach (21)

(28) $$[\Delta_{II}]_\nu \pm \sqrt{\delta_{02}{}^2 - \sum_1^\nu{}^\mu (\gamma^{(\nu)}{}_{\mu 0}\, \delta_{\mu,1})}\,,$$

und es wird daher

für lineare Regression

$$[\Delta_{II}]_1 = \pm\, \delta_{02}\,,$$

für quadratische Regression

$$\begin{aligned}[\Delta_{II}]_2 &= \pm\sqrt{\delta_{02}{}^2 - (\gamma^{(2)}{}_{20})^2 \varepsilon^{(3)}{}_{30}} \\ &= \pm\sqrt{\delta_{02}{}^2 - \gamma^{(2)}{}_{20}\, \delta_{21}}\,,\end{aligned}$$

für kubische Regression

$$\begin{aligned}[\Delta_{II}]_3 &= + \sqrt{\delta_{02}{}^2 - (\gamma^{(2)}{}_{20})^2\, \varepsilon^{(3)}{}_{30} - (\gamma^{(3)}{}_{30})^2\, \frac{\varepsilon^{(3)}{}_{30}\, \varepsilon^{(4)}{}_{40} - (\varepsilon^{(3)}{}_{40})^2}{\varepsilon^{(3)}{}_{30}}} \\ &= \pm\sqrt{\delta_{02}{}^2 - \gamma^{(3)}{}_{20}\, \delta_{21} - \gamma^{(3)}{}_{30}\, \delta_{31}}\,,\end{aligned}$$

für Regression vierten Grades

$$[\Delta_{II}]_4 = \pm\sqrt{\delta_{02}{}^2 - \gamma^{(4)}{}_{20}\, \delta_{21} - \gamma^{(4)}{}_{30}\, \delta_{31} - \gamma^{(4)}{}_{40}\, \delta_{41}}\,.$$

Dann ist notwendige Bedingung für die Anwendbarkeit einer Regressionsgleichung ν-ten Grades, daß $[\Delta]_\nu{}^2$ einen positiven Zahlenwert ergibt. Wird $[\Delta]_\nu{}^2$ negativ, so bedeutet dies, daß die durch tatsächlichen Vergleich der beobachteten und der mittels der Gleichung ν-ten Grades berechneten Regressionswerte erhaltene mittlere quadratische Differenz größer als $[\Delta]_1$ ist, mithin diese Regressionsgleichung schlechtere Resultate als die lineare liefert.

3. Numerische Beispiele.

Die vorstehenden Ausführungen seien an drei Beispielen erläutert, von denen das erste die zunehmende Übereinstimmung zwischen Beobachtung und Berechnung bei steigendem Grad der Regressionsgleichungen ersten bis dritten Grades ersichtlich macht, das zweite einen in mehrfacher Hinsicht interessanten, konstruierten Fall darstellt, während das dritte zeigt, wie die zu hoch gewählte Regressionsgleichung dritten Grades eine schlechtere Übereinstimmung ergibt, als selbst diejenige ersten Grades.

Wicksell (1918) behandelt das männliche (I) und weibliche (II) Heiratsalter bei Erst- und Wiederverehelichungen in Schweden für die Zeiträume 1891—1900 und 1901—1910 nach eigener, von der Pearsons abweichender Methode. In acht Tabellen gibt er die Kombinationsschemata der vier Möglichkeiten aa, ab, ba und bb (a = Erst-, b = Wiederverehelichung) für jede der beiden Perioden. Hier sei nur die Regression des männlichen auf das weibliche Heiratsalter bei Erstverehelichungen von 256 940 Paaren in 1891—1900 (Wicksell 1918 p. 39 Tab. I) dargestellt.

Die Bestimmungswerte dieses Kombinationsschemas sind:

Tabelle 1.

	I. ♂	II. ♀	Produkt-Momentquotienten	
A	28.61900 Jahre	26.18280 Jahre	β_{11}	0.428622
s	6.08505 „	5.62900 „	β_{12}	0.701390
β_3	1.46446	1.11916	β_{21}	0.663281
β_4	6.46839	5.23016	β_{13}	2.962583
β_5	24.35079	17.03841	β_{31}	2.732156
β_6	121.19004	80.49693	n	256940

Es handelt sich also, wie sachlich selbstverständlich, um hypergeometrische, hochgradig positiv asymmetrische Verteilungen, insbesondere bei den Männern.

Tabelle 2 enthält in der ersten Spalte das Heiratsalter der Mädchen, in der zweiten deren Anzahl, in der dritten die empirischen, in der vierten die berechneten zugeordneten Mittel des männlichen Heiratsalters, in der fünften, sechsten und siebenten die relativen Abweichungen der Werte der ersten, dritten und vierten Spalte von ihren totalen Mitteln. Den Werten der dritten und vierten Spalte dieser Tabelle entsprechen die der zweiten Spalten der Tabellen E und F bei Wicksell (l. c. p. 16), deren Übereinstimmung weniger gut ist als die der hier vorliegenden Werte. Auch finden sich gelegentlich Unstimmigkeiten zwischen Wicksells und meinen Rechnungsergebnissen.

Tabelle 2.

1.	2.	3.	4.	5.	6.	7.
V_{II}	f_{II}	A'_I	$R_3 \pm s_I \Delta_I$	$x_{II} : s_{II}$	$x'_I : s_I$	$R_3 + \Delta_I$
17.5	20330	26.09	26.08 + 0.01	—1.54251	—0.41544	—0.41703 + 0.00159
22.5	105744	26.90	26.91 — 0.01	—0.65425	—0.28244	—0.28134 — 0.00110
27.5	78807	28.70	28.71 — 0.01	0.23400	0.01320	0.01511 — 0.00191
32.5	33918	31.26	31.27 — 0.01	1.12226	0.43480	0.43567 — 0.00087
37.5	12203	34.45	34.43 + 0.02	2.01051	0.95850	0.95568 + 0.00282
42.5	4135	37.99	38.02 — 0.03	2.89877	1.53975	1.54496 — 0.00521
47.5	1250	41.43	41.80 — 0.37	3.78702	2.10565	2.16545 — 0.05980
52.5	409	46.09	45.62 + 0.47	4.67528	2.87181	2.79390 + 0.07791
57.5	93	50.30	49.29 + 1.01	5.56353	3.56228	3.39718 + 0.16510
62.5	32	52.50	52.62 — 0.12	6.45179	3.92453	3.94466 — 0.02013
67.5	12	61.67	55.43 + 6.24	7.34005	5.43095	4.40565 + 1.02530
72.5	7	47.50	57.52 — 10.02	8.22830	3.10284	4.74952 — 1.64668
	256940 $= n$				0.196115 $= \tau_I^2$	+ 0.01275 $= [\Delta_I]_3$

Mit Hilfe der Spalten 2 und 6 der Tabelle 2 findet man

$$\tau_I^2 = 0.196115$$

daher

$$\delta_{20}^2 = \tau_I^2 - \beta_{11}^2 = 0.012389,$$

ferner aus Tabelle 1 die Hilfsgrößen

$$\begin{aligned}
\varepsilon^{(3)}_{03} &= \beta_{04} - \beta_{03}^2 - 1 &&= 2.97764 \\
\varepsilon^{(3)}_{04} &= \beta_{05} - \beta_{03}\beta_{04} - \beta_{03} &&= 10.06586 \\
\varepsilon^{(4)}_{04} &= \beta_{06} - \beta_{04}^2 - \beta_{03}^2 &&= 51.88983 \\
\delta_{12} &= \beta_{12} - \beta_{11}\beta_{03} &&= 0.183584 \\
\delta_{13} &= \beta_{13} - \beta_{11}\beta_{04} &&= 0.490395
\end{aligned}$$

und mittels dieser

$$\gamma^{(2)}_{02} = \delta_{12} : \varepsilon^{(3)}_{03} = 0.061654$$

$$\gamma^{(3)}_{03} = \frac{\delta_{13}\,\varepsilon^{(3)}_{03} - \delta_{12}\,\varepsilon^{(3)}_{04}}{\varepsilon^{(3)}_{03}\,\varepsilon^{(4)}_{04} - (\varepsilon^{(3)}_{04})^2} = -\,0.007290$$

$$\gamma^{(3)}_{02} = \gamma^{(2)}_{02} - \gamma^{(3)}_{03}\,\frac{\varepsilon^{(3)}_{04}}{\varepsilon^{(3)}_{03}} = 0.086302,$$

so daß

$$[\varDelta_I]_1 = \pm\,\delta_{20} = \pm\,0.1113$$

$$[\varDelta_I]_2 = \pm\sqrt{\delta_{20}^{\,2} - \gamma'^{(2)}_{02}\,\delta_{12}} = \pm\,0.0328$$

$$[\varDelta_I]_3 = \pm\sqrt{\delta_{20}^{\,2} - \gamma^{(3)}_{20}\,\delta_{21} - \gamma^{(3)}_{30}\,\delta_{31}} = \pm\,0.0114.$$

Von den drei Regressionsgleichungen ersten bis dritten Grades

$$R_1:\quad \frac{x'_I}{s_I} = 0 + 0.428622\,\frac{x_{II}}{s_{II}}$$

$$R_2:\quad \frac{x'_I}{s_I} = -0.061654 + 0.359621\,\frac{x_{II}}{s_{II}} + 0.061654\,\frac{x_{II}^{\,2}}{s_{II}^{\,2}}$$

$$R_3:\quad \frac{x'_I}{s_I} = -0.078143 + 0.370164\,\frac{x_{II}}{s_{II}} + 0.086302\,\frac{x_{II}^{\,2}}{s_{II}^{\,2}} - 0.007290\,\frac{x_{II}^{\,3}}{s_{II}^{\,3}}$$

ist daher die dritte die zutreffendste. Ihre Auswertung findet man in Spalte 7 der Tabelle 2, aus welcher Spalte 4 derselben Tabelle abgeleitet ist. Die Bestimmung von $[\varDelta_I]_3$ aus den Einzeldifferenzen der Spalte 7 ergibt

$$[\varDelta_I]_3 = \pm\,0.0127,$$

so daß eine geringe Unstimmigkeit der beiden Rechnungsweisen, wohl infolge logarithmischer und dezimaler Abrundungen, vorliegt. Die Zahlenwerte der Spalte 6 (empirische Werte = Kreispunkte) und 7 (berechnete Werte = Linienzug) sind in Fig. 1 graphisch dargestellt;

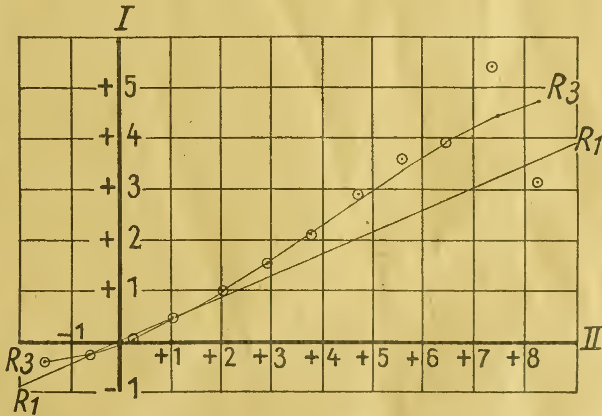

Abb. 1.

ferner ist die Gerade der linearen Regression angegeben. Die beiden bedeutend abweichenden, maximal extremen Werte beruhen auf nur je 12 und 7 unter 256940 Beobachtungen.

Die entsprechenden reziproken Regressionsgleichungen des weiblichen auf das männliche Heiratsalter sind

$$R_1:\quad \frac{x'_{II}}{s_{II}} = \quad 0 \qquad + 0.428622\,\frac{x_I}{s_I}$$

$$R_2:\quad \frac{x'_{II}}{s_{II}} = -0.022171 + 0.396154\,\frac{x_I}{s_I} + 0.022171\,\frac{x_I^2}{s_I^2}$$

$$R_3:\quad \frac{x'_{II}}{s_{II}} = -0.034125 + 0.398770\,\frac{x_I}{s_I} + 0.040934\,\frac{x_I^2}{s_I^2} - 0.004650\,\frac{x_I^3}{s_I^3};$$

sie ergeben die mittleren quadratischen Differenzen

$$[\Delta_{II}]_1 = \pm\, 0.0644,\ [\Delta_{II}]_2 = \pm\, 0.0501,\ [\Delta_{II}]_3 = \pm\, 0.0448.$$

Die Ergebnisse der quadratischen und der kubischen Regressionsgleichung weichen hier von denen der linearen weniger ab, als bei dem männlichen Material.

Das zweite Beispiel ist von W. Johannsen (1913 p. 335) erdacht, um zu zeigen, daß Korrelation selbst dort vorliegen kann, wo ihre Messung durch den Korrelationskoeffizienten Null ergibt. Da dies Beispiel übereinstimmende, streng symmetrische Variation der beiden bei Johannsen mit x (= I) und y (= II) bezeichneten „Merkmale" bietet, erfährt an ihm die Berechnung der Regressionskoeffizienten wesentliche Vereinfachungen. Seine Bestimmungswerte sind

Tabelle 3.

I = II		Produkt-Momentquotienten	
A	0.00000	β_{11}	0.000000
s	1.74929	β_{12}	0.000000
β_3	0.00000	β_{21}	0.553727
β_4	2.70537	β_{13}	0.000000
β_5	0.00000	β_{31}	0.000000
β_6	10.82833	β_{14}	0.000000
β_7	0.00000	β_{41}	2.428630
β_8	54.11560	n	500

Bei symmetrischer Variation eines Merkmals werden alle seine ungeraden Momentquotienten sowie diejenigen Werte $\varepsilon^{(\iota)}{}_{\varkappa}$ zu Null, für welche die Summe $\iota + \varkappa$ ungerade Zahlen ergibt; daher auch

$$\delta_{2\mu,\,1} = \beta_{2\mu,\,1} \text{ und } \delta_{1,\,2\nu} = \beta_{1,\,2\nu}.$$

So erhält man aus den Bestimmungswerten

$$\begin{aligned}
\varepsilon^{(3)}{}_{30} &= \beta_{40} - 1 &&= 1.70537\\
\varepsilon^{(3)}{}_{50} &= \beta_{60} - \beta_{40} &&= 8.12296\\
\varepsilon^{(4)}{}_{40} &= \beta_{60} - \beta_{40}{}^2 &&= 3.50930\\
\varepsilon^{(5)}{}_{50} &= \beta_{80} - \beta_{40}{}^2 &&= 46.79657\\
\delta_{21} &= \beta_{21} &&= 0.553727\\
\delta_{41} &= \beta_{41} &&= 2.428630
\end{aligned}$$

und ferner

$$\tau_{II}{}^2 = \frac{s_{II}{}^2 - [s'_{II}]^2}{s_{II}{}^2} = \delta_{02}{}^2$$
$$= 0.190627.$$

Hieraus findet man

$$\gamma^{(1)}_{1} = 0$$
$$\gamma^{(2)}_{20} = \beta_{21} : \varepsilon^{(3)}_{30} = 0.324696$$
$$\gamma^{(3)}_{30} = 0$$
$$\gamma^{(4)}_{40} = \frac{\beta_{41}\,\varepsilon^{(3)}_{30} - \beta_{21}\,\varepsilon^{(3)}_{50}}{\varepsilon^{(3)}_{30}\,\varepsilon^{(5)}_{50} - (\varepsilon^{(3)}_{50})^2} = -0.025768$$
$$\gamma^{(4)}_{30} = 0$$
$$\gamma^{(4)}_{20} = \gamma^{(2)}_{20} - \frac{\gamma^{(4)}_{40}\,\varepsilon^{(3)}_{50}}{\varepsilon^{(3)}_{30}} = 0.447434$$
$$\gamma^{(4)}_{10} = 0$$
$$\gamma^{(4)}_{0} = -\gamma^{(4)}_{20} - \gamma^{(4)}_{40}\,\beta_{40} = -0.377721,$$

mithin

$$[\Delta_{II}]_1 = \pm \tau_{II} = \pm 0.43661$$
$$[\Delta_{II}]_2 = \pm \sqrt{\tau_{II}^2 - \gamma^{(2)}_{20}\,\beta_{21}} = \pm 0.10409$$
$$[\Delta_{II}]_3 = [\Delta_{II}]_2$$
$$[\Delta_{II}]_4 = \sqrt{\tau_{II}^2 - \gamma^{(4)}_{20}\,\beta_{21} - \gamma^{(4)}_{40}\,\beta_{41}} = \pm 0.07383.$$

Von den vier ersten Regressionsgleichungen des zweiten Merkmals auf das erste

$$R_1:\quad \frac{x'_{II}}{s_{II}} = 0$$

$$R_2:\quad \frac{x'_{II}}{s_{II}} = -0.324696 + 0.324696\,\frac{x_I^2}{s_I^2}$$

$$R_3:\quad \frac{x'_{II}}{s_{II}} = -0.324696 + 0.324696\,\frac{x_I^2}{s_I^2}$$

$$R_4:\quad \frac{x'_{II}}{s_{II}} = -0.377721 + 0.447434\,\frac{x_I^2}{s_I^2} - 0.025768\,\frac{x_I^4}{s_I^4}$$

erweist sich also die vierte als die geeignetste.

Tabelle 4 enthält die relativen Abweichungen der den Varianten -5 bis $+5$ (Spalte 1) des ersten Merkmals zugeordneten Mittel des zweiten, und zwar nach direkter Berechnung aus dem Kombinationsschema (Spalte 4) wie nach den Regressionsgleichungen zweiten und vierten Grades (Spalte 5 und 7). Spalte 2 enthält die Frequenz der bedingenden Varianten, Spalte 3 ihre relativen Abweichungen, Spalte 6 und 8 die Produkte $f_I\,\Delta_{II}^2$. Die fünfstelligen Dezimalen der rechnerischen Ergebnisse sind in der Tabelle auf dreistellige abgerundet.

Tabelle 4.

1	2	3	4	5	6	7	8
V_I	f_I	$x_I : s_I$	$x'_{II} : \Delta_{II}$	$R_2 \pm \Delta_{II}$	$f_I \Delta_{II}^2$	$R_4 + \Delta_{II}$	$f_I \Delta_{II}^2$
0	112	0.000	−0.490	−0.325 − 0.165	3.060	−0.378 − 0.112	1.412
±1	95	±0.572	−0.162	−0.219 + 0.057	0.598	−0.234 + 0.072	0.979
±2	60	±1.143	0.181	0.100 + 0.081	0.793	0.163 + 0.018	0.038
±3	29	±1.715	0.651	0.630 + 0.021	0.024	0.715 − 0.064	0.244
±4	9	±2.287	1.270	1.373 − 0.103	0.190	1.257 + 0.013	0.003
±5	1	±2.858	1.715	2.328 − 0.613	0.752	1.558 + 0.157	0.049
	500			±0.104	5.417	±0.074	2.725
	$= n$			$= [\Delta_{II}]_2$	$= \Sigma$	$= [\Delta_{II}]_4$	$= \Sigma$

Fig. 2 stellt die Werte der Tabelle 4 graphisch dar; R_1 fällt mit der Abszissenachse der Figur zusammen.

Sämtliche reziproken Regressionsgleichungen lauten übereinstimmend

$$R_\nu: \qquad \frac{x'_I}{s_I} = 0,$$

und man findet dementsprechend als Korrelationsquote des ersten Merkmals

$$\tau_I^2 = 0.$$

Im dritten Beispiel, welches des allzukleinen Materials (332 Beobachtungen) wegen an sich keine günstigen Ergebnisse erwarten läßt, handelt es sich um die Regression der Totallänge (I) auf das Körper-

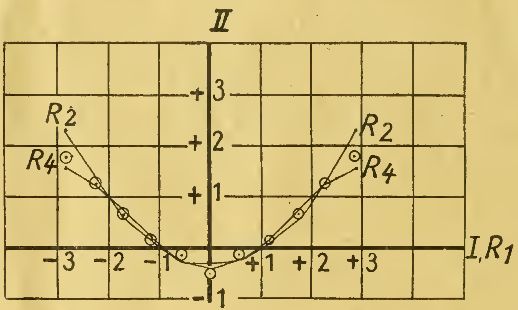

Abb. 2.

gewicht (II) des Kaulbarsches (*Acerina cernua* L.). Die reziproke Regression ist bei dieser Art von praktischer Bedeutung und eine solche dritten Grades. Aus dem an anderer Stelle zu veröffentlichenden (Duncker l. c.) Kombinationsschema der beiden Merkmale erhält man die Bestimmungswerte

Tabelle 5.

	I. Länge	II. Gewicht	Produkt-Momentquotienten	
A	12.66867 cm	29.87950 g	β_{11}	0.900581
s	1.73079 „	11.91655 „	β_{12}	0.668919
β_3	−0.11437	1.14860	β_{21}	0.270057
β_4	3.49836	5.14550	β_{13}	3.898532
β_5	−0.63183	15.04051	β_{31}	3.159070
β_6	18.99604	58.40810	n	332

Da

$$\tau_I^2 = 0.883148,$$

ist

$$\delta_{20}^2 = 0.072102 \pm 0.019659$$

und

$$[\Delta_I]_1 = \pm 0.2685.$$

Ferner sind

$$\gamma^{(2)}{}_{02} = -0.129320$$
$$\gamma^{(3)}{}_{03} = \quad 0.206023,$$

daher

$$[\Delta_I]_2{}^2 = \quad 0.024837$$
$$[\Delta_I]_3{}^2 = -0.062954.$$

Die Anwendung der Regressionsgleichung zweiten Grades bedeutet demnach eine Verbesserung gegenüber der der linearen; dagegen führt die Auswertung der Regressionsgleichung dritten Grades zu gänzlich unhaltbaren Resultaten (s. folgende Tabelle).

Tabelle 6.

1	2	3	4	5	6
V_{II}	f_{II}	$x_{II} : s_{II}$	$x'_I : s_I$	R_2	$R_3 \pm \Delta_I$
10	11	−1.668	−2.303	−1.707	−2.528 + 0.225
15	21	−1.249	−1.666	−1.178	−0.976 − 0.690
20	65	−0.829	−0.684	−0.694	0.029 − 0.713
25	62	−0.409	−0.228	−0.255	0.578 − 0.806
30	67	0.010	0.110	0.138	0.761 − 0.651
35	28	0.430	0.625	0.486	0.671 − 0.046
40	36	0.849	0.914	0.788	0.398 + 0.516
45	16	1.269	0.950	1.045	0.033 + 0.917
50	13	1.688	1.503	1.256	−0.331 + 1.834
55	2	2.108	1.347	1.422	−0.604 + 1.951
60	6	2.528	1.828	1.542	−0.694 + 2.522
65	—	2.947	—	1.617	−0.510 ± ?
70	2	3.367	2.214	1.646	0.039 + 2.175
75	1	3.786	2.214	1.629	1.045 + 1.169
80	2	4.206	3.080	1.567	2.598 + 0.492
	332		0.883		± 0.844
	$= n$		$= \tau_I{}^2$		$= [\Delta_I]_3$

Die quadratische Regressionsgleichung lautet

$$R_2: \quad \frac{x'_I}{s_I} = 0.129320 + 0.885791 \frac{x_{II}}{s_{II}} - 0.129320 \frac{x_{II}{}^2}{s_{II}{}^2}$$

und führt zu

$$[\Delta_I]_2 = + 0.1576.$$

Dagegen ergibt der Vergleich der empirischen mit den nach der kubischen Regressionsgleichung

$$R_3: \quad \frac{x'_I}{s_I} = 0.760504 + 0.090164 \frac{x_{II}}{s_{II}} - 0{,}784067 \frac{x_{II}{}^2}{s_{II}{}^2} + 0.206023 \frac{x_{II}{}^3}{s_{II}{}^3}$$

berechneten Werten

$$[\Delta_I]_3 = \pm 0.8436,$$

also eine mittlere quadratische Differenz, die mehr als das dreifache von $[\Delta_I]_1$ beträgt. Gleichungen vierten und höheren Grades würden wachsende negative Werte von $[\Delta_I]^2$ bedingen und dementsprechend noch weniger zur Wiedergabe der empirischen Befunde geeignet sein.

Literatur.

Betz, W., Über Korrelation. Beih. Zeitschr. ang. Psychol. u. psychol. Sammelf. Nr. 3. Leipzig 1911. 88 pp.

Blakeman, J., On tests for linearity of regression in frequency distributions. Biometrika. Vol. IV Nr. 3 1905 p. 332—350.

Collier, W. A., Einführung in die Variationsstatistik. Berlin 1921. 8°. VI 73 pp.

Exner, F. M., Über die Korrelationsmethode. Jena 1913. 8°. 36 pp.

Goldschmidt, R., Einführung in die Vererbungswissenschaft. — Leipzig 1911. 8°. X + 502 pp.

Johannsen, W., Elemente der exakten Erblichkeitslehre. — Jena 1913. 8°. 2. Aufl. XII + 724 pp.

Lang, A., Experimentelle Vererbungslehre in der Zoologie seit 1900. Erste Hälfte. Jena 1914. gr. 8°. 892 pp.

Pearson, K., Mathematical contributions to the theory of evolution. XIV: On the general theory of skew correlation and non-linear regression. Draper's Co. Res. Mem. Biom. Ser. II. Lond. 1905. 4°. 54 pp. 3 pl.

Wicksell, S. D., Das Heiratsalter in Schweden 1891—1910. Eine korrelationsstatistische Untersuchung. Festskr. Lunds Universitet 250 årsjubil. 1918. Act. R. Soc. Physiograph. Lundens N. F. Bd. 29 Handl. Nr. 18. 46 pp. (14 Tab.).

Zur Ähnlichkeit der Kuckuckseier.

Von Dr. **Horst Wachs**-Rostock.

Mit der Frage der Mimikry der Kuckuckseier beschäftigt sich eine Abhandlung, die Friedrich von Lucanus im Journal für Ornithologie 1921, S. 239 ff. veröffentlicht. Verfasser untersuchte die Sammlung an Kuckuckseiern des Berliner Museums, die 728 Gelege 30 verschiedener Vogelarten mit zusammen 765 Kuckuckseiern enthält. Weitaus die meisten Kuckuckseier waren gezeichnet, nur 17 Stück waren einfarbig; von diesen durchlaufen 16 alle Abstufungen vom tiefen Blaugrün bis zur milchweißen Farbe, ein Ei aus einem Rotkehlchen-Gelege ist lehmgelb.

Die Gruppierung nach „ähnlichen" und „unähnlichen" Eiern ergab eine fast vollkommene Übereinstimmung für dieGruppe der Sylvien (Grasmücken), der weitaus die meisten Gelege angehörten: in 481 Gelegen der Gartengrasmücke waren alle zugehörigen 502 Kuckuckseier als „sehr ähnlich" anzusprechen, desgleichen 16 Kuckuckseier in 15 Gelegen der Dorngrasmücke, 2 bei der Zaun- und 4 bei der Orpheusgrasmücke. Nur bei der Mönchsgrasmücke war von 14 Gelegen in 4 Fällen das Kuckucksei unähnlich. Zeigt sich sonach in dieser Gruppe eine ganz außerordentliche Übereinstimmung, so ist das Gegenteil der Fall bei den Gelegen von *Phylloscopus* (Laubsänger) und *Troglodytes* (Zaunkönig): hier sind alle gefundenen Kuckuckseier als „unähnlich" anzusprechen, und zwar 4 in 4 Gelegen beim Waldlaubsänger, 6 in 5 Gelegen beim Weidenlaubsänger und 120 Kuckuckseier in 109 Gelegen beim Zaun-