Biometrika Trust

Nochmals Über "The Elimination of Spurious Correlation Due to Position in Time or Space

Author(s): Von O. Anderson

Source: *Biometrika*, Vol. 10, No. 2/3 (Nov., 1914), pp. 269-279

Published by: [Biometrika Trust](#)

Stable URL: <http://www.jstor.org/stable/2331784>

Accessed: 22/06/2014 10:47

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Biometrika Trust is collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*.

<http://www.jstor.org>

NOCHMALS ÜBER "THE ELIMINATION OF SPURIOUS CORRELATION DUE TO POSITION IN TIME OR SPACE."

VON O. ANDERSON, St. Petersburg, Rußland.

1. Im Aprilheft der *Biometrika*, hat "Student" gezeigt*, daß das von Cave und Hooker vorgeschlagene Verfahren, den Korrelationskoeffizienten zweier oscillierender Variablen durch Berechnung erster Differenzen (also durch Ersetzung der Reihe $x_1, x_2, \dots x_n$ durch die Reihe: $\Delta'x_1 = x_1 - x_2, \Delta'x_2 = x_2 - x_3, \dots \Delta'x_{n-1} = x_{n-1} - x_n$) vom evolutorischen Element zu befreien, eine Verallgemeinerung zuläßt. Das Verfahren ist nämlich, streng genommen, nur dann richtig, wenn die evolutorische Komponente durch eine lineare Gleichung darstellbar ist. Findet letzteres nicht statt, kann also, z. B., jener nur eine parabolische Gleichung höherer Ordnung genügen, so muß man zweite, dritte u.s.w. Differenzen nehmen (also statt $\Delta'x_1, \Delta'x_2, \dots$ nehme man $\Delta''x_1 = \Delta'x_1 - \Delta'x_2, \Delta''x_2 = \Delta'x_2 - \Delta'x_3, \dots$ etc.) und danach Korrelationskoeffizienten berechnen. Letztere können bald einen konstanten Grenzwert erreichen, der das gewünschte Resultat darstellt.

Unterzeichneter ist schon vor etwa 2 Jahren zu ähnlichen Schlüssen gekommen. Durch von ihm unabhängige Gründe wurde er aber bis jetzt vom Drucke seiner diesbezüglichen Schrift abgehalten. Da er bei seiner Untersuchung Wege einschlägt, die von denen des "Students" sehr verschieden sind, und auch zu manchen Schlüssen kommt, welche letzterem unbekannt geblieben zu sein scheinen, so könnte vielleicht eine kurzgefaßte Darstellung der wichtigsten Resultate seiner Untersuchung für die Leser der *Biometrika* von einigem Interesse sein.

2. *Methode.* Die englische statistische Schule vernachlässigt in ihren Untersuchungen ein Verfahren, das von russischen und deutschen Gelehrten oft angewandt wird (Tchebycheff, Markoff, v. Bortkiewicz, u.s.w.) und neben großer Strenge und Exaktheit noch den Vorzug hat recht elementar zu sein—die Methode der mathematischen Erwartung nämlich. Mathematische Erwartung einer Größe (A) heißt bekanntlich soviel als das Produkt aus dieser Größe und ihrer Wahrscheinlichkeit (w), also Aw . Wenn eine Variable eine Reihe einander ausschließender

* *Biometrika*, Vol. x, Part 1, S. 179, "The Elimination of Spurious Correlation due to Position in Time or Space." By "Student."

Größen annehmen kann, so ist deren math. Erwartung als die Summe der Erwartungen aller dieser Größen definiert. Wir werden hier die mathem. Erwartung überall durch das Symbol $E()$ bezeichnen. $E(A)$ ist also, z. B., gleich Aw .

Die hauptsächlichsten Sätze über mathematische Erwartungen dürften als bekannt angenommen werden. Um aber die Nachprüfung der Formeln dieser Schrift zu erleichtern, werden wir hier die für uns wichtigsten Sätze noch kurz andeuten:

$$(1) \quad E(x + y - z + u - t \dots) = E(x) + E(y) - E(z) + E(u) - E(t) \dots$$

(2) Wenn $x, y, z, \dots$ von einander unabhängig sind, so ist

$$E(x \cdot y \cdot z \cdot \dots) = E(x) \cdot E(y) \cdot E(z) \dots$$

(3) $E(kx) = kE(x)$, wo k const. ist; und daher auch:

$$E(k) = k.$$

(4) Wenn eine Variable X die Werte $x_1, x_2, \dots, x_n$ annehmen kann, so ist die Wahrscheinlichkeit W , daß die Differenz $x_i - E(x)$ zwischen den Grenzen $-\alpha \sqrt{E(x^2) - [E(x)]^2}$ und $+\alpha \sqrt{E(x^2) - [E(x)]^2}$ enthalten sei, größer als $1 - \frac{1}{\alpha^2}$, wo α größer als 1 sein muß (ein Theorem von Tchebycheff).

In unserer Untersuchung werden wir überall statt des wahrscheinlichsten Wertes einer Größe deren mathematische Erwartung berechnen.

Bestimmen wir zuerst, wie sich der Korrelationskoeffizient zweier oscillirender Reihen verhält, wenn man deren Größen durch Differenzen $\Delta'x, \Delta''x, \Delta'''x \dots, \Delta'y, \Delta''y, \Delta'''y, \dots$ ersetzt, und darauf untersuchen wir die Frage von den Grenzen der Anwendbarkeit der verallgemeinerten Cave-Hookerschen Methode. Um Raum zu sparen, werden wir nur die endgiltigen Resultate der Berechnungen anführen, ausgenommen die 3 ersten Formeln, deren Bestimmung als Beispiel der Rechnungsmethode dienen möge.

3. *Definition.* Unter einer oscillatorischen Reihe werden wir eine solche Reihe

$$x_1, x_2, x_3, \dots, x_i, \dots, x_n$$

verstehen, bei der

$$E(x_1) = E(x_2) = \dots = E(x_i) = \dots = E(x_n) = E(x) = \text{const.}$$

und alle einzelnen Glieder von einander völlig unabhängig sind, so daß

$$E(x_i x_j) = E(x_i) \cdot E(x_j), \text{ wobei } i \neq j.$$

Solchen Bedingungen würde zum Beispiel eine Reihe genügen, deren Glieder die Resultate einer Versuchsreihe mit konstanter Wahrscheinlichkeit darstellen, etwa Resultate von Ziehungen aus einer Urne mit m weißen und n schwarzen Kugeln.

4. Mittleres Fehlerquadrat.

Bezeichnen wir $x_i - E(x)$ durch ξ_i , so ist

$$E(\xi_i) = E[x_i - E(x)] = E(x) - E(x) = 0.$$

Da die einzelnen ξ von einander völlig unabhängig sind, so ist

$$E(\xi_i \xi_j) = E(\xi_i) \cdot E(\xi_j) = 0.$$

Das mittlere Fehlerquadrat der Reihe x ist gleich

$$\frac{\sum_1^n [x_i - E(x)]^2}{n}$$

Seine mathematische Erwartung wollen wir (nicht ganz in Übereinstimmung mit der üblichen Bezeichnung) σ_x^2 nennen.

$$\sigma_x^2 = E \left\{ \frac{\sum_1^n [x_i - E(x)]^2}{n} \right\} = \frac{1}{n} \left\{ E \left[\sum_1^n x_i^2 \right] - 2E \left[\sum_1^n x_i \right] \cdot E(x) + n[E(x)]^2 \right\} \\ = E(x^2) - [E(x)]^2,$$

ein Ausdruck, der oben im Satze 4 (§ 2) unter dem Zeichen der Quadratwurzel steht.

$$\text{Andererseits ist aber } E \left\{ \frac{\sum_1^n [x_i - E(x)]^2}{n} \right\} \text{ gleich } E \left\{ \frac{\sum_1^n \xi_i^2}{n} \right\} \text{ und daher} \\ \sigma_x^2 = E(\xi^2).$$

Untersuchen wir den Ausdruck $E \left[\sum_1^n (x_i - M_x)^2 \right]$, wo M_x das arithmetische

Mittel der Reihe x , also $\frac{\sum_1^n x_i}{n}$ bedeutet. Da

$$x_i - M_x = E(x) + \xi_i - \frac{E(x) + \xi_1 + E(x) + \xi_2 + \dots + E(x) + \xi_n}{n} = \xi_i - M_\xi$$

(wenn man M_ξ für $\frac{\sum_1^n \xi_i}{n}$ einsetzt), so haben wir:

$$E \left[\sum_1^n (x_i - M_x)^2 \right] = E \left[\sum_1^n (\xi_i - M_\xi)^2 \right] = E \left[\sum_1^n \xi_i^2 - nM_\xi^2 \right] \\ = nE(\xi^2) - nE(M_\xi^2) = nE(\xi^2) - nE \left\{ \frac{\sum_1^n \xi_i^2 - 2\sum_1^n \xi_i \xi_j}{n^2} \right\} = nE(\xi^2) - E(\xi^2).$$

$$E \left[\sum_1^n (x_i - M_x)^2 \right] = (n-1) E(\xi^2).$$

Um $E(\xi^2)$ zu bekommen, muß man diesen Ausdruck durch $(n-1)$ dividieren.

Es ist also auch

$$\sigma_x^2 = E \left\{ \frac{\sum_1^n (x_i - M_x)^2}{n-1} \right\}$$

Um das Fehlerquadrat $\sigma_{\Delta'x}^2$ für die erste Differenz $\Delta'x$ zu erhalten, berücksichtigen wir, daß

$$\Delta'x_i = (x_i - x_{i+1}) = (\xi_i - \xi_{i+1})$$

und

$$E(\Delta'x_i) = E(\xi_i) - E(\xi_{i+1}) = 0.$$

Daher haben wir:

$$\begin{aligned} \sigma_{\Delta'x}^2 &= E \left\{ \frac{\sum_1^{n-1} [\Delta'x_i - E(\Delta'x)]^2}{n-1} \right\} = E \left\{ \frac{\sum_1^{n-1} (\xi_i - \xi_{i+1})^2}{n-1} \right\} \\ &= \frac{1}{n-1} \left\{ E \left[\sum_1^{n-1} \xi_i^2 \right] - 2E \left[\sum_1^{n-1} \xi_i \xi_{i+1} \right] + E \left[\sum_2^n \xi_i^2 \right] \right\} \\ &= \frac{1}{n-1} \{ (n-1) E(\xi_i^2) - 0 + (n-1) E(\xi_i^2) \} = 2E(\xi_i^2). \end{aligned}$$

Es ist also

$$\sigma_{\Delta'x}^2 = 2\sigma_x^2.$$

Nach demselben Rechenschema ergibt sich für das mittlere Fehlerquadrat

der zweiten Differenz $\Delta''x_i$	der Ausdruck	$6\sigma_x^2$
„ dritten „ $\Delta'''x_i$	„ „	$20\sigma_x^2$
„ vierten „ $\Delta^{(iv)}x_i$	„ „	$70\sigma_x^2$
.....		
„ k-ten „ $\Delta^{(k)}x_i$	„ „	$\frac{2k!}{k! \cdot k!} \sigma_x^2$

Wir können daher folgende Gleichung aufstellen

$$\sigma_x^2 = \frac{\sigma_{\Delta'x}^2}{2} = \frac{\sigma_{\Delta''x}^2}{6} = \frac{\sigma_{\Delta'''x}^2}{20} = \frac{\sigma_{\Delta^{(iv)}x}^2}{70} = \dots = \frac{\sigma_{\Delta^{(k)}x}^2}{\binom{2k!}{k! \cdot k!}} = E(\xi^2), \dots \dots (1)$$

welche exakt ist, und folgende

$$\frac{\sum_1^n (x_i - M_x)^2}{n-1} = \frac{\sum_1^{n-1} \Delta'x_i^2}{2(n-1)} = \frac{\sum_1^{n-2} \Delta''x_i^2}{6(n-2)} = \frac{\sum_1^{n-3} \Delta'''x_i^2}{20(n-3)} = \frac{\sum_1^{n-4} \Delta^{(iv)}x_i^2}{70(n-4)} = \dots = \frac{\sum_1^{n-k} \Delta^{(k)}x_i^2}{\frac{2k!}{k! \cdot k!} (n-k)} = \sigma_x^{2*},$$

..... (1 a)

welche nur annähernd richtig ist.

* Es ist vorteilhafter $\frac{\sum_1^{n-k} \Delta^{(k)}x_i^2}{\frac{2k!}{k! \cdot k!} (n-k)}$ und nicht $\frac{\sum_1^{n-k} (\Delta^{(k)}x_i - M_{\Delta^{(k)}x})^2}{\frac{2k!}{k! \cdot k!} (n-k)}$ zu berechnen.

5. Das mittlere Produkt $\frac{1}{n} \sum_{i=1}^n x_i y_i$

Wenn zwei Reihen

$$x_1, x_2, x_3, \dots, x_n,$$

und

$$y_1, y_2, y_3, \dots, y_n,$$

beide im Sinne des § 3 oscillatorisch sind, und eine Korrelation nur zwischen Größen mit gleichen Indexen, also zwischen x_1 und y_1 , x_2 und y_2 , x_3 und y_3 , u.s.w. bestehen kann, so ist es leicht ersichtlich, daß

$$E \{ [x_i - E(x)] [y_j - E(y)] \} = 0, \text{ wenn } i \neq j.$$

Bezeichnen wir $y_i - E(y)$ durch ψ_i , $x_i - E(x)$ wieder durch ξ_i , so finden wir leicht folgende Ausdrücke:

$$p_{xy} = E \left\{ \frac{1}{n} \sum_{i=1}^n \xi_i \psi_i \right\} = E(\xi_i \psi_i),$$

$$E \left\{ \frac{1}{n-1} \sum_{i=1}^n (x_i - M_x)(y_i - M_y) \right\} = E(\xi_i \psi_i) = p_{xy},$$

$$p_{\Delta'x\Delta'y} = E \left\{ \frac{1}{n-1} \sum_{i=1}^{n-1} \Delta'x_i \Delta'y_i \right\} = 2p_{xy},$$

$$p_{\Delta''x\Delta''y} = E \left\{ \frac{1}{n-2} \sum_{i=1}^{n-2} \Delta''x_i \Delta''y_i \right\} = 6p_{xy},$$

.....

$$p_{\Delta^{(k)}x, \Delta^{(k)}y} = E \left\{ \frac{1}{n-k} \sum_{i=1}^{n-k} \Delta^{(k)}x_i \Delta^{(k)}y_i \right\} = \frac{2k!}{k!k!} p_{xy}.$$

Wir können daher wieder zwei Gleichungssysteme:

$$p_{xy} = \frac{p_{\Delta'x\Delta'y}}{2} = \frac{p_{\Delta''x\Delta''y}}{6} = \frac{p_{\Delta'''x\Delta'''y}}{20} = \frac{p_{\Delta^{(iv)}x\Delta^{(iv)}y}}{70} = \dots = \frac{p_{\Delta^{(k)}x\Delta^{(k)}y}}{\frac{2k!}{k!k!}} = E(\xi_i \psi_i), \dots (2)$$

und

$$\begin{aligned} \frac{\sum_{i=1}^n [x_i - E(x)][y_i - E(y)]}{n-1} &= \frac{\sum_{i=1}^{n-1} \Delta'x_i \Delta'y_i}{2(n-1)} = \frac{\sum_{i=1}^{n-2} \Delta''x_i \Delta''y_i}{6(n-2)} = \frac{\sum_{i=1}^{n-3} \Delta'''x_i \Delta'''y_i}{20(n-3)} \\ &= \frac{\sum_{i=1}^{n-4} \Delta^{(iv)}x_i \Delta^{(iv)}y_i}{70(n-4)} = \dots = \frac{\sum_{i=1}^{n-k} \Delta^{(k)}x_i \Delta^{(k)}y_i}{\frac{2k!}{k!k!} \cdot (n-k)} = p_{xy}, \dots (2a) \end{aligned}$$

aufstellen, von denen das erste exakt und das zweite angenähert ist, und welche den Gleichungen für σ_x^2 (also auch für σ_y^2) des § 4 genau analog sind.

6. *Das Fehlerquadrat der Fehlerquadrate.*

Betrachten wir jetzt den Bereich der Schwankungen der Größen der Systeme (1a) und (2a) um deren mathematisch zu erwartenden Größen in (1) und (2). Mit anderen Worten, gehen wir (mit Rücksicht auf Satz 4 § 2) zur Darstellung der math. Erwartung der Fehlerquadrate der genannten Größen über.

$$\text{Für } \frac{\sum_1^n [x_i - E(x)]^2}{n} \text{ ergibt sich das Fehlerquadrat } \frac{E(\xi^4) - [E(\xi^2)]^2}{n}.$$

$$\text{Für } \frac{\sum_1^n [x_i - M_x]^2}{n-1} \text{ ergibt sich das Fehlerquadrat } \frac{E(\xi^4) - [E(\xi^2)]^2}{n} + \frac{2[E(\xi^2)]^2}{n(n-1)}.$$

$$\text{Für } \frac{\sum_1^{n-1} [\Delta' x_i]^2}{2(n-1)} \text{ ergibt sich das Fehlerquadrat } \frac{(2n-3)\{E(\xi^4) - [E(\xi^2)]^2\} + 2(n-1)[E(\xi^2)]^2}{2(n-1)^2}.$$

$$\text{Für } \frac{\sum_1^{n-2} [\Delta'' x_i]^2}{6(n-2)} \text{ ergibt sich das Fehlerquadrat } \frac{(9n-23)\{E(\xi^4) - [E(\xi^2)]^2\} + (17n-42)[E(\xi^2)]^2}{9(n-2)^2}.$$

$$\text{Und endlich für } \frac{\sum_1^{n-k} (\Delta^{(k)} x_i)^2}{\frac{2k!}{k! \cdot k!} (n-k)} \text{ ergibt sich der recht komplizierte Ausdruck:}$$

$$\frac{1}{A_0^2 (n-k)^2} \left\{ A_0^2 (n-2k) \{E(\xi^4) - [E(\xi^2)]^2\} + 4[E(\xi^2)]^2 [A_1^2 (n-2k+1) + A_2^2 (n-2k+2) + A_3^2 (n-2k+3) + \dots + A_k^2 (n-2k+k)] + 2B_0^2 \{E(\xi^4) - [E(\xi^2)]^2\} + 8[E(\xi^2)]^2 \cdot (B_1^2 + B_2^2 + \dots + B_{k-1}^2) \right\}.$$

Wenn $b_0, b_1, b_2, \dots, b_k$ die Koeffizienten der Zerlegung des Binoms $(1+1)^k$ darstellen, also $b_0=1, b_1=k, b_2=\frac{k \cdot (k-1)}{1 \cdot 2}$, u.s.w., so ist hier

$$\begin{aligned} A_0^2 &= (b_0^2 + b_1^2 + b_2^2 + \dots + b_k^2)^2 = \left[\frac{2k!}{k! \cdot k!} \right]^2, \\ A_1^2 &= (b_0 b_1 + b_1 b_2 + b_2 b_3 + \dots + b_{k-1} b_k)^2 = \left[\frac{2k!}{k-1! \cdot k+1!} \right]^2, \\ A_2^2 &= (b_0 b_2 + b_1 b_3 + \dots + b_{k-2} b_k)^2 = \left[\frac{2k!}{k-2! \cdot k+2!} \right]^2, \\ &\dots\dots\dots \\ A_j^2 &= (b_0 b_j + b_1 b_{j+1} + \dots + b_{k-j} b_k)^2 = \left[\frac{2k!}{k-j! \cdot k+j!} \right]^2, \\ &\dots\dots\dots \\ A_k^2 &= 1, \end{aligned}$$

Es ist also klar, daß zusammen mit dem endlichen Differenzieren die Unsicherheit der Bestimmung von σ_x^2 stetig wächst, anfangs etwa im Verhältnis

$$\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{5} \dots$$

7. *Das Fehlerquadrat des mittleren Produktes.*

Für $\frac{\sum_1^n \xi_i \psi_i}{n}$ ergibt sich das Fehlerquadrat $\frac{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2}{n}$.

Für $\frac{\sum_1^n (x_i - M_x)(y_i - M_y)}{n-1}$ ergibt sich das Fehlerquadrat

$$\frac{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2}{n} + \frac{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2}{n(n-1)}$$

Für $\frac{\sum_1^{n-1} \Delta' x_i \Delta' y_i}{2(n-1)}$ ergibt sich das Fehlerquadrat

$$\frac{(2n-3) \{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2\} + (n-1) \{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2\}}{2(n-1)^2}$$

Im allgemeinen Fall $\frac{\sum_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{\frac{2k!}{k!k!} (n-k)}$ erhalten wir für das Fehlerquadrat folgenden

Ausdruck :

$$\frac{1}{A_0^2 (n-k)^2} \left\{ A_0^2 (n-2k) \{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2\} \right. \\ \left. + 2 \{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2\} [A_1^2 (n-2k+1) + A_2^2 (n-2k+2) + \dots + A_k^2 (n-2k+k)] \right. \\ \left. + 2B_0^2 \{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2\} + 4 \{[E(\xi_i \psi_i)]^2 + \sigma_x^2 \sigma_y^2\} (B_1^2 + B_2^2 + \dots + B_{k-1}^2) \right\},$$

wo die Koeffizienten $A_0^2, A_1^2, \dots, B_0^2, B_1^2, \dots$ dieselbe Bedeutung haben, wie in § 6.

Wenn x_i und y_i einander vollkommen gleich sind, so ist

$$[E(\xi_i \psi_i)]^2 = \{E(\xi^2)\}^2 = \sigma_x^4; \quad E(\xi_i^2 \psi_i^2) = E(\xi_i^4); \quad \sigma_x^2 \sigma_y^2 = \sigma_x^4,$$

und obiger Ausdruck fällt mit dem in § 6 zusammen.

Für den Fall der "normalen" Verteilung können wir auch alle diese Ausdrücke beträchtlich vereinfachen, besonders wenn wir $E(\xi_i \psi_i)$ und $E(\xi_i^2 \psi_i^2)$ als Funktionen von r_{xy} darstellen. Ohne aber hier darauf einzugehen, wollen wir jetzt über den Korrelationskoeffizienten ins Klare kommen.

8. *Definition des Korrelationskoeffizienten.*

Der Korrelationskoeffizient R wird gewöhnlich nach der Formel

$$R_0 = \frac{\frac{\sum_1^n (x_i - M_x)(y_i - M_y)}{1}}{\sqrt{\frac{\sum_1^n (x_i - M_x)^2}{1} \frac{\sum_1^n (y_i - M_y)^2}{1}}} \text{ berechnet.}$$

Zu welchem Ausdruck ist er nun als empirische Annäherung aufzufassen,

$$\text{zu } E \left\{ \frac{\sum_1^n (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_1^n (x_i - M_x)^2 \sum_1^n (y_i - M_y)^2}} \right\} \text{ oder zu } \frac{E \left[\sum_1^n (x_i - M_x)(y_i - M_y) \right]}{\sqrt{E \left[\sum_1^n (x_i - M_x)^2 \right] \cdot E \left[\sum_1^n (y_i - M_y)^2 \right]}} ?$$

Beide Formeln sind durchaus nicht mit einander zu identifizieren und fallen nur in erster Annäherung zusammen. Da die zweite aber bedeutend leichter zu handhaben ist und dies auch mehr den üblichen Rechnungsmethoden der englischen Schule entspricht, so *definieren* wir r_{xy} als

$$\frac{E \left[\sum_1^n (x_i - M_x)(y_i - M_y) \right]}{\sqrt{E \left[\sum_1^n (x_i - M_x)^2 \right] \cdot E \left[\sum_1^n (y_i - M_y)^2 \right]}}$$

Anders ausgedrückt ist $r_{xy} = \frac{p_{xy}}{\sigma_x \sigma_y}$, wo p_{xy} , σ_x , σ_y die Bedeutungen haben, welche wir ihnen oben in §§ 4 und 5 beigemessen haben.

9. *Das Verhalten des Korrelationskoeffizienten zweier oscillierender Reihen x und y , wenn man deren Größen durch Differenzen ersetzt.*

Für die k -te endliche Differenz von x und y haben wir

$$r_{\Delta^{(k)}x, \Delta^{(k)}y} = \frac{E \left(\sum_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i \right)}{\sqrt{E \left[\sum_1^{n-k} \Delta^{(k)} x_i^2 \right] \cdot E \left[\sum_1^{n-k} \Delta^{(k)} y_i^2 \right]}} = \frac{p_{\Delta^{(k)}x, \Delta^{(k)}y}}{\sqrt{\sigma^2_{\Delta^{(k)}x} \cdot \sigma^2_{\Delta^{(k)}y}}} = \frac{\frac{2k!}{k! \cdot k!} p_{xy}}{\sqrt{\frac{2k!}{k! \cdot k!} \sigma_x^2 \cdot \frac{2k!}{k! \cdot k!} \sigma_y^2}},$$

$$r_{\Delta^{(k)}x, \Delta^{(k)}y} = \frac{p_{xy}}{\sigma_x \sigma_y} = r_{xy}.$$

Wir haben also ganz allgemein das *genaue* Resultat :

$$r_{xy} = r_{\Delta'x, \Delta'y} = r_{\Delta''x, \Delta''y} = r_{\Delta'''x, \Delta'''y} = \dots = r_{\Delta^{(k)}x, \Delta^{(k)}y}.$$

Da aber diese r unbekannt bleiben und wir für ein beliebiges $r_{\Delta^{(i)}x, \Delta^{(i)}y}$ nur

$$\text{dessen Annäherungsformel } R_i = \frac{\sum_1^{n-i} \Delta^{(i)} x_i \Delta^{(i)} y_i}{\sqrt{\sum_1^{n-i} \Delta^{(i)} x_i^2 \cdot \sum_1^{n-i} \Delta^{(i)} y_i^2}}$$

wiederum feststellen, inwiefern man sich in der Praxis auf die Übereinstimmung der empirischen Koeffizienten mit deren mathematischen Erwartungen verlassen kann, wie groß also die Unsicherheit ihrer Bestimmung zu schätzen ist.

10. *Mittleres Fehlerquadrat des Korrelationskoeffizienten der endl. Differenzen zweier Reihen.*

Aus der Formel $R_k = \frac{\sum_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{\sqrt{\sum_1^{n-k} \Delta^{(k)} x_i^2 \cdot \sum_1^{n-k} \Delta^{(k)} y_i^2}}$, kann man folgenden Ausdruck

ableiten:

$$\frac{R_k - r_{xy}}{r_{xy}} = \frac{\frac{\sum_1^{n-k} \Delta^{(k)} x_i \Delta^{(k)} y_i}{n-k} - p_{\Delta^{(k)} x, \Delta^{(k)} y}}{p_{\Delta^{(k)} x, \Delta^{(k)} y}} - \frac{\frac{\sum_1^{n-k} \Delta^{(k)} x_i^2}{n} - \sigma^2_{\Delta^{(k)} x}}{2\sigma^2_{\Delta^{(k)} x}} - \frac{\frac{\sum_1^{n-k} \Delta^{(k)} y_i^2}{n} - \sigma^2_{\Delta^{(k)} y}}{2\sigma^2_{\Delta^{(k)} y}},$$

der nur in erster Annäherung und, wenn alle 4 Brüche des Ausdrucks echte sind, richtig ist; dies ist bei großem n der Fall. Und ferner ergibt sich daraus die Formel:

$$\sigma^2_{R_k} = \frac{(1 - r^2_{xy})^2}{(n-k)^2} \left\{ (n-k) + 2(n-k-1) \left(\frac{k}{k+1} \right)^2 + 2(n-k-2) \left(\frac{k \cdot (k-1)}{(k+1) \cdot (k+2)} \right)^2 \right. \\ \left. + 2(n-k-3) \left(\frac{k \cdot (k-1) \cdot (k-2)}{(k+1) \cdot (k+2) \cdot (k+3)} \right)^2 + \dots \right. \\ \left. + 2(n-k-k) \left(\frac{k \cdot (k-1) \cdot (k-2) \dots 2 \cdot 1}{(k+1) \cdot (k+2) \cdot (k+3) \dots (k+k)} \right)^2 \right\} \\ \frac{\sum_1^{n-k} (\Delta^{(k)} x_i)^2}{k! \cdot k! (n-k)}$$

(vergl. dazu die Formel für $\frac{1}{2k!} \frac{\sum_1^{n-k} (\Delta^{(k)} x_i)^2}{(n-k)}$ in § 6).

Die Formel für $\sigma^2_{R_k}$ ist immer nur dann gültig, wenn man

$$\frac{r^2_{xy}}{n} \left\{ \frac{E(\xi_i^2 \psi_i^2) - [E(\xi_i \psi_i)]^2}{[E(\xi_i \psi_i)]^2} + \frac{E(\xi^4) - [E(\xi^2)]^2}{4[E(\xi^2)]^2} + \frac{E(\psi^4) - [E(\psi^2)]^2}{4[E(\psi^2)]^2} \right. \\ \left. - \frac{E(\xi_i^3 \psi_i) - E(\xi_i \psi_i) \cdot E(\xi^2)}{E(\xi_i \psi_i) \cdot E(\xi^2)} - \frac{E(\xi_i \psi_i^3) - E(\xi_i \psi_i) \cdot E(\psi^2)}{E(\xi_i \psi_i) \cdot E(\psi^2)} \right. \\ \left. + \frac{E(\xi_i^2 \psi_i^2) - E(\xi^2) \cdot E(\psi^2)}{2E(\xi^2) \cdot E(\psi^2)} \right\}$$

gleich $\frac{(1 - r^2_{xy})^2}{n}$ setzen darf (vergl. *Biometrika*, Vol. IX. p. 4).

Aus der Formel für $\sigma^2_{R_k}$ erhalten wir:

$$\sigma^2_{R_0} = \frac{(1 - r^2_{xy})^2}{n}, \\ \sigma^2_{R_1} = \frac{(1 - r^2_{xy})^2}{n-1} \cdot \frac{3n-4}{2(n-1)}, \\ \sigma^2_{R_2} = \frac{(1 - r^2_{xy})^2}{n-2} \cdot \frac{35n-88}{18(n-2)}, \\ \sigma^2_{R_3} = \frac{(1 - r^2_{xy})^2}{n-3} \cdot \frac{231n-843}{100(n-3)},$$

u. s. w.

Die Fehlerquadrate der Korrelationskoeffizienten aufeinanderfolgender Differenzenordnungen verhalten sich folglich zueinander ungefähr wie $2 : 3 : 4 : 5 \dots$.

Die Unsicherheit wächst also mit zunehmender Differenzenordnung etwa im Verhältnis

$$\sqrt{2} : \sqrt{3} : \sqrt{4} : \sqrt{5}, \dots$$

11. *Korrelationskoeffizient zweier zusammengesetzter Reihen, die aus oscillatorischen und evolutorischen Elementen bestehen.*

Da "Student" diese Frage treffend dargelegt hat, können wir uns kurz fassen. Wenn wir in Betracht ziehen, daß für uns die evolutorische Komponente einer Reihe schon dann in der Praxis verschwunden ist, wenn sie im Verhältnis zur oscillatorischen Komponente so klein geworden ist, daß sie nur die 3^{ten}, 4^{ten}, u.s.w. Zahlenstellen des Ausdruckes für R beeinflussen kann, so kommen wir zum Schluß, daß nicht nur Komponenten, die durch eine Parabel höherer Ordnung darstellbar sind, sondern auch solche, denen nur transzendente Gleichungen (z. B. Sinusreihen) genügen, beim endlichen Differenzieren eliminiert werden. Ja mehr noch, man kann beweisen, daß überhaupt alle mehr oder minder "glatten Reihen," alle bei denen eine genügende positive Korrelation zwischen den Nachbargliedern bemerkbar ist, für die Praxis beim endlichen Differenzieren verschwinden. Das verallgemeinerte Cave-Hookersche Verfahren ist daher augenscheinlich ein sehr universales Mittel, die Korrelation oscillatorischer Elemente aus zusammengesetzten Reihen herauszuschälen. Es hat aber einen Haken, auf den hier noch hingewiesen werden muß.

12. *Kann man aus dem Verhalten der Reihe $R_0, R_1, R_2, \dots R_k$ bestimmen, ob wir den Korrelationskoeffizienten rein oscillatorischer Reihen vor uns haben?* "Student" scheint zu glauben, daß wenn irgendein R_i seinem Vorgänger R_{i-1} gleich ist, wir es sicher mit dem Korrelationskoeffizienten oscillierender Elemente zu tun haben. Vor einem solchen Schluß ist nachdrücklich zu warnen. Wie es meine (für diesen Artikel etwas zu langwierigen) Berechnungen zeigen, können zwei Nachbarkoeffizienten R_i, R_{i-1} auch bei stark evolutorischen Reihen einander ungefähr gleich sein, und die Wahrscheinlichkeit eines solchen Zusammentreffens ist gar nicht sehr gering einzuschätzen. Nur wenn wir, von irgendeinem R_j angefangen, immer dieselbe Größe für R erhalten, also $R_j = R_{j+1} = R_{j+2} = R_{j+3} = \dots$, wird ein solcher Schluß berechtigt sein, und je länger die Reihe gleicher R , desto wahrscheinlicher wird dieser Schluß.