Die Dispersion der Knabenquote bei Zwillingsgeburten.

Von Dr. L. v. Borkiewicz,
ordentlichem Professor der Staatswissenschaften an der Universität Berlin.

Die Anwendung der von *Lexis* begründeten Theorie der Dispersion statistischer Zahlenwerte auf den Fall der Knabenquote bei Zwillingsgeburten begegnet einer gewissen Schwierigkeit, die dadurch verursacht wird, dass zwischen zwei aus ein und derselben Zwillingsgeburt hervorgegangenen Kindern in bezug auf deren Geschlecht keine Unabhängigkeit im Sinne der Wahrscheinlichkeitsrechnung besteht. Es sei p die Wahrscheinlichkeit, dass ein neugeborenes Kind ein Knabe und q , dass es ein Mädchen ist, und es seien bei Zwillingsgeburten w_2, w_1, w_0 der Reihe nach die Wahrscheinlichkeiten der drei Geschlechtskombinationen 2 Knaben, 1 Knabe und 1 Mädchen, 2 Mädchen. Man hat demgemäss

$$(1) \quad p + q = 1,$$

$$(2) \quad w_2 + w_1 + w_0 = 1,$$

$$(3) \quad p = w_2 + \frac{w_1}{2},$$

$$(4) \quad q = w_0 + \frac{w_1}{2}.$$

Es würden zugleich die Beziehungen

$$(5) \quad w_2 = p^2, \quad w_1 = 2 p q, \quad w_0 = q^2$$

bestehen, wenn je zwei zusammengehörende Zwillinge in bezug auf das Geschlecht unabhängig voneinander wären. Der „Unabhängigkeitshypothese“ widersprechen aber die Ergebnisse der Statistik.

Es sei für einen in z gleiche Abschnitte eingeteilten Zeitraum die Zahl der auf den i^{ten} Zeitabschnitt entfallenden Zwillingsgeburten n_i , und es seien a_i, b_i, c_i die Zahlen derjenigen unter diesen n_i Zwillingsgeburten, welche auf die Geschlechtskombinationen 2 Knaben, 1 Knabe und 1 Mädchen, 2 Mädchen kommen. Demnach ist

$$(6) \quad a_i + b_i + c_i = n_i,$$

und setzt man

$$\frac{a_i}{n_i} = \alpha_i, \quad \frac{b_i}{n_i} = \beta_i, \quad \frac{c_i}{n_i} = \gamma_i,$$

so erhält man

$$(7) \quad \alpha_i + \beta_i + \gamma_i = 1.$$

Bezeichnet man ferner mit s_i die Zahl der im i^{ten} Zeitabschnitt als Zwillinge geborenen Kinder, mit x_i die Zahl der Knaben unter ihnen und mit y_i den Quotienten $\frac{x_i}{s_i}$, so hat man:

$$(8) \quad s_i = 2 n_i,$$

$$(9) \quad x_i = 2 a_i + b_i$$

und

$$(10) \quad y_i = \alpha_i + \frac{\beta_i}{2}.$$

Im folgenden soll für die Summen der z Werte n_i, a_i, b_i, c_i, s_i und x_i, N, A, B, C, S und X geschrieben werden. Dementsprechend stellen sich die aus den z Einzelwerten $\alpha_i, \beta_i, \gamma_i, y_i$ unter Berücksichtigung ihrer Gewichte (n_i bzw. s_i) gebildeten arithmetischen Mittelwerte, die mit $\alpha_0, \beta_0, \gamma_0, y_0$ bezeichnet werden sollen, in der Form

$$\alpha_0 = \frac{A}{N}, \quad \beta_0 = \frac{B}{N}, \quad \gamma_0 = \frac{C}{N}, \quad y_0 = \frac{X}{S}$$

dar.

Ist nun die Zahl N einigermaßen gross, so erhält man stets

$$(11) \quad \alpha_0 > y_0^2, \quad \beta_0 < 2 y_0 (1 - y_0), \quad \gamma_0 > (1 - y_0)^2,$$

und da die statistischen „Häufigkeiten“ $\alpha_0, \beta_0, \gamma_0, y_0$ der Reihe nach als empirische Werte der Wahrscheinlichkeiten w_2, w_1, w_0, p betrachtet werden können, so ist aus (11) auf

$$(12) \quad w_2 > p^2, \quad w_1 < 2 p q, \quad w_0 > q^2$$

zu schliessen. Für das Deutsche Reich 1901—1911 findet man z. B.: $\alpha_0 = 0,3211, \beta_0 = 0,3771, \gamma_0 = 0,3018, y_0 = 0,5097, y_0^2 = 0,2598, 2 y_0 (1 - y_0) = 0,4998, (1 - y_0)^2 = 0,2404$. Man kann dieses durch die Ungleichungen (11) charakterisierte typische Verhalten der Zahlenwerte $\alpha_0, \beta_0, \gamma_0$ als „Geschlechtsattraktion“ bezeichnen.

Die Dispersion der Häufigkeit y_i untersuchen, heisst an die Schwankungen der y_i -Werte diejenige Norm anlegen, welche die Wahrscheinlichkeitsrechnung

hierfür aufstellt. Würde es sich hierbei, statt um Zwillingsgeburten, um einfache Geburten handeln und demnach s_i die Zahl der Einzelgeborenen, x_i die Zahl der Knaben unter ihnen und y_i , nach wie vor, den Quotienten $x_i : s_i$ bedeuten, so wäre nach Lexis in der Weise zu verfahren, dass man die *wirkliche* mittlere Abweichung von y_i , die mit σ bezeichnet werden möge, mit der entsprechenden erwartungsmässigen oder *theoretischen* mittleren Abweichung, die man mit δ bezeichnen wolle, vergleicht. Je nachdem σ nahezu gleich δ , merklich kleiner als δ oder merklich grösser als δ ausfällt, spricht Lexis von normaler, unternormaler oder übernormaler Dispersion. Bildet man aber aus σ und δ den sogenannten *Divergenzkoeffizienten* $Q = \frac{\sigma}{\delta}$, so erhält man als Kriterien der drei Arten der Dispersion: $Q =$ nahezu 1, Q merklich kleiner als 1, Q merklich grösser als 1.

Bei $s_i = \text{const.} = s$ ist

$$(13) \quad \sigma^2 = \frac{1}{z-1} \sum_{i=1}^z (y_i - y_0)^2$$

und

$$(14) \quad \delta^2 = \frac{y_0(1-y_0)}{s}$$

zu setzen. In Formel (13) ist der Subtrahendus 1 im Nenner durch den Umstand bedingt, dass die Abweichungen $y_i - y_0$ nicht „wahre“, sondern „scheinbare“ Fehler sind, und in Formel (14) vertreten y_0 und $1 - y_0$ die ihnen zugrunde liegenden Wahrscheinlichkeiten p bzw. q . Die Zahlen s_i sind aber in der Wirklichkeit niemals konstant, und es wird diesem Umstand nach dem Vorgang von Lexis meist in der Weise Rechnung getragen, dass man, ohne an der Formel (13) etwas zu ändern, in Formel (14) für s das arithmetische Mittel s_0 der z Werte s_i , somit $S : z$ einsetzt. Man kommt so auf die Formel

$$(15) \quad Q^2 = \frac{s_0 \sum_{i=1}^z (y_i - y_0)^2}{(z-1) y_0 (1-y_0)}$$

Genauer ist jedoch die Formel

$$(16) \quad Q^2 = \frac{z \sum_{i=1}^z (x_i - s_i y_0)^2}{(z-1) X (1-y_0)}$$

Es möge in diesem Zusammenhange auf zwei Punkte hingewiesen werden, die für die Beurteilung des Charakters der Dispersion an dem Wert von Q nicht

ohne Bedeutung sind. Erstens kommt es hierbei, streng genommen, nicht sowohl darauf an, ob Q , wie Lexis annimmt, sondern ob Q^2 mehr oder weniger, sei es in plus; sei es in minus, von 1 abweicht, weil die mathematische Erwartung von Q^2 gleich 1, diejenige von Q aber etwas kleiner als 1 ist. Sie lässt sich näherungsweise durch $1 - \frac{1}{4(z-1)}$ ausdrücken. Zweitens empfiehlt es sich, was Lexis unterlässt, stets in methodischer Weise darauf Rücksicht zu nehmen, inwiefern die Abweichung des ermittelten Q -Wertes von dessen mathematischer Erwartung zufälligen Charakter trägt. Dies geschieht zweckmässigerweise so, dass man den absoluten Betrag der betreffenden Abweichung zu dem mittleren Fehler von Q , der sich näherungsweise durch $\frac{1}{\sqrt{2(z-1)}}$ ausdrücken lässt, in Beziehung setzt. Im folgenden soll dementsprechend das Verhältnis jener positiv genommenen Abweichung zu dem mittleren Fehler von Q unter der Bezeichnung η jeweils mitangegeben werden.

Eine Anwendung der Formeln (13)–(16) auf den Fall der Knabenquote (y_i) bei Zwillingsgeburten ist nun deshalb nicht angängig, weil dies bedeuten würde, dass man an der Unabhängigkeitshypothese, trotzdem sie von der Erfahrung widerlegt wird, festhält oder, anders ausgedrückt, dass man der statistisch verbürgten Tatsache der Geschlechtsattraktion keine Rechnung trägt. Es bietet sich allerdings die Möglichkeit, die Unabhängigkeitshypothese dadurch auszuschalten, dass man statt der Knabenquote y_i die Häufigkeiten $\alpha_i, \beta_i, \gamma_i$ zum Gegenstand der Untersuchung macht. Dabei erscheint als Versuchszahl n_i statt s_i , und bezeichnet man das arithmetische Mittel der z -Werte n_i mit n_0 , so erhält man als Gegenstück zu (15) und (16), wenn es sich z. B. um die Häufigkeit α_i handelt:

$$(17) \quad Q^2 = \frac{n_0 \sum_{i=1}^z (\alpha_i - \alpha_0)^2}{(z-1) \alpha_0 (1-\alpha_0)}$$

$$(18) \quad Q^2 = \frac{z \sum_{i=1}^z (\alpha_i - n_i \alpha_0)^2}{(z-1) A (1-\alpha_0)}$$

Lexis selbst hat sich denn auch im Fall der Zwillingsgeburten auf die Betrachtung der verschiedenen in Frage kommenden Geschlechtskombinationen beschränkt, somit als Versuchszahlen die Zahlen der gesamten herangezogenen Zwillingspaare und als Ereigniszahlen die Zahlen der Zwillingspaare von be-

stimmter Geschlechtszusammensetzung verwendet¹⁾. In dem von Lexis behandelten Beispiel, das sich auf preussisches Material für 1862—1873 gründet, stellte sich ein Resultat heraus, dem ein Wert 1,255 von Q entspricht. Da es sich aber hierbei um eine einzige Reihe von nur 12 Jahresergebnissen handelte, so war Lexis zu der Behauptung berechtigt, dass die Annahme einer normalen Dispersion in diesem Falle „noch keineswegs ausgeschlossen“ sei. In der Tat findet man in diesem Fall $\eta = 1,30$.

Im Unterschied von Lexis hat alsdann sein Schüler Herrl neben der Häufigkeit der verschiedenen Geschlechtskombinationen die Knabenquote bei sämtlichen Zwillingsgeburten zum Gegenstand dispersionstheoretischer Untersuchungen gemacht²⁾. Er hat hierbei als Versuchszahlen durchweg, d. h. auch im Fall der Knabenquote bei sämtlichen Zwillingsgeburten die Zahlen nicht der geborenen Zwillinge (s_i), sondern der Zwillingsgeburten (n_i) benützt und erhielt für die verschiedenen Geschlechtskombinationen Werte von Q , die um 1 oszillieren, dagegen für die Knabenquote bei sämtlichen Zwillingsgeburten ausnahmslos Werte von Q , die hinter 1 zurückbleiben, so für ganz Preussen 1867—1881 $Q = 0,773$ und für 15 preussische Gebietsteile (teils Provinzen, teils Regierungsbezirke) 1867—1881 im Durchschnitt 0,723. Diese niedrigen Werte von Q haben Herrl nicht überrascht. „Indem wir“, meint er, „um das Geschlechtsverhältnis bei sämtlichen Zwillingsgeburtsfällen festzustellen, zu den gleichgeschlechtlichen Zwillingsgeburtsfällen die zweigeschlechtlichen zu je gleichen Teilen addieren, entsteht in unserer Reihe eine ausgleichende Tendenz“, wodurch eben die Reihe scheinbar den Charakter einer gebundenen erhalte. In der so formulierten Erklärung des gewonnenen Resultats, die durchaus zutreffend ist, liegt das Eingeständnis, dass die zur Bestimmung von Q angewandte Methode nicht die richtige war. Wie

¹⁾ W. Lexis, Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft. Freiburg i. B. 1877, S. 74—78. Hier wird eine Berechnung nur für α_i und γ_p , nicht aber auch für β_p , ausgeführt, und zwar bezieht sich diese Berechnung nicht direkt

auf α_i und γ_p , sondern auf den Quotienten $\frac{\alpha_i}{\gamma_i}$, was ein Analogon

dazu ist, dass Lexis bei den Einzelgeborenen sich immer an den Quotienten $\frac{y_i}{1-y_i}$ statt an y_i hält. Sehr zweckmässig ist diese

Verfahrungsweise nicht, weil sie eine neue Ungenauigkeit in die Rechnung hineinträgt, aber das Wesen der Sache wird dadurch nicht berührt.

²⁾ G. A. Herrl, Über die Stabilität des Geschlechtsverhältnisses bei Mehrlingsgeburten. Freiburg i. B. 1884 (Freiburger philos. Dissertation). Auch Herrl operiert wie Lexis (siehe Fuss-

note 1) mit dem Quotienten $\frac{y_i}{1-y_i}$.

aber in diesem Fall zu verfahren sei, lässt Herrl „in Ermangelung genügend sicherer Hypothesen“ dahingestellt sein. Ja, er ist sogar der Meinung, dass die Zwillingsgeburten in ihrer Gesamtheit überhaupt nicht gleichmässig mit den gleichgeschlechtlichen behandelt werden können¹⁾.

Eine grössere Zuversicht in dieser Beziehung legt Horowicz an den Tag, der wohl ebenfalls von Lexis dazu angeregt worden ist, dem Problem des Geschlechtsverhältnisses bei Zwillingsgeburten eine besondere Studie zu widmen²⁾. Sofern es sich um die Knabenquote (y_i) handelt, konstruiert Horowicz eine rechnungsmässige Versuchszahl g_p , die zwischen n_i und s_i liegt, um sie in die Formel der erwartungsmässigen mittleren Abweichung einzusetzen. Er geht dabei davon aus, dass es eineiige und zweieiige Zwillingsgeburten gibt, von denen die ersteren stets Kinder gleichen Geschlechts, die letzteren aber Kinder teils gleichen, teils verschiedenen Geschlechts liefern. Zugleich nimmt er an: 1. dass die Wahrscheinlichkeit für einen Zwilling, männlichen Geschlechts zu sein, in den beiden Kategorien der Zwillingsgeburten die gleiche ist, und 2. dass bei den zweieiigen Zwillingsgeburten das Geschlecht des einen Zwillinges von dem Geschlecht des anderen unabhängig ist. Diesen beiden Annahmen zufolge sind bei den zweieiigen Zwillingsgeburten die Wahrscheinlichkeiten jeder der drei Geschlechtskombinationen 2 Knaben, 1 Knabe und 1 Mädchen, 2 Mädchen der Reihe nach durch p^2 , $2 p q$, q^2 gegeben.

Es seien nun für den i^{ten} Zeitabschnitt: n'_i und n''_i die Zahlen der eineiigen und der zweieiigen neugeborenen Zwillingspaare, s'_i und s''_i die Zahlen der aus jeder der beiden Kategorien von Zwillingsgeburten hervorgegangenen Kinder, so dass $s'_i = 2 n'_i$, $s''_i = 2 n''_i$, und x'_i bzw. x''_i die Zahlen der Knaben unter ihnen. Man setze ferner

$$\frac{n'_i}{n_i} = \varepsilon'_i, \quad \frac{n''_i}{n_i} = \varepsilon''_i$$

und bezeichne mit e' bzw. e'' die Wahrscheinlichkeit, dass sich eine erfolgte Zwillingsgeburt aus einem Ei bzw. aus zwei Eiern entwickelt hat. Nach Horowicz ist die Zahl der (voneinander unabhängigen) Versuche in den Fällen eineiiger Befruchtung durch die Zahl der Zwillingspaare, hingegen in den Fällen zweieiiger Befruchtung durch die Zahl der betreffenden Kinder dargestellt. Man müsste dementsprechend $g_i = n'_i + s''_i$ setzen, woraus sich $g_i = n'_i + 2 n''_i$ oder auch $g_i = n_i (\varepsilon'_i + 2 \varepsilon''_i)$ und schliesslich wegen $\varepsilon'_i + \varepsilon''_i = 1$

$$(19) \quad g_i = (1 + \varepsilon''_i) n_i$$

¹⁾ A. a. O., S. 38—39.

²⁾ K. J. Horowicz, Über das Geschlechtsverhältnis bei Zwillingsgeburten. Göttingen 1912 (Göttinger philos. Dissertation).

ergäbe. *Horowicz* setzt aber $\varepsilon_i'' = \text{const.} = \varepsilon''$, erhält demnach

$$(20) \quad g_i = (1 + \varepsilon'') n_i$$

und bestimmt ε'' auf der Grundlage der seinen Annahmen entsprechenden Gleichung

$$(21) \quad w_1 = \varepsilon'' \cdot 2 p q,$$

die sich durch Substitution der empirischen Werte $\beta_0, \varepsilon'', y_0, 1 - y_0$ für w_1, ε'', p, q in

$$(22) \quad \beta_0 = 2 \varepsilon'' y_0 (1 - y_0)$$

verwandelt, woraus

$$(23) \quad \varepsilon'' = \frac{\beta_0}{2 y_0 (1 - y_0)}$$

und, wenn man mit *Horowicz* $y_0 = \frac{1}{2}$ setzt,

$$(24) \quad \varepsilon'' = 2 \beta_0$$

hervorgeht. Für Preussen 1887—1907 findet *Horowicz* $\alpha_0 = 0,3209, \beta_0 = 0,3780, \gamma_0 = 0,3011$, somit auf Grund von (24) $\varepsilon'' = 0,7560$ und auf Grund von (20) $g_i = 1,7560 n_i$. Es ergibt sich $Q = 1,11545^1$.

Die von *Horowicz* vorgeschlagene Methode zur Bestimmung von Q im Fall der Knabenquote bei Zwillingsgeburten wäre nur dann stichhaltig, wenn die statistische Grösse, um deren Schwankungen es sich hierbei handelt, nicht der Quotient y_i , welcher auch als

$$(25) \quad y_i = \frac{x_i' + x_i''}{s_i' + s_i''}$$

dargestellt werden kann, sondern der — einer statistischen Feststellung unzugängliche — Quotient

$$(26) \quad h_i = \frac{\frac{1}{2} x_i' + x_i''}{\frac{1}{2} s_i' + s_i''}$$

wäre, d. h. wenn man bei Berechnung der empirischen Werte der Knabenquote die aus eineiigen Befruchtungen

¹⁾ A. a. O., S. 28—29 und 32—35. Mit der Berechnung von Q bis auf die 5. Dezimale steht die „vereinfachende“ Annahme $y_0 = \frac{1}{2}$ nicht ganz im Einklang. In diesem Fall ist $y_0 = 0,5099$, und Formel (23) ergibt $\varepsilon'' = 0,7563$. Erhöht man demgemäss den von *Horowicz* berechneten Wert von Q im Verhältnis von $\sqrt{1,7563}$ zu $\sqrt{1,7560}$, so erhält man $Q = 1,11555$. Ich sehe im Text davon ab, dass *Horowicz*, darin *Lewis* und *Herrl* folgend, die Dispersion nicht von y_i , sondern von $\frac{y_i}{1 - y_i}$ untersucht. Siehe Fussnoten 1 und 2 auf Seite 237, erste Spalte.

hervorgehenden Kinder nur mit ihrer halben Zahl in Ansatz bringen würde. Stellt man die beiden Grössen y_i und h_i in der Form

$$(27) \quad y_i = \frac{\varepsilon_i' \frac{x_i'}{s_i'} + \varepsilon_i'' \frac{x_i''}{s_i''}}{\varepsilon_i' + \varepsilon_i''}$$

und

$$(28) \quad h_i = \frac{\frac{1}{2} \varepsilon_i' \frac{x_i'}{s_i'} + \varepsilon_i'' \frac{x_i''}{s_i''}}{\frac{1}{2} \varepsilon_i' + \varepsilon_i''}$$

dar, so sieht man direkt ein, dass ihre mittleren Fehler nicht übereinstimmen können. Denn $\frac{x_i'}{s_i'}$ schwankt, relativ (d. h. unter Berücksichtigung der Verschiedenheit der Zahlen s_i' und s_i'') stärker als $\frac{x_i''}{s_i''}$, und dieser Umstand bringt es, der verschiedenen Art wegen, wie einerseits y_i und andererseits h_i aus $\frac{x_i'}{s_i'}$ und $\frac{x_i''}{s_i''}$ zusammengesetzt sind, mit sich, dass von diesen beiden Grössen y_i stärker schwankt als h_i . Hätte also *Horowicz* die Werte h_i an Stelle der Werte y_i zum Gegenstand der Untersuchung machen können und tatsächlich gemacht, wie es seiner Konstruktion der Versuchszahl g_i entsprechen haben würde, so wäre er erwartungsgemäss auf einen kleineren Wert von σ und demgemäss auf einen kleineren Wert von Q gekommen. Der von ihm berechnete Wert von Q ist somit zu hoch ausgefallen. Was aber das Ausmass des von *Horowicz* begangenen methodologischen Fehlers anlangt, so wird es sich aus der nachfolgenden positiven Darlegung unschwer bestimmen lassen.

Man führe die Hilfsgrösse ϑ_j ein, welche die Knabenquote bei einem einzelnen, und zwar dem j^{ten} Zwillingspaar in der Reihe der n_i Zwillingspaare angeben soll. Offenbar kann ϑ_j nur einen der drei Werte 0, $\frac{1}{2}$ und 1 annehmen. Da diesen Werten der Reihe nach die Wahrscheinlichkeiten w_0, w_1 und w_2 zukommen, so ist die mathematische Erwartung von ϑ_j durch $\frac{1}{2} w_1 + w_2$ oder, laut Formel (3), durch p gegeben. Ferner erhält man als mathematische Erwartung von ϑ_j^2 den Ausdruck $\frac{1}{4} w_1 + w_2$ oder $p - \frac{1}{4} w_1$. Hieraus folgt, dass das Quadrat der theoretischen mittleren Abweichung von ϑ_j gleich $p - \frac{w_1}{4} - p^2$ oder gleich $p q - \frac{w_1}{4}$ ist. Nun lässt sich aber y_i als arith-

metisches Mittel der n_i Werte ϑ_j auffassen. Die theoretische mittlere Abweichung von y_i wird daher durch

$$(29) \quad \sqrt{\frac{pq - \frac{w_1}{4}}{n_i}}$$

oder näherungsweise durch

$$(30) \quad \sqrt{\frac{y_0(1-y_0) - \frac{\beta_0}{4}}{n_i}}$$

ausgedrückt. Der Formel

$$(31) \quad \sigma^2 = \frac{1}{z-1} \sum_{i=1}^z (y_i - y_0)^2$$

entspricht also die Formel

$$(32) \quad \delta^2 = \frac{y_0(1-y_0) - \frac{\beta_0}{4}}{n},$$

wenn $n_i = \text{const.} = n$. Trifft aber diese Bedingung nicht zu, so kann man ebenso, wie es bei Aufstellung der Formel (15) geschehen ist, für n das arithmetische Mittel n_0 einsetzen. Auf diese Weise gelangt man zu

$$(33) \quad Q^2 = \frac{n_0 \sum_{i=1}^z (y_i - y_0)^2}{(z-1) \left\{ y_0(1-y_0) - \frac{\beta_0}{4} \right\}}$$

Den Vorzug verdient indessen die den Formeln (16) und (18) analoge Formel

$$(34) \quad Q^2 = \frac{z \sum_{i=1}^z (x_i - y_0 s_i)^2}{(z-1) \left\{ 2X(1-y_0) - B \right\}}$$

Die von *Herrl* zur Berechnung von Q^2 benützte Formel unterscheidet sich von Formel (33) durch das Fehlen des Ausdrucks $-\frac{\beta_0}{4}$ im Nenner. Mithin sind seine Q -Werte, wenn sie richtiggestellt werden sollen, mit dem Faktor

$$(35) \quad \sqrt{\frac{4y_0(1-y_0)}{4y_0(1-y_0) - \beta_0}}$$

zu multiplizieren¹⁾. Auf der Grundlage der von *Herrl* für ganz Preussen 1867—1881 festgestellten Werte $y_0 = 0,5093$ und $\beta_0 = 0,3694$ berechnet sich der Ausdruck (35) zu 1,26, so dass sich die im obigen angeführten *Herrlschen* Q -Werte von 0,773 auf 0,974 und von 0,723 auf 0,911 erhöhen. Der mittlere Fehler des ersten dieser beiden Werte beträgt $\frac{1}{\sqrt{2 \cdot 14}} =$

0,189, der des zweiten $\frac{1}{\sqrt{15 \cdot 2 \cdot 14}} = 0,049$. Dabei

ist zu bedenken, dass die mathematische Erwartung von Q nicht 1, sondern in beiden Fällen $1 - \frac{1}{4 \cdot 14} =$

0,982 ist. Demnach weichen die betreffenden Werte von Q um 0,008 bzw. 0,071 von ihrer mathematischen Erwartung in minus ab, und man erhält $\eta = 0,04$ bzw. $\eta = 1,45$. Bei den unkorrigierten Werten von Q würde man aber $\eta = 1,11$ bzw. $\eta = 5,29$ erhalten. Was insbesondere den Fall betrifft, wo sich Q im Durchschnitt auf 0,723 (unkorrigiert) bzw. 0,911 (korrigiert) stellt, so erreicht von den 15 unkorrigierten Einzelwerten, die dem Durchschnitt zugrunde liegen, kein einziger den Wert 0,982, während von den 15 korrigierten Einzelwerten 7 den Wert 0,982 und zugleich 1 übersteigen. Sonach kann hier von keiner ausgesprochenen unternormalen Dispersion die Rede sein.

Die Beziehung zwischen der auf Formel (30) beruhenden und der von *Horowicz* vorgeschlagenen Methode zur Bestimmung von Q tritt klar zutage, wenn man eine seiner Grösse g_i entsprechende Grösse f_i im Sinne einer fingierten Versuchszahl einführt und demgemäss

¹⁾ Dies gilt, wohlbemerkt, nur von denjenigen unter den *Herrlschen* Q -Werten, die sich auf die Knabenquote bei (sämtlichen) Zwillingsgeburten beziehen. Die Q -Werte hingegen, welche die verschiedenen Geschlechtskombinationen bei den Zwillingsgeburten betreffen, bedürfen keiner Berichtigung, weil da, wie eingangs dargelegt worden ist, die üblichen Formeln anwendbar sind — mit der Massgabe allerdings, dass man als Versuchszahlen die Zahlen der Zwillingspaare betrachtet, was *Herrl* auch getan hat. Er sagt ausdrücklich (S. 16): „Wir gebrauchen . . . für g nicht die Zahl der Geborenen, sondern der *Fälle* von Zwillingsgeburten.“ Vgl. S. 43, wo das Wort „Fälle“ sogar durch fetten Druck hervorgehoben wird. Es beruht daher auf einem ganz unbegreiflichen Missverständnis, wenn *A. A. Tschuprow* (Studien zur Theorie der Statistik, 2. Aufl., russisch, St. Petersburg, 1910, S. 351—353) gegen *Herrl* den Vorwurf erhebt, mit den Zahlen der Geborenen als Versuchszahlen operiert zu haben, und wenn er demgemäss sich veranlasst sieht, die *Herrlschen* Q -Werte in der Weise umzurechnen, dass er sie durch $\sqrt{2}$ dividiert. Im übrigen ist *Tschuprow* der Meinung, dass es bei einer Untersuchung der Häufigkeit der verschiedenen Geschlechtskombinationen sein Bewenden haben müsse (a. a. O., S. 351). Er hält somit die Frage einer korrekten Bestimmung von Q im Fall der Knabenquote bei (sämtlichen) Zwillingsgeburten für unlösbar!

$$(36) \quad \frac{y_0(1-y_0) - \frac{\beta_0}{4}}{n_i} = \frac{y_0(1-y_0)}{f_i}$$

setzt. Hieraus ergibt sich

$$(37) \quad f_i = \frac{4 y_0 (1-y_0) n_i}{4 y_0 (1-y_0) - \beta_0}$$

Demgegenüber hat man wegen (20) und (23)

$$(38) \quad g_i = \frac{\{2 y_0 (1-y_0) + \beta_0\} n_i}{2 y_0 (1-y_0)},$$

und man erhält:

$$\frac{f_i}{g_i} = \frac{8 y_0^2 (1-y_0)^2}{8 y_0^2 (1-y_0)^2 + \beta_0 \{2 y_0 (1-y_0) - \beta_0\}}$$

Also ist ein Wert von Q , der nach *Horowicz's* Methode berechnet ist, um richtiggestellt zu werden, mit dem Faktor

$$(39) \quad \sqrt{\frac{8 y_0^2 (1-y_0)^2}{8 y_0^2 (1-y_0)^2 + \beta_0 \{2 y_0 (1-y_0) - \beta_0\}}}$$

zu multiplizieren, d. h. entsprechend zu verringern, weil ja das Produkt

$$\beta_0 \{2 y_0 (1-y_0) - \beta_0\}$$

in allen in Frage kommenden Fällen positiv und nicht Null ist. Negativ kann dieses Produkt wegen (22) nicht werden, da ϵ'' nicht über 1 hinausgehen kann. Wäre $\beta_0 = 0$, so fände die übliche Berechnungsweise Anwendung, wobei als Versuchszahl die Zahl der Zwillingspaare betrachtet werden müsste. Bei $2 y_0 (1-y_0) = \beta_0$ erhielte man laut Formel (23) $\epsilon'' = 1$, somit $\epsilon' = 0$, und es würde wiederum die übliche Formel des mittleren Fehlers in Kraft treten, und zwar mit der Massgabe, dass die Versuchszahl der Zahl der Geborenen gleichzusetzen wäre. Unterstellt man aber einen dieser beiden Spezialfälle, so wird *Horowicz's* Konstruktion offenbar gegenstandslos. Es bestätigt sich also, dass seine Methode stets einen zu grossen Wert von Q liefert. Setzt man in (39) die Werte $y_0 = 0,5099$ und $\beta_0 = 0,3780$, wie sie *Horowicz* für Preussen 1887—1907 ermittelt hat, ein, so erhält man 0,957. Dementsprechend sinkt der betreffende von *Horowicz* berechnete Q -Wert von 1,115 auf 1,067 herab. In diesem Fall beträgt die mathematische Erwartung des Quotienten Q näherungsweise 0,988 und dessen mittlerer Fehler 0,158, so dass sich die festgestellte Abweichung auf 0,079, somit auf 50 % des zugehörigen mittleren Fehlers, beläuft ($\eta = 0,50$), während bei dem unkorrigierten Wert von Q die entsprechende Abweichung 0,127, somit 80 % des mittleren Fehlers, beträgt ($\eta = 0,80$).

Eine grössere Beachtung als die numerische Ungenauigkeit der Resultate, zu denen die von *Horowicz* in Vorschlag gebrachte Methode führt, verdient der Umstand, dass, wie obige Ableitung der Formel (29) und der aus ihr folgenden Formeln (33) und (34) zeigt, die korrekte Lösung der hier zur Diskussion stehenden Aufgabe — der Meinung *Horowicz's* und auch *Herrls* entgegen — von irgendwelchen Hypothesen oder Theorien biologischen Charakters über das Zustandekommen von Zwillingsgeburten ganz und gar unabhängig ist¹⁾. Obige Ableitung gestaltet sich denn auch, wenn man sie mit der Unterscheidung der eineiigen und zweieiigen Zwillingsgeburten und mit den beiden bei Wiedergabe der Konstruktion *Horowicz's* unter 1 und 2 angeführten Annahmen verbindet, nur etwas komplizierter, ohne dass das Ergebnis irgendwie davon tangiert würde. Man erhält nämlich in diesem Fall Formel (29) nicht direkt, sondern findet zunächst als mittleren Fehler von y_i den Ausdruck

$$\sqrt{\frac{\left\{ e' p + e'' \left(p^2 + \frac{p q}{2} \right) \right\}}{n_i} - p^2},$$

woraus sodann vermöge der Substitutionen $e' = 1 - e''$ und $1 - p = q$ der Ausdruck

$$\sqrt{\frac{p q - \frac{e'' p q}{2}}{n_i}}$$

und schliesslich wegen (21) Formel (29) hervorgeht. Insbesondere erweist sich auch die an sich richtige Ansicht, dass die „rechnungsmässige“ Versuchszahl, wofür im obigen f_i geschrieben wurde, zwischen n_i und s_i liegt, als unabhängig davon, ob man die Tatsache der Geschlechtsattraktion so oder anders erklärt. Stellt man nämlich Formel (37), welche bei $\beta_0 > 0$ die Ungleichung $f_i > n_i$ ergibt, in der Form

¹⁾ Um Missverständnissen vorzubeugen, bemerke ich ausdrücklich, dass meine Kritik der von *Horowicz* zur Berechnung von Q benützten Methode nicht als Widerlegung seiner These, wonach von den Zwillingsgeburten 24.4 % als eineiige und 75.6 % als zweieiige anzusehen seien, gemeint ist. Diese These an sich geht den Wahrscheinlichkeitstheoretiker nichts an. Als solcher kann ich aber nicht umhin, auf einen groben Fehler aufmerksam zu machen, den sich *Horowicz* (S. 36) bei der Beurteilung eines für seine These nicht sehr günstigen statistischen Ergebnisses zuschulden kommen lässt. Er behauptet nämlich, dass die Abweichung, die sich da zwischen einem direkt ermittelten Prozentsatz zweieiiger Geburten (78.4) und 75.6 ergibt, mit Rücksicht auf die relative Kleinheit der beobachteten Fälle (876), sehr wohl eine zufällige sein könne, weil es sich ausrechnen lasse, „dass diese Abweichung mit der Wahrscheinlichkeit 0,94 vorkommen kann“. In Wirklichkeit ist 0,94 in diesem Fall die Wahrscheinlichkeit einer Abweichung, die (ihrem absoluten Betrag nach) hinter der festgestellten Abweichung zurückbleibt. Demnach ist die Wahrscheinlichkeit einer so grossen oder noch grösseren Abweichung, wie die festgestellte, 0,06, und eben nach dieser Wahrscheinlichkeit ist die in Frage stehende Unstimmigkeit zu beurteilen.

$$(40) \quad f_i = \frac{2 y_0 (1 - y_0) s_i}{2 y_0 (1 - y_0) + \{2 y_0 (1 - y_0) - \beta_0\}}$$

dar, so sieht man unmittelbar ein, dass $f_i < s_i$, wenn die Ungleichungen (11) erfüllt sind. Wäre aber umgekehrt β_0 grösser als $2 y_0 (1 - y_0)$, läge somit „Geschlechtsrepulsion“ vor, so würde f_i nach oben nicht begrenzt sein. Im Grenzfall, d. h. bei $\beta_0 = 1$, wäre $y_0 = \frac{1}{2}$ und $f_i = \infty$; die wirkliche, sowohl wie die theoretische mittlere Abweichung, würde gleich Null sein, und man erhielte demgemäss $Q = \frac{0}{0}$.

Es ist ein leichtes, die theoretische mittlere Abweichung für den Fall der Knabenquote bei Mehrlingsgeburten einer beliebigen „Klasse“ abzuleiten. Man bezeichne mit m die Zahl der Kinder, die aus je einer Mehrlingsgeburt hervorgehen. Die Wahrscheinlichkeit der Geschlechtskombination k Knaben und $m - k$ Mädchen soll (wie im besonderen Fall der Zwillingsgeburten) mit w_k und die Zahl der auf den i^{ten} Zeitabschnitt entfallenden Mehrlingsgeburten von der Geschlechtszusammensetzung k Knaben und $m - k$ Mädchen mit $l_{k,i}$ bezeichnet werden. Man hat demnach: $s_i = m n_i$, $S = m N$, $\sum_{k=0}^m w_k = 1$, $\sum_{k=0}^m l_{k,i} = n_i$, $\sum_{k=0}^m k l_{k,i} = x_i$ und

$$(41) \quad p = \sum_{k=0}^m \frac{k w_k}{m}.$$

Führt man noch die Bezeichnungen $\lambda_{k,i}$ für den Quotienten $l_{k,i} : n_i$ und L_k für die Summe $\sum_{i=1}^z l_{k,i}$ ein, so lässt sich das gewogene arithmetische Mittel der z -Werte, die $\lambda_{k,i}$ bei einem gegebenen k für $i = 1$ bis z annimmt, durch

$$(42) \quad \lambda_{k,0} = \frac{L_k}{N}$$

ausdrücken.

Um die mittlere Abweichung von y_i zu bestimmen, soll auch jetzt von ϑ_j ausgegangen werden, wobei der Sinn dieses Symbols, wie es im Wesen der neuen Aufgabe liegt, dahin generalisiert werden muss, dass ϑ_j die Knabenquote bei der j^{ten} Mehrlingsgeburt (statt Zwillingsgeburt) in der Reihe der n_i Mehrlingsgeburten zu bedeuten hat. Da ϑ_j dementsprechend $m + 1$ verschiedene Werte, nämlich die Werte $0, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}, 1$ annehmen kann, denen die Wahrscheinlichkeiten $w_0, w_1, w_2, \dots, w_{m-1}, w_m$ zukommen, so ist die mathe-

mathematische Erwartung von ϑ_j laut Formel (41) durch p und das Quadrat der theoretischen mittleren Abweichung von ϑ_j durch

$$(43) \quad \frac{1}{m^2} \sum_{k=0}^m k^2 w_k - p^2$$

gegeben. Wegen (41) sind aber die beiden Ausdrücke p und $\frac{1}{m^2} \sum_{k=0}^m k m w_k$ identisch. Folglich lässt sich (43) auch in der Form

$$p - p^2 - \left(\frac{1}{m^2} \sum_{k=0}^m k m w_k - \frac{1}{m^2} \sum_{k=0}^m k^2 w_k \right)$$

oder, was dasselbe ist, in der Form

$$(44) \quad p q - \frac{1}{m^2} \sum_{k=1}^{m-1} k (m-k) w_k$$

darstellen. Mithin erhält man als theoretische mittlere Abweichung der Grösse y_i , die auch jetzt nichts anderes als das arithmetische Mittel der n_i -Grössen ϑ_j ist:

$$(45) \quad \sqrt{\frac{p q - \frac{1}{m^2} \sum_{k=1}^{m-1} k (m-k) w_k}{n_i}}$$

oder näherungsweise

$$(46) \quad \sqrt{\frac{y_0 (1 - y_0) - \frac{1}{m^2} \sum_{k=1}^{m-1} k (m-k) \lambda_{k,0}}{n_i}},$$

und die verallgemeinerten Formeln (33) und (34) sind:

$$(47) \quad Q^2 = \frac{n_0 \sum_{i=1}^z (y_i - y_0)^2}{(z-1) \left\{ y_0 (1 - y_0) - \frac{1}{m^2} \sum_{k=1}^{m-1} k (m-k) \lambda_{k,0} \right\}}$$

und

$$(48) \quad Q^2 = \frac{z \sum_{i=1}^z (x_i - y_0 s_i)^2}{(z-1) \left\{ m X (1 - y_0) - \sum_{k=1}^{m-1} k (m-k) L_k \right\}}$$

Im besonderen Fall der Drillingsgeburten ist $m = 3$ zu setzen und man erhält aus (47) und (48):

$$(49) \quad Q^2 = \frac{n_0 \sum_{i=1}^z (y_i - y_0)^2}{(z-1) \left\{ y_0 (1 - y_0) - \frac{2}{9} (\lambda_{1,0} + \lambda_{2,0}) \right\}}$$

bzw.

$$(50) \quad Q^2 = \frac{z \sum_{i=1}^z (x_i - y_0 s_i)^2}{(z-1) \{ 3 X (1 - y_0) - 2 (L_1 + L_2) \}}$$

Es möge noch gezeigt werden, dass Formel (45) in die übliche für Einzelgeborene gültige Formel

$$(51) \quad \sqrt{\frac{p q}{s_i}}$$

übergeht, wenn die Mehrlinge in bezug auf das Geschlecht unabhängig voneinander sind. Formel (43) lässt sich wegen (41) auch in der Form

$$(52) \quad \frac{1}{m^2} \sum_{k=0}^m w_k (k - m p)^2$$

schreiben. Führt man ferner die Bezeichnung

$$(53) \quad \binom{m}{k} p^k q^{m-k} = w'_k$$

ein und beachtet man, dass

$$(54) \quad \sum_{k=0}^m w'_k (k - m p)^2 = m p q,$$

so findet man, dass (52), somit auch (41), mit

$$\frac{1}{m^2} \left\{ m p q + \sum_{k=0}^m (w_k - w'_k) (k - m p)^2 \right\}$$

oder

$$(55) \quad \frac{p q}{m} + \frac{1}{m^2} \sum_{k=0}^m (w_k - w'_k) (k - m p)^2$$

identisch ist. Folglich kann der mittlere Fehler von y_i wegen $m n_i = s_i$ ebensogut, wie durch (45), durch

$$(56) \quad \sqrt{\frac{p q}{s_i} + \frac{1}{m s_i} \sum_{k=0}^m (w_k - w'_k) (k - m p)^2}$$

ausgedrückt werden, woraus eben zu ersehen ist, dass man bei $w_k = w'_k$, d. h. wenn die Mehrlinge in bezug auf das Geschlecht unabhängig voneinander sind, auf Formel (51) zurückkommt.

Die hier gegebene Lösung der Frage, wie die Dispersion der Knabenquote bei Mehrlingsgeburten zu

bestimmen sei, beruht darauf, dass die Knabenquote (y_i) aufgefasst wird nicht als Quotient $\left(\frac{x_i}{s_i}\right)$, als welcher sie unmittelbar erscheint, auch nicht als arithmetisches Mittel von s_i Grössen, welche die Werte 1 und 0 annehmen (je nachdem der Neugeborene ein Knabe oder ein Mädchen ist), sondern als arithmetisches Mittel von n_i Grössen, welche, die Knabenquote bei je einer Mehrlingsgeburt darstellend, $m + 1$ verschiedene Werte $\frac{k}{m}$ annehmen¹⁾. Es gibt aber noch einen anderen Weg, der zu demselben Ziele, d. h. zu der grundlegenden Formel (45) führt, und dieser — etwas weitere — Weg ist dadurch gekennzeichnet, dass dabei die Versuchszahl nicht der Zahl der Geburten (n_i), sondern, wie im Fall der Einzelgeborenen, der Zahl der Geborenen (s_i) gleichgesetzt wird, mit der Massgabe jedoch, dass für je m unter den in Frage kommenden Versuchen eine gegenseitige Abhängigkeit angenommen wird, die nach Richtung und Mass in dem *Korrelationskoeffizienten* ihren Ausdruck findet.

Liegen je r verschiedene Werte U_h und V_h zweier Grössen U und V vor, so ist der Korrelationskoeffizient durch

$$r = \frac{\sum_{h=1}^r u_h v_h}{v \mu_1 \mu_2}$$

dargestellt, wo, unter Anwendung der abkürzenden Bezeichnungen

$$\frac{1}{v} \sum_{h=1}^r U_h = U_0, \quad \frac{1}{v} \sum_{h=1}^r V_h = V_0,$$

die Grössen u_h, v_h, μ_1, μ_2 durch

$$u_h = U_h - U_0, \quad v_h = V_h - V_0,$$

$$\mu_1 = \sqrt{\frac{1}{v} \sum_{h=1}^r u_h^2}, \quad \mu_2 = \sqrt{\frac{1}{v} \sum_{h=1}^r v_h^2}$$

¹⁾ Dass die *Lexische* Dispersionstheorie überhaupt auf jede statistische Grösse anwendbar ist, die unter den Begriff eines arithmetischen Mittels fällt, habe ich bereits in meinen „Kritischen Betrachtungen zur theoretischen Statistik“ (2. Artikel, Jahrbücher für Nationalökonomie und Statistik, 3. Folge, Bd. X, 1895, S. 332 bis 343) nachgewiesen. Vgl. meine Ausführungen in der Enzyklopädie der mathematischen Wissenschaften I D 4a, S. 835—836. Die Knabenquote bei Mehrlingsgeburten ist nichts anderes als ein Spezialfall, und so liessen sich denn auch die für diesen Spezialfall geltenden Formeln leicht aus den von mir damals aufgestellten — höchst einfachen — allgemeinen Formeln herleiten. *Czuber* (Wahrscheinlichkeitsrechnung usw., II, 2. Auflage, S. 77—78) spricht unter Bezugnahme auf mich von einer „Ausdehnung des Dispersionsbegriffs auf extensive Grössen“. Es müsste eigentlich heissen: „auf beliebige statistische Durchschnittszahlen“.

definiert sind. Im Fall der Mehrlingsgeburten ist U_h und V_h gleich 1 oder gleich 0 zu setzen, je nachdem das betreffende Kind ein Knabe oder ein Mädchen ist (so dass U_h bzw. V_h die Zahl der Knaben ist, die auf *einen* Geborenen kommt), wobei r zunächst nach Massgabe nicht irgendeiner empirischen, sondern der theoretischen Häufigkeit des Vorkommens der verschiedenen Geschlechtskombinationen bestimmt werden soll. Dementsprechend hat man $U_0 = V_0 = p$ und u_h bzw. $v_h = q$ oder u_h bzw. $v_h = -p$, je nachdem es sich um einen Knaben oder um ein Mädchen handelt; zugleich ergibt sich (bei einem beliebigen v) $\mu_1 = \mu_2 = \sqrt{p q^2 + q p^2} = \sqrt{p q}$, somit $\mu_1 \mu_2 = p q$. Liegt nun eine Mehrlingsgeburt mit k Knaben und $m-k$ Mädchen vor, so lassen sich $\frac{m(m-1)}{2}$ Produkte $u_h v_h$ bilden, von denen $\frac{k(k-1)}{2}$ den Wert q^2 , $\frac{(m-k)(m-k-1)}{2}$ den Wert p^2 und $k(m-k)$ den Wert $-p q$ haben. Als arithmetisches Mittel dieser $\frac{m(m-1)}{2}$ Produkte $u_h v_h$ erhält man

$$\frac{k(k-1)q^2 + (m-k)(m-k-1)p^2 - 2k(m-k)pq}{m(m-1)}$$

oder auch

$$\frac{(k-m p)^2 + k(p-q) - m p^2}{m(m-1)}$$

Um nunmehr r zu bestimmen, braucht man nur diese Ausdrücke mit w_k zu multiplizieren, die so erhaltenen Produkte nach k in den Grenzen von 0 bis m zu summieren und die gefundene Summe durch $p q$ zu dividieren. Man gelangt auf diese Weise, unter Berücksichtigung von (41), zu

$$(57) \quad r = \frac{\sum_{k=0}^m w_k (k-m p)^2 - m p q}{m(m-1) p q}$$

und erhält $r = 1$ bei $w_0 = q$, $w_1 = w_2 = \dots = w_{m-1} = 0$, $w_m = p$ und $r = 0$ bei $w_k = w'_k$. Die untere Grenze von r ist $-\frac{1}{m-1}$ (sie kommt nur unter der Bedingung, dass das Produkt $m p$ eine ganze Zahl ist, in Frage), somit im Fall der Zwillingsgeburten -1 . Hier ergibt sich $r = -1$, wenn $w_0 = w_2 = 0$, w_1 und dementsprechend $p = \frac{1}{2}$.

Nach einem bekannten Lehrsatz der mathematischen Statistik erhöht oder vermindert sich die mittlere Abweichung einer Grösse dadurch, dass zwischen je m Elementen der statistischen Masse, auf die sie sich

bezieht, eine gegenseitige Abhängigkeit (Korrelation) besteht, im Verhältnis von $\sqrt{1 + (m-1)r}$ zu 1¹⁾. Demgemäss erhält man im Falle der Knabenquote bei Mehrlingsgeburten als theoretische mittlere Abweichung an Stelle des Ausdrucks (51) den Ausdruck

$$(58) \quad \sqrt{\frac{p q \{1 + (m-1)r\}}{s_i}}$$

Aus (57) ergibt sich aber

$$1 + (m-1)r = \frac{\sum_{k=0}^m w_k (k-m p)^2}{m p q},$$

so dass (57) zunächst in

$$(59) \quad \sqrt{\frac{\sum_{k=0}^m w_k (k-m p)^2}{m s_i}},$$

sodann wegen $s_i = m n_i$ und wegen

$$\sum_{k=0}^m w_k (k-m p)^2 = \sum_{k=0}^m k^2 w_k - m^2 p^2$$

in

$$\sqrt{\frac{\frac{1}{m^2} \sum_{k=0}^m k^2 w_k - p^2}{n_i}}$$

und schliesslich auf Grund der bei Ableitung von (44) aus (43) angegebenen Umformung in (45) übergeht.

Nach dem Vorstehenden kann (57) auch in einer der beiden Formen

$$r = \frac{\sum_{k=0}^m (w_k - w'_k) (k-m p)^2}{m(m-1) p q}$$

oder

$$r = 1 - \frac{\sum_{k=1}^{m-1} k(m-k) w_k}{m(m-1) p q}$$

geschrieben werden. Ersetzt man aber in letzterer Formel w_k durch $\lambda_{k,0}$ und p bzw. q durch y_0 bzw. $1-y_0$, so erhält man als empirischen Wert von r

¹⁾ G. Udny Yule, An introduction to the theory of statistics. London 1911, S. 207-208, 282, 346.

$$(60) \quad r = 1 - \frac{\sum_{k=1}^m k(m-k)\lambda_{k,0}}{m(m-1)y_0(1-y_0)},$$

somit im Fall der Zwillingengeburt

$$(61) \quad r = \frac{2y_0(1-y_0) - \beta_0}{2y_0(1-y_0)}$$

und im Fall der Drillingsgeburt

$$(62) \quad r = \frac{3y_0(1-y_0) - (\lambda_{1,0} + \lambda_{2,0})}{3y_0(1-y_0)}.$$

Beispiele.

A. Deutsches Reich 1901—1911.

I. *Zwillinge*. Hierzu die Tabellen 1 und 2. Man erhält hier den in der letzten Zeile der Tabelle 1 aufgeführten Zahlen zufolge: $\alpha_0 = 0,3211$, $\beta_0 = 0,3771$, $\gamma_0 = 0,3018$, $y_0 = 0,5097$, und nach Formel (34) ergibt sich

$$Q^2 = \frac{11 \cdot 181678}{10(2 \cdot 290737 \cdot 0,4903 - 107567)} = 1,126,$$

woraus $Q = 1,061$ folgt. Demgegenüber erhält man, wenn man mit $E(Q)$ die mathematische Erwartung von Q bezeichnet, $E(Q) = 1 - \frac{1}{4 \cdot 10} = 0,975$, so dass $Q - E(Q) = 0,086$ und $\eta = 0,086 : \frac{1}{\sqrt{20}} = 0,38$. Formel (33) führt zu $Q = 1,072$, $\eta = 0,43$. Nach Formel (61) findet man $r = 0,2455$.

II. *Drillinge*. Hierzu die Tabellen 3 und 4. Die Zahlen der letzten Zeile in Tabelle 3 liefern die Werte: $\lambda_{3,0} = 0,2232$, $\lambda_{2,0} = 0,2492$, $\lambda_{1,0} = 0,2795$, $\lambda_{0,0} = 0,2481$, $y_0 = 0,4825$, und der Formel (50) zufolge erhält man:

$$Q^2 = \frac{11 \cdot 2849,24}{10(3 \cdot 4066 \cdot 0,5175 - 2 \cdot 1485)} = 0,937$$

bzw. $Q = 0,968$. Man hat auch hier: $E(Q) = 0,975$, daher $Q - E(Q) = -0,007$ und $\eta = 0,007 : \frac{1}{\sqrt{20}} = 0,03$. Formel (62) ergibt $r = 0,2942$.

B. Berlin 1879—1911.

I. *Zwillinge*. Hierzu die Tabellen 5 und 6. In derselben Weise wie in dem Beispiel A I findet man:

$\alpha_0 = 0,3284$, $\beta_0 = 0,3563$, $\gamma_0 = 0,3153$, $y_0 = 0,5065$, ferner nach Formel (34) $Q^2 = 1,077$, $Q = 1,038$, alsdann $E(Q) = 1 - \frac{1}{4 \cdot 32} = 0,992$, $Q - E(Q) = 0,046$,

$\eta = 0,046 : \frac{1}{\sqrt{64}} = 0,37$ und schliesslich nach Formel

$$(61) \quad r = 0,2972.$$

II. *Drillinge*. Hierzu Tabelle 7. In derselben Weise wie in dem Beispiel A II erhält man: $\lambda_{3,0} = 0,272$, $\lambda_{2,0} = 0,171$, $\lambda_{1,0} = 0,272$, $\lambda_{0,0} = 0,285$, $y_0 = 0,477$. Zur Berechnung von Q^2 empfiehlt es sich hier, mit Rücksicht auf die Kleinheit der betreffenden absoluten Zahlen, den Zähler des durch Formel (50) gegebenen Ausdrucks von Q^2 in der Form

$$z \left(\sum_{i=1}^z x_i^2 - 6y_0 \sum_{i=1}^z n_i x_i + 9y_0^2 \sum_{i=1}^z n_i^2 \right)$$

darzustellen. Man findet demnach:

$$Q^2 = \frac{33 \left\{ 2038 - 6 \cdot \frac{226}{474} \cdot 1292 + 9 \cdot \left(\frac{226}{474} \right)^2 \cdot 920 \right\}}{32 \left(3 \cdot 226 \cdot \frac{248}{474} - 2 \cdot 70 \right)} = 1,077$$

und $Q = 1,038$. Weil aber $E(Q) = 1 - \frac{1}{4 \cdot 32} = 0,992$, so hat man $Q - E(Q) = 0,046$ und $\eta = 0,046 : \frac{1}{\sqrt{64}} = 0,37$. Formel (62) ergibt $r = 0,408$.

Als arithmetisches Mittel der vier Werte von Q^2 , die sich nach den Formeln (34) und (50) in obigen Beispielen ergeben haben, erhält man 1,054. Der mittlere Fehler dieser Grösse ist durch $\frac{1}{2} \sqrt{\frac{1}{2} \left(\frac{2}{10} + \frac{2}{32} \right)} = 0,181$ gegeben. Somit beträgt die Abweichung des festgestellten arithmetischen Mittels von dessen mathematischer Erwartung, welche 1 ist, 0,054 oder 30% des zugehörigen mittleren Fehlers.

Die mitgeteilten Ergebnisse sind, wie man sieht, an und für sich mit der Annahme einer normalen Dispersion nicht unvereinbar. Möglicherweise liegt aber eine schwach ausgesprochene übernormale Dispersion vor.

Es handelt sich hier um den ersten Fall einer korrekten Untersuchung der Dispersion einer statistischen Grösse von nicht ganz so einfacher Struktur, wie diejenigen Grössen es sind, auf welche die *Lexische* Dispersionstheorie in ihrer ursprünglichen Form eingestellt gewesen ist.

Tabellen.

Tabelle 1.
Deutsches Reich 1901—1911. Zwillinge.

i	a_i	b_i	c_i	n_i
1	2	3	4	5
1	8 519	9 791	7 817	26 127
2	8 353	9 786	7 829	25 978
3	8 309	9 831	7 853	25 993
4	8 558	10 113	8 080	26 751
5	8 421	9 664	7 847	25 932
6	8 444	10 173	7 918	26 535
7	8 345	9 858	7 769	25 972
8	8 538	9 933	7 843	26 314
9	8 236	9 716	7 941	25 893
10	8 057	9 414	7 614	25 085
11	7 805	9 288	7 552	24 645
1—11	91 585	107 567	86 073	285 225

Tabelle 2.
Deutsches Reich 1901—1911. Zwillinge.

i	x_i	$s_i y_0$	$x_i - s_i y_0$		$(x_i - s_i y_0)^2$
			—	+	
1	2	3	4	5	6
1	26 829	26 632	.	197	38 809
2	26 492	26 480	.	12	144
3	26 449	26 495	46	.	2 116
4	27 229	27 268	39	.	1 521
5	26 506	26 433	.	73	5 329
6	27 061	27 048	.	13	169
7	26 548	26 474	.	74	5 476
8	27 009	26 823	.	186	34 596
9	26 188	26 393	205	.	42 025
10	25 528	25 570	42	.	1 764
11	24 898	25 121	223	.	49 729
1—11	290 737	290 737	555	555	181 678

Tabelle 3.
Deutsches Reich 1901—1911. Drillinge.

i	$l_{3,i}$	$l_{2,i}$	$l_{1,i}$	$l_{0,i}$	n_i
1	2	3	4	5	6
1	47	64	72	58	241
2	59	54	83	81	277
3	48	66	90	66	270
4	69	70	77	75	291
5	61	56	68	56	241
6	62	77	69	58	266
7	59	52	60	61	232
8	56	72	80	53	261
9	61	67	68	65	261
10	57	60	57	69	243
11	48	62	61	55	226
1—11	627	700	785	697	2809

Tabelle 4.
Deutsches Reich 1901—1911. Drillinge.

i	x_i	$s_i y_0$	$x_i - s_i y_0$		$(x_i - s_i y_0)^2$
			—	+	
1	2	3	4	5	6
1	341	348,9	7,9	.	62,41
2	368	401,0	33,0	.	1089,00
3	366	390,8	24,8	.	615,04
4	424	421,1	.	2,9	8,41
5	363	348,9	.	14,1	198,81
6	409	385,0	.	24,0	576,00
7	341	335,8	.	5,2	27,04
8	392	377,8	.	14,2	201,64
9	385	377,8	.	7,2	51,84
10	348	351,8	3,8	.	14,44
11	329	327,1	.	1,9	3,61
1—11	4066	4066,0	69,5	69,5	2848,24

Tabelle 5.
Berlin 1879—1911. Zwillinge.

i	a_i	b_i	c_i	n_i
1	2	3	4	5
1	197	205	183	585
2	157	195	198	550
3	173	185	163	521
4	181	156	204	541
5	164	180	169	513
6	180	207	160	547
7	169	163	156	488
8	186	190	177	553
9	196	204	168	568
10	211	183	160	554
11	166	221	178	565
12	177	206	153	536
13	209	199	173	581
14	164	178	167	509
15	145	163	150	458
16	164	204	179	547
17	142	169	136	447
18	181	211	149	541
19	157	158	145	460
20	187	205	201	593
21	164	193	175	532
22	165	196	171	532
23	182	174	171	527
24	194	198	193	585
25	171	187	153	511
26	178	213	180	571
27	214	192	179	585
28	185	220	169	574
29	202	210	181	593
30	193	204	173	570
31	188	199	172	559
32	147	175	159	481
33	155	199	167	521
1—33	5844	6342	5612	17 798

Tabelle 6.
Berlin 1879—1911. Zwillinge.

i	x _i	s _i y ₀	x _i - s _i y ₀		(x _i - s _i y ₀) ²
			-	+	
1	599	592,6	.	6,4	40,96
2	509	557,2	48,2	.	2 323,24
3	531	527,8	.	3,2	10,24
4	518	548,1	30,1	.	906,01
5	508	519,7	11,7	.	136,89
6	567	554,1	.	12,9	166,41
7	501	494,4	.	6,6	43,56
8	562	560,2	.	1,8	3,24
9	596	575,4	.	20,6	424,36
10	605	561,2	.	43,8	1 918,44
11	553	572,4	19,4	.	376,36
12	560	543,0	.	17,0	289,00
13	617	588,6	.	28,4	806,56
14	506	515,6	9,6	.	92,16
15	453	464,0	11,0	.	121,00
16	532	554,1	22,1	.	488,41
17	453	452,8	.	0,2	0,04
18	573	548,1	.	24,9	620,01
19	472	466,0	.	6,0	36,00
20	579	600,7	21,7	.	470,89
21	521	538,9	17,9	.	320,41
22	526	538,9	12,9	.	166,41
23	538	533,9	.	4,1	16,81
24	586	592,6	6,6	.	43,56
25	529	517,7	.	11,3	127,69
26	569	578,4	9,4	.	88,36
27	620	592,6	.	27,4	750,76
28	590	581,5	.	8,5	72,25
29	614	600,7	.	13,3	176,89
30	590	577,4	.	12,6	158,76
31	575	566,3	.	8,7	75,69
32	469	487,3	18,3	.	334,89
33	509	527,8	18,8	.	353,44
1—33	18 030	18 030,0	257,7	257,7	11 959,70

Tabelle 7.
Berlin 1879—1911. Drillinge.

i	l _{3,i}	l _{2,i}	l _{1,i}	l _{0,i}	n _i	n _i ²	x _i	x _i ²	n _i x _i
2	4	2	1	1	8	64	17	289	136
3	4	—	—	—	4	16	12	144	48
4	—	—	1	—	1	1	1	1	1
5	1	2	2	2	7	49	9	81	63
6	—	3	—	2	5	25	6	36	30
7	2	1	—	2	5	25	8	64	40
8	—	—	1	1	2	4	1	1	2
9	1	2	2	—	5	25	9	81	45
10	4	—	1	—	5	25	13	169	65
11	—	2	2	1	5	25	6	36	30
12	1	—	—	1	2	4	3	9	6
13	1	1	1	2	5	25	6	36	30
14	1	1	—	—	2	4	5	25	10
15	1	1	2	1	5	25	7	49	35
16	2	1	2	4	9	81	10	100	90
17	1	1	2	2	6	36	7	49	42
18	1	1	2	2	6	36	7	49	42
19	2	—	—	1	3	9	6	36	18
20	2	—	—	3	5	25	6	36	30
21	2	—	3	1	6	36	9	81	54
22	3	2	1	1	7	49	14	196	98
23	1	1	4	—	6	36	9	81	54
24	—	1	6	3	10	100	8	64	80
25	1	—	2	1	4	16	5	25	20
26	2	—	2	4	8	64	8	64	64
27	2	1	2	2	7	49	10	100	70
28	1	2	—	1	4	16	7	49	28
29	—	1	1	1	3	9	3	9	9
30	—	—	1	1	2	4	1	1	2
31	—	1	—	1	2	4	2	4	4
32	1	—	—	1	2	4	3	9	6
33	—	—	—	2	2	4	—	—	—
1—33	43	27	43	45	158	920	226	2038	1292