ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 164, 2003

*Tadeusz Gerstenkorn**, *Joanna Gerstenkorn***

DIE GINI-MITTELDIFFERENZ IN DER STATISTISCHEN PRAXIS[1]. ANWENDUNG AUF EINIGE INFLATIONISTISCHE VERTEILUNGEN

Zusammenfassung. In der Arbeit sind einige interessante Eigenschaften der Gini-Mitteldifferenz dargastellt. Es wird in diesem Artikel die entsprechende Literatur zu diesem Thema zitiert (Artikel, Bücher). Die Anwendung der Mitteldifferenz auf die inflationische binomiale Verteilung wird ausführlich beschrieben, was für die Statistik interessant sein kann.

Stichwörter: Gini-Mitteldifferenz, Gini-Konzentrationskoeffizient, Lorenz-Konzentrationskurve, Mitteldifferenz der inflationistischen Verteilungen.

1. EINLEITUNG

Im 1911 hat Prof. Corrado G i n i das sehr umfangreiche statistische Studium *Variabilita e Mutabilita* in der ökonomisch-juristischen Zeitschrift „Studi Economico-Giuridici delle R. Universita di Cagliari“ (Band III, Teil II, S. 3–159) publiziert. Es ist bemerkenswert, daß spätere Autoren, die sich mit dem Problem der Mitteldifferenzy beschäftigten, sich nicht auf diese Arbeit bezogen haben. Im Buch von M. G. K e n d a l l und A. S t u a r t (1963) finden wir zwar den Namen Gini und im Literaturverzeichnis auch eine Gini-Arbeit, aber mit anderem Datum und ohne nähere Angaben. Man kann also vermuten, daß die Abhandlung schwierig zu erreichen war.

Die Zeit des I. Weltkriegs hat sicherlich die Verbreitung der Gini-Ideen behindert. Es ist uns nicht gelungen festzustellen, ob sich jemand während der zwanzigen Jahren mit der Mitteldifferenz beschäftigt hat. Vielleicht

* Professor, Fakultät für Mathematik der Universität Łódź.
** Doctor, Ökonomisch-Soziologischen Fakultät der Universität Łódź.

[1] Vortrag gehalten an der Europa-Universität Viadrina in Frankfurt (O) anläßlich der Gemeinsamen Tagung der Deutschen Statistischen Gesellschaft und der Polnischen Statistischen Gesellschaft vom 14.–16. Juni 2000.

genoß sie keine Popularität im Kreise der theoretischen Statistiker. Sie ist jedoch beachtenswert, weil sie, anders als viele Charakteristiken, die zum Messen der Streuung dienen, von dem beliebigen Zentralmaß unabhängig ist, wie man aus Definition der Mitteldifferenz erkennen kann:

$$\Delta_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x-y| dF(x) dF(y) \tag{1.1}$$

für stetige Zufallsvariable oder im Falle einer diskreten Zufallsvariable

$$\Delta_1 = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} |x_i - x_j| p_i p_j \tag{1.2}$$

wobei $p_i = P(X = x_i)$, $p_j = P(X = x_j)$ ist. Es wird darauf hingewiesen, daß für die Mitteldifferenz in der Literatur auch andere Bezeichnungen als Δ_1 verwendent werden.

Die analytische Untersuchung dieser Charakteristik ist wegen des in der Formel sich befindlichen absoluten Betrages ziemlich schwierig. Andererseits er leichtert dies Berechung bei realen Daten, wie dies auch bei der mittleren absoluten Abreichung von Mittelwert der Fall ist. Das ist vermutlich die Ursache, daß man auch solche Untersuchungen vornimmt, die mittlere Differenz zusammen mit der mittleren Abweichung zu betrachten. So ist das, z.B. in der Arbeit von T. A. Ramasubban (1958). Diese Arbeit stützt sich auf Methoden, die auch in den Arbeiten von N. L. Johnson (1957) und E. L. Crow (1958) verwendet werden. Die mittlere Abweichung für die binomiale Verteilung hat früher J. S. Frame (1945) untersucht.

Die Schwierigkeiten, die mit dem absuluten Betrag verbunden sind, kann man relativ leicht mit den unvollständigen Momenten überwinden. Das wurde in der Arbeit von T. Gerstenkorn (1975) gezeigt.

Die bisherigen Ergebnisse, die mittlere Differenz betreffen und mit Eigenschaften dieser Statistik, die mit der Stichprobe verbunden sind, scheinen nicht genügend sein. Jedoch kann man im Fall einer normalverteilten Zufallsvariablen einige interessante Resultate notieren. So hat vermutlich, für die Normalverteilung den genauen Standardfehler für die mittlere Differenz als erster U. S. Nair (1936) vorgelegt aber mit einer ziemlich komplizierten Methode. Viel später, schon nach dem II. Weltkrieg hat Z. A. Lomnicki (1952) dasselbe Resultat mit einfacheren Methode erlangt. Ein Jahr später hat A. R. Kamat (1953) das dritte Moment exakt berechnet und (aus der Betrachtung des Schiefemaßes β_1 entnommen), daß die Verteilung der mittleren Differenz für große n durch χ-Verteilung approximiert werden kann.

Kamat folgend hat T. A. Ramasubban (1956) ungefähre Werte für das vierte Moment erzielt und gezeigt, daß das Wölbungsmaß β_2, auf dieser Basis berechnet und mit den Werten für β_1 genommen, scheint die Genauigkeit χ-Approximation zu zeigen und mindestens für die Stichproben mit $n > 10$.

Derselbe Autor beabsichtigte die empirische Verteilung von Δ_1 für kleinere Stichproben ($n < 10$) zu erzielen, wir wissen allerdings nicht, ob das Resultat veröffentlicht wurde.

Die Normalverteilung kann man unter gewissen Voraussetzungen als Grenzverteilung der Binomialverteilung oder der Poisson-Verteilung betrachten. Es ist also ganz natürlich, die Stichprobenmomente der Mitteldifferenz für diese diskreten Verteilungen zu bestimmen, sogar wenn man die genaue Verteilung nicht finden kann. Um diese Arbeit zu erleichtern, leitet man im ersten Schritt Formeln für die mittlere absolute Differenz, d.h. für

$$\Delta_r = \sum_i \sum_{j \neq i} |r_i - r_j|^r p_i p_j \tag{1.3}$$

her.

Dieses Problem ist in der Arbeit von T. A. Ramasubban (1959) besprochen. Die Erweiterung dieses Problems kann man in der Arbeit von S. K. Katti (1960) finden.

Die Mitteldifferenz hat keine größere Anerkennung durch die Textbuchautoren gefunden. Wir haben sie in keinem polnischen Handbuch angetroffen. Sie wird kompakt in der Monographie von M. G. Kendall und A. Stuart (1963) besprochen. In der deutschen Literatur sehen wir die mittlere Differenz im Textbuch von Prof. H. Rinne (1974), wo die praktische Berechnung gezeigt wird.

2. EIGENSCHAFTEN DER MITTLEREN DIFFERENZ

Wenn die diskrete Zufallsvariable nur eine endliche Zahl von Werten annimmt, dann schreiben wir die Mitteldifferenz in der Form

$$\Delta_1 = \frac{1}{N^2} \sum_{i=1}^{l} \sum_{j=1}^{l} |x_i - x_j| n_i n_j, \tag{2.1}$$

wobei $\frac{n_i}{N} \approx p_i = P(X = x_i)$, $\frac{n_j}{N} \approx p_j = P(X = x_j)$ und $n_1 + n_2 + ... + n_l = N$. Das ist die sog. Differenz mit Wiederholung.

Manchmal definieren wir die Mitteldifferenz etwas anders:

$$\Delta_1 = \frac{1}{N(N-1)} \sum_{i=1}^{l} \sum_{j=1}^{l} |x_i - x_j| n_i n_j, \quad i \neq j \tag{2.2}$$

Das ist die sog. Differenz ohne Wiederholung.

Manchmal werden die obigen Formeln ohne die Gewichte n_i, n_j geschrieben, d, h$\cdot n_i = n_j = 1$ für alle i, j und damit wird $l = N$, sowie

$$\Delta_1 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - x_j|, \quad i, j = 1, 2, \ldots, N, \tag{2.1a}$$

$$\Delta_1 = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - x_j|, \quad i \neq j. \tag{2.2a}$$

Am meisten bedienen wir uns der folgenden Formel

$$\Delta_1 = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} |x_i - x_j| \tag{2.2b}$$

oder

$$\Delta_1 = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=1}^{j-1} |x_i - x_j|. \tag{2.2c}$$

Diese Formeln ilustrieren wir durch ein Beispiel aus dem Buch von H. Rinne (S. 119). $\sum_{v=1}^{n} \sum_{\substack{k=1 \\ k>v}}^{n} |x_v - x_k|$

x_0 \ x_k	1	5	6	6	8	10	13	13	16	22	$\sum_{k=v+1}^{n} \lvert x_v - x_k \rvert$
1	–	4	5	5	7	9	12	12	15	21	90
5	–	–	1	1	3	5	8	8	11	17	54
6	–	–	–	0	2	4	7	7	10	16	46
6	–	–	–	–	2	4	7	7	10	16	46
8	–	–	–	–	–	2	5	5	8	14	34
10	–	–	–	–	–	–	3	3	6	12	24
13	–	–	–	–	–	–	–	0	3	9	12
13	–	–	–	–	–	–	–	–	3	9	12
16	–	–	–	–	–	–	–	–	–	6	6
22	–	–	–	–	–	–	–	–	–	–	–
$\sum_{v=1}^{k-1} \lvert x_v - x_k \rvert$	–	4	6	6	14	24	42	42	66	120	324

Aus der Tabelle erhalten wir

$$\Delta_1 = \frac{2}{10 \cdot 9} 324 = 7{,}2.$$

Es könnte scheinen, daß die Schwierigkeiten mit dem absoluten Betrag verschwinden werden, wenn man an Stelle der Mitteldifferenz den Koeffizient

$$E^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-y)^2 dF(x) dF(y)$$

einführt.

Nach einer einfachen Berechnung erhält man jedoch

$$E^2 = 2\mu_2,$$

d.h. die doppelte Varianz. Dieser interessante Zusammenhang zeigt nämlich, daß man die Varianz als die Hälfte von E^2 definieren kann, d.h. ohne die Notwendigkeit, die Abweichungen hinsichtlich des Zentralwerts (des Mittelwerts) zu betrachten. Diese Bemerkung haben wir nur bei M. G. Kendall und A. Stuart (1963, S. 47, russ. Ausgabe S. 74) gefunden.

Die Darstellung der Formel (2.1) kann man jedoch so modifizieren, daß darin kein Absolutbetrag mehr vorkommt. Bemerken wir, daß

$$\sum_{i=1}^{l} \sum_{j=1}^{l} |x_i - x_j| n_i n_j = 2 \sum_{i=j+1}^{l} \sum_{j=1}^{l-1} |x_i - x_j| n_i n_j \tag{2.3}$$

gilt.

Wir nehmen an, die Beobachtungen x_i seinen so nummeriert, daß

$$x_1 < x_2 < \ldots < x_N.$$

Dann kann man die Formel (2.3) in der folgenden Form schreiben:

$$2 \sum_{i=j+1}^{l} \sum_{j=1}^{l-1} (x_i - x_j) n_i n_j. \tag{2.3a}$$

Die Formeln (2.1) und (2.2) nehmen dann die folgende Form an:

$$\Delta_1 = \frac{2}{N^2} \sum_{i=j+1}^{l} \sum_{j=1}^{l-1} (x_i - x_j) n_i n_j, \tag{2.4}$$

$$\Delta_1 = \frac{2}{N(N-1)} \sum_{i=j+1}^{l} \sum_{j=1}^{l-1} (x_i - x_j) n_i n_j. \tag{2.5}$$

Den Formeln (2.1a) und (2.2a) kann man noch eine andere Gestalt geben. Wir bemerken, daß wir nach sorgfältiger Berechnung folgende Ausdrücke bekommen können

$$\sum_{i=j+1}^{N} \sum_{j=1}^{N-1} (x_i - x_j) = \sum_{k=1}^{N-1} k(N-k)(x_{k+1} - x_k),$$

weswegen wir die mittlere Differenz in der Form notieren:

$$\Delta_1 = \frac{2}{N^2} \sum_{k=1}^{N-1} k(N-k)(x_{k+1} - x_k) \tag{2.6}$$

oder

$$\Delta_1 = \frac{2}{N(N-1)} \sum_{k=1}^{N-1} k(N-k)(x_{k+1} - x_k) \tag{2.7}$$

schreiben können.

Diese Formeln sind bei äquidistanten Abständen, also Konstanten $x_{k+1} - x_k$, besonderes bequem.

Man kann eine weitere Vereinfachung der Formeln erlangen, wenn wir die Verteilungsfunktion

$$F_k = P(X \leqslant x_k) = F(x_k)$$

einführen.

Im Fall, wenn $F_k = \frac{k}{N}$ ist und die Abstände einander gleich und gleich eins sind, erhalten wir

$$\Delta_1 = \frac{2}{N^2} \sum_{k=1}^{N-1} NF_k(N - NF_k) = 2 \sum_{k=1}^{N-1} F_k(1 - F_k). \tag{2.8}$$

Wenn wir mit $G_k = NF_k$ die kummulative Häufigkeit dann erhalten wir

$$\Delta_1 = \frac{2}{N^2} \sum_{k=1}^{N-1} G_k(N - G_k). \tag{2.9}$$

Diese Formel ist für die praktische Berechnung bequem. Dies zeigt die Tabelle von M. G. Kendall und A. Stuart, (S. 50–51, russ. Ausg. S. 78).

Height, inches	Frequency	G_h	$N - G_h$	$G_hN - G_h$
57–	2	2	8 583	17 166
58–	4	6	8 579	51 474
59–	14	20	8 565	171 300
60–	41	61	8 524	519 964
61–	83	144	8 441	1 215 504
62–	169	313	8 272	2 589 136
63–	394	707	7 878	5 569 746
64–	669	1 376	7 209	9 919 584
65–	990	2 366	6 219	14 714 154
66–	1 223	3 589	4 996	17 930 644
67–	1 329	4 918	3 667	18 034 306
68–	1 230	6 148	2 437	14 982 676
69–	1 063	7 211	1 374	9 907 914
70–	646	7 857	728	5 719 896
71–	392	8 249	336	2 771 664
72–	202	8 451	134	1 132 434
73–	79	8 530	55	469 150
74–	32	8 562	23	196 926
75–	16	8 578	7	60 046
76–	5	8 583	2	17 166
77–	2	8 585	–	–
Totals	8 585	–	–	105 990 850

Aus der Tabelle erhalten wir

$$\Delta_1 = \frac{2 \cdot 105\ 990\ 850}{8\ 585^2} = 2{,}88.$$

Für die Vollständigkeit der Darstellung ist es zweckmäßig den sog. Gini-Konzentrationskoeffizient zu erwähnen

$$G = \frac{\Delta_1}{2m}, \quad m = E(X), \text{ wenn er existiert,}$$

oder

$$G = \frac{\Delta_1}{2\bar{x}},$$

eine offenbar dimensionslose Größe.

In der statistischen Praxis benutzen wir auch die sog. Lorenz-Konzentrationskurve (1905). Das ist die Kurve mit Punkten $(F(x), \Phi(x))$, wobei

$$\Phi(x) = \frac{1}{m} \int_{-\infty}^{x} x dF(x)$$

das unvollständige Moment darstellt $(0 \leqslant \Phi(x) \leqslant 1)$.

Die Kurve ist konvex. Man kann zeigen, daß der Flächeninhalt S zwischen der Kurve und der Geraden $\Phi = F$ gleich $\frac{1}{2}G$ ist. Den Beweis kann man bei M. G. Kendall und A. Stuart finden (S. 49, russ. Ausg. SS. 76–77). Im Buch von H. Rinne ist ein schönes Beispiel für die Anwendung dieser Kurve (S. 146) zu finden.

3. DIE MITTELDIFFERENZ FÜR DIE INFLATIONISTISCHE BINOMIALE VERTEILUNG

In vielen wissenschaftlichen Disziplinen benutzt man bekannte diskrete Verteilungen. Es treten jedoch Situationen auf, in welchen wir bereit sind zu anerkennen, daß eine Erscheinung der typischen Verteilung folgt, jedoch unter der Bedingung, daß wir diese Verteilung einer Modifikation unterwerfen. Am meisten benutzen wir in solchem Fall die sog. Mischung von Verteilungen. Als einfachste können wir die sog. inflationistische Verteilung darstellen. Sie besteht aufs der Faltung einer beliebigen diskreten Verteilung mit der degenerierten, d.h. 1-Punktverteilung.

Wir führen die folgende Bezeichnung für eine diskrete Verteilung ein:

$$P(X = i) = h(i), \quad i = 0, 1, 2, \ldots$$

Dann haben wir:

Definition 3.1. Wir sagen, daß die diskrete Zufallsvariable Y einer **inflationistischen Verteilung** folgt (mit Deformierung im Punkt $i = 0$), wenn ihre Wahrscheinlichkeitsfunktion die folgende Form hat:

$$P(Y = i) = \begin{cases} \beta + \alpha h(0) & \text{für} \quad i = 0, \\ \alpha\, h(i) & \text{für} \quad i > 1, \end{cases} \tag{3.1}$$

wobei $\alpha \in (0, 1]$ und $\beta = 1 - \alpha$.

Die Deformierung der Verteilung kann auch an beliebiger Stelle stattfinden. Dann gelte:

Definition 3.2. Wir sagen, daß die diskrete Zufallsvariable Y einer **verallgemeinerten inflationistischen** Verteilung folgt (mit Deformierung im Punkt $i = l$), wenn

$$P(Y=i)=\begin{cases}\beta+a\,h(l) & \text{für} \quad i=l,\\ a\,h(i) & \text{für} \quad i=0,\ 2,\ \ldots,\ l-1,\ l+1,\ \ldots,\ n,\end{cases} \tag{3.2}$$

wobei $a \in (0, 1]$ und $\beta = 1 - a$.

Im Spezialfall der Binomialverteilung haben wir:

Definition 3.3. Wir sagen, daß die Zufallsvariable Y der inflationistischen Binomialverteilung $P(X = i)$ (deformiert im Punkt $i = 0$) folgt, wenn ihre Wahrscheinlichkeitsfunktion durch die folgende Formel gegeben ist:

$$P(Y=i)=\begin{cases}\beta+a\,q^n & \text{für} \quad i=0\\ a\binom{n}{i}p^i q^{n-i} & \text{für} \quad i=1,\ 2,\ \ldots,\ n,\end{cases} \tag{3.3}$$

wobei $a \in (0, 1]$ und $\beta = 1 - a$, $0 < p < 1$, $p + q = 1$.

Wenn $a = 1$ ist, dann führt die obige Verteilung auf die binomiale Verteilung zurück:

$$P(X=i)=\binom{n}{i}p^i q^{n-i} \quad \text{für} \quad i=0,\ 1,\ 2,\ \ldots,\ n.$$

Definition 3.4. Wir sagen, daß die diskrete Zufallsvariable Y der verallgemeinerten inflationistischen Binomialverteilung folgt, wenn ihre Wahrscheinlichkeitsfunktion durch die folgende Formel gegeben ist

$$P(Y=i)=\begin{cases}\beta+a\binom{n}{l}p^l q^{n-l} & \text{für} \quad i=l,\\ a\binom{n}{i}p^i q^{n-i} & \text{für} \quad i=0,\ 1,\ 2,\ \ldots,\ l-1,\ l+1,\ \ldots,\ n,\end{cases} \tag{3.4}$$

wobei $0 < a \leqslant 1$, $a + \beta = 1$, $0 < p < 1$, $p + q = 1$.

Die Formeln (3.1)–(3.4) sind natürlich Wahrscheinlichkeitsverteilungen, was aus einer einfachen Berechnung folgt.

Die inflationistischen Verteilungen sind in die Literatur von S. N. Singh (1963) für den Poissonschen Fall eingeführt und danach von M. P. Singh für die Binomialverteilung, deformiert im Punkt $i = 0$ (1965/1966) und in einem beliebigen (1966), untersucht worden.

Mit inflationistischen Verteilungen haben sich viele Autoren beschäftigt. Es sind viele Arbeiten über verschiedene Themen entstanden. Die Problematik dieser Verteilungen wurde in einem Artikel von T. Gerstenkorn (1977) besprochen.

Die Mitteldifferenz für die inflationistische Binomialverteilung (3.3) wurde in einem Artikel von T. Gerstenkorn (1997) dargestellt.

Hier stellen wir die Mitteldifferenz für die verallgemeinerte inflationistische Binomialverteilung (3.4) dar. So haben wir:

Satz 3.1. Für die Mitteldifferenz der verallgemeinerten inflationistischen Verteilung ergibt sich.

$$\Delta_1 = 2\alpha\beta l\,[2F(l+1)-1] - 2\alpha\beta m_1 + 4\alpha\beta m_l(l+1) +$$

$$+ 2\alpha^2 \sum_{j=1}^{} \sum_{i=0}^{j-1} (j-i)h(i)h(j) - 2\alpha^2 \sum_{j=1}^{l-1} \sum_{i=0}^{j-1} (j-i)h(i)h(j), \tag{3.5}$$

wobei die darin auftretenden Größen folgendes bedeuten:

m_1 – Erwartungswert der Verteilung ohne Deformierung;

$m_1(l+1)$ – unvollständiges rechtsseitige Moment, d.h. mit dem Abschneiden des Zufallswertes bis $x = l$ einschließlich für die Verteilung ohne Deformierung;

$F(l+1)$ – Verteilungsfunktion der Verteilung ohne Deformierung im Punkt $x = l+1$.

Der Beweis ist ziemlich lang und rechnerisch kompliziert, er wird daher er hier nicht dargestellt. Statt dessen beweisen wir die entsprechenden Formeln für die inflationistische binomiale Verteilung (3.3) und (3.4) mit Hilfe der Formel (3.5) und einiger Relationen aus der Arbeit von T. Ramasubban ((2.4), S. 550 oder (2.8), S. 550) und zusätzlich der Formel (1.14) für das unvollständige Moment der Binomialverteilung, das in der Arbeit von T. Gerstenkorn (1971) nach R. Risser und C. E. Traynard (S. 320–321 oder 1957, S. 92–93) zitiert ist. Nämlich ist

$$m_1(l+1) = (l+1)\binom{n}{l+1} p^{l+1} q^{n-l} + np \cdot m_0(l+1).$$

Das obige in Betracht ziehend, haben wir

$$\Delta_1 = 4\alpha\beta\, lF(l+1) - 2\alpha\beta l - 2\alpha\beta np + 4\alpha\beta \left[(l+1)\binom{n}{l+1} p^{l+1} q^{n-l} + np \cdot m_0(l+1) \right] +$$

$$+ 2\alpha^2 npq \left[\sum_{i=1}^{n-1} \binom{n-1}{j-1}\binom{n-1}{i} p^{2i-1} q^{2n-2i-1} + \sum_{i=0}^{n-1} \binom{n-1}{i}^2 p^{2i} q^{2n-2i-2} \right] -$$

$$- 2\alpha^2 \sum_{j=1}^{l-1} \sum_{i=0}^{j-1} (j-i)\, h(i)h(j) =$$

$$= 4\alpha\beta l F(l+1) - 2\alpha\beta l - 2\alpha\beta np + 4\alpha\beta(l+1)\binom{n}{l+1} p^{l+1} q^{n-l} +$$

$$+4\alpha\beta np\cdot m_0(l+1)+2\alpha^2npq\left[\sum_{i=1}^{n-1}\binom{n-1}{i-1}\binom{n-1}{i}p^{2i-1}q^{2n-2i-1}+\right.$$

$$\left.+\sum_{i=0}^{n-1}\binom{n-1}{i}^2p^{2i}q^{2n-2i-2}\right]-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j).$$

Schließlich bekommen wir:

Folgesatz 3.1. Im Falle der verallgemeinerten inflationistischen binomialen Verteilung haben wir

$$\Delta_1=2\alpha\beta\left[2lF(l+1)-l+np\,(2m_0(l+1)-1+2(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}\right]+$$

$$+2\alpha^2npq\left[\sum_{i=1}^{n-1}\binom{n-1}{i-1}\binom{n-1}{i}p^{2i-1}q^{2n-2i-1}+\sum_{i=0}^{n-1}\binom{n-1}{i}^2p^{2i}q^{2n-2i-2}\right]-$$

$$-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)\,h(i)h(j). \tag{3.6}$$

Der in (3.6) vorkommende Wert $F(l+1)$ der binomialen Verteilunosfunktion kann man entsprechenden statistischen Tafel entnechmen, z.B. R. Zieliński (1972, S. 150).

Man kann auch eine etwas andere Form dieser Formel erreichen, wenn man die Relation (2.8), S. 550 aus der Arbeit von Ramasubban heranzieht

$$\Delta_1=4\alpha\beta lF(l+1)-2\alpha\beta l-2\alpha\beta np+4\alpha\beta(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}+$$

$$+4\alpha\beta np+2\alpha^2pq\sum_{i=0}^{n-1}(-1)^i\binom{n}{i+1}\binom{2i}{i}p^iq^i-$$

$$-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)\,h(i)h(j).$$

Dann gilt:

Folgesatz 3.2. Die Gini-Mitteldifferenz der verallgemeinerten inflationistischen Binomialverteilung hat die Form

$$\Delta_1=2\alpha\beta\left[2lF(l+1)-l+np(2m_0(l+1)-1)+2(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}\right]+$$

$$+2\alpha^2\left[pq\sum_{i=0}^{n-1}(-1)^i\binom{n}{i+1}\binom{2i}{i}p^iq^i-\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)\,h(i)h(j)\right]. \tag{3.7}$$

Nach etwas mühsamer Rechnung gelingt es auch zu zeigen, daß die Mitteldifferenz für die Binomialverteilung ohne Deformierung die folgende Form hat:

$$\Delta_1' = 2pq \sum_{i=0}^{n-1} (-1)^i \binom{n}{i+1}\binom{2i}{i} p^i q^i =$$

$$= 2npq\left[1 - \frac{n-1}{1!}\left(\frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{4pq}{1!} + \frac{(n-1)(n-2)}{2!}\cdot\frac{1}{3}\left(\frac{1}{2}\cdot\frac{3}{2}\right)\cdot\frac{(4pq)^2}{2!} - \right.$$

$$\left. - \ldots + (-1)^{n-1}\frac{(n+1)\cdot\ldots\cdot(2n-2)}{(n-1)!}\cdot\frac{1}{n}\cdot\left(\frac{1}{n-1}\cdot\frac{n}{n-1}\right)\cdot\frac{(4pq)^{n-1}}{(n-1)!}\right].$$

Führen wir die Bezeichnungen

$$\alpha = -(\mathrm{n}-1), \quad \beta = \frac{1}{2}, \quad \gamma = 2, \quad x = 4pq,$$

ein, so wird

$$\Delta_1' = 2npq\left[1 + \frac{\alpha\beta}{\gamma}\cdot\frac{x}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\cdot\frac{x^2}{2!} + \ldots + \right.$$

$$\left. + \frac{\alpha(\alpha+1)\ldots(\alpha+n-2)\cdot\beta(\beta+1)\ldots(\beta+n-2)}{\gamma(\gamma+1)\ldots(\gamma+n-2)}\cdot\frac{x^{n-1}}{(n-1)!}\right],$$

was wir noch etwas einfacher schreiben können als:

$$\Delta_1' \cong 2npqF(\alpha, \beta, \gamma, x) = 2npqF\left[(-n+1), \frac{1}{2}, 2, 4pq\right], \tag{3.8}$$

d.h. in Form der hypergeometrischen Reihe

$$F(\alpha, \beta, \gamma, x) = 1 + \sum_{n=1}^{\infty} \frac{\alpha^{[n,-1]}\beta^{[n,-1]}}{\gamma^{[n,-1]}}\cdot\frac{x^n}{n!},$$

wobei $\alpha^{[n-1]} = \alpha(\alpha+1)(\alpha+2)\ldots(\alpha+n-1)$ das faktorielle Polynom darstellt.

Schließlich (3.8) in Betracht ziehend, haben wir:

Folgesatz 3.3. Die Gini-Mitteldifferenz der verallgemeinerten inflationistischen Binomialverteilung hat die asymptotische Form

$$\Delta_1 \cong 2\alpha\beta\left[2lF(l+1) - l + np(2m_0(l+1) - 1) + 2(l+1)\binom{n}{l+1}p^{l+1}q^{n-1}\right] +$$

$$+ 2\alpha^2\left\{npqF\left[(-n+1), \frac{1}{2}, 2, 4pq\right] - \sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)\right\}. \tag{3.9}$$

Im Fall $i = l = 0$ nehmen die oben angegebenen Formeln eine einfachere Gestalt an. Die vollständigen Ausdrücke dafür finden sich in der Arbeit von T. Gerstenkorn (1997).

LITERATUR

Crow E. L. (1958), *The Mean Deviation of the Poisson Distribution*, „Biometrika" **45**(3–4), 556–559.

Frame J. S. (1945), *Mean Deviation of the Binomial Distribution*, Amer. Math. Monthly **52**, 377–379.

Gerstenkorn T. (1971), *The Recurrence Relations for the Moments of the Discrete Probability Distributions*, Dissert. Mathem. **83**, 1–46.

Gerstenkorn T. (1975), *Bemerkungen über die zentralen unvollständigen und absoluten Momente der Pólya-Verteilung*, „Zastosowania Matematyki" – Applic. Math. **14**(4), 579–597.

Gerstenkorn T. (1977), *Jednowymiarowe rozkłady dyskretne ze zniekształceniem*, [in:] *Metody statystyczne w sterowaniu jakością*, Sprawozdanie z Konferencji PAN w Jabłonnej 24–28 listopada 1975 r., red. Sz. Firkowicz, Ossolineum, Wrocław, 195–208.

Gerstenkorn T. (1997), *The Gini's Mean Difference of an Inflated Discrete Distribution*, Proc. 16th Intern. Conf. on Multivariate Statistical Analysis MSA' 97, November 27–29 1997, Łódź University, Chair of Statistical Methods, ed. Cz. Domański, D. Parys, 147–151.

Gini C. (1910), *Indicici di concentrazione e di dipendenza, atti della III Riunione della Societa Italiana per il progresso delle scienze*, Padua 1910, 453–469.

Gini C. (1911), *Variabilità e mutabilità – contributo allo studio delle distribuzioni e delle relazioni statististiche*, „Studi Economico–Giuridici delle R. Universita di Cagliari", vol. III, Parte II, 3–159.

Johnson N. L. (1957), *A Note on the Mean Deviation of the Binomial Distribution*, „Biometrika" **44**(3–4), 532–533.

Kamat A. R. (1953), *The Third Moment of Gini's Mean Difference*, „Biometrika" **40**(3–4), 451–452.

Katti S. K. (1960), *Moments of the Absolute Difference and Absolute Deviation of Discrete Distributions*, Ann. Math. Stat. **31**(1), 78–85.

Kendall M. G., Stuart A. (1963), *The Advanced Theory of Statistics*, Vol. 1, Distribution Theory, II ed., Charles Griffin & Comp., London, Sec. 2.20–2.23; Russische Ausgabe: *Tieorija raspriedielenij*, Izd. Nauka, Moskwa 1966.

Lomnicki Z. A. (1952), *The Standard Error of Gini's Mean Difference*, Ann. Math. Stat. **23**, 635–637.

Lorenz M. O. (1905), *Methods of Measuring the Concentration of Wealth*, J. Amer. Statist. Assoc. **9**, 209 ff.

Nair U. S. (1936), *Standard Error of Gini's Mean Difference*, „Biometrika" **28**, 428 ff.

Ramasubban T. A. (1956), *A χ-Approximation to Gini's Mean Difference*, J. Indian Soc. Agric. Statist. **8**, 116 ff.

Ramasubban T. A. (1958), *The Mean Difference and the Mean Deviation of some Discontinuous Distributions*, „Biometrika" **45**(3–4), 549–556.

Ramasubban T. A. (1959), *The Generalized Mean Differences of the Binomial and Poisson Distributions*, „Biometrika" **46**(1–2), 223–229.

Rinne H. (1974), *Statistik I: Vorlesungsunterlagen für das Grundstudium der Wirtschaftswissenschaften im Fach Mathematik*, Band I, Giessen, Justus-Liebig-Universität.

Risser R., Traynard C. E. (1933, 1957), *Les Principes de la Statistique Mathématique, Livre I*, Séries Statistiques, Paris 1933; 2ed. 1957.

Singh S. N. (1963), *A Note on Inflated Poisson Distribution*, J. Ind. Stat. Assoc. **1**(3), 140–144.

Singh M. P. (1965/66), *Inflated Binomial Distribution*, J. Sci. Res. Banares Hindu University **16**(1), 87–90.

Singh M. P. (1966), *A Note on Generalized Inflated Binomial Distribution*, „Sankhyã, The Indian Journal of Statistics" **28**(1), 99.

Zieliński R. (1972), *Tablice statystyczne*, PWN, Warszawa.

Tadeusz Gerstenkorn, Joanna Gerstenkorn

WSPÓŁCZYNNIK GINIEGO W PRAKTYCE STATYSTYCZNEJ. ZASTOSOWANIE DO PEWNYCH ROZKŁADÓW INFLACYJNYCH

(Streszczenie)

W pracy podane są interesujące własności średniej różnicy Giniego. Zacytowana jest odpowiednia literatura uwzględniająca publikacje książkowe i artykuły. Przedstawione zostaje zastosowanie średniej różnicy do rozkładów inflacyjnych (ze zniekształceniem) ważnych i interesujących w problematyce statystycznej.