

Information systems research

УДК 519.65.1:336.153

doi: 10.20998/2522-9052.2019.2.09

В. Ю. Дубницкий¹, Г. Г. Зубрицкая², А.И. Ходырев¹¹ Харьковский учебно-научный институт ГВУЗ Университета банковского дела, Харьков, Украина² Ягеллонский университет, Краков, Польша

ОПРЕДЕЛЕНИЕ ИНДЕКСА ДЖИНИ С УЧЁТОМ ПОГРЕШНОСТЕЙ ВЫБОРОЧНЫХ НАБЛЮДЕНИЙ

Аннотация. В статье рассматривается случайная величина, полученная в результате измерения величины конкретного физического свойства. Исследуются статистические и вычислительные аспекты задачи оценки отклонения распределения данной случайной величины от равномерного распределения. Рассматриваемая случайная величина нормируется. Одним из способов определения меры отклонения полученной таким образом функции распределения от равномерного распределения используют индекс Джини. **Целью статьи** является разработка предложений по учету погрешностей наблюдений при определении индекса Джини и построении кривой Лоренца. **Результаты.** Рассмотрена задача вычисления индекса Джини и параметров кривой Лоренца с учетом погрешностей выборочных наблюдений. Показано, что эта задача возникает в различных предметных областях, в том числе и социологии. Для учёта ошибок выборочных наблюдений использованы интервальные вычисления в системе центр – радиус. Для вычисления индекса Джини применено численное интегрирование по формуле трапеций с использованием интервальных чисел. **Вывод.** Показано, что отказ от учёта ошибок выборочных наблюдений может привести к ошибочным выводам об уровне социального расслоения в обществе и оценке наличия в нём среднего класса.

Ключевые слова: индекс Джини; кривая Лоренца; средний класс; интервальные вычисления; численное интегрирование интервально определённой функции.

Введение

Постановка задачи. Пусть задано множество физически реализуемых объектов Ω , содержащее J объектов ω_j . Каждому из объектов ω_j поставлена в соответствие величина его физически измеримого свойства S , равная s_j . Величину этого свойства для конкретного объекта предполагаем случайной и принимаем, что погрешность определения численного значения свойства s_j пренебрежимо мала в сравнении со значением измеряемого свойства. Примем, что $0 \leq S < \infty$. Требуется оценить отклонение распределения этого свойства от равномерного распределения. Такого рода задачи возникают при изучении или управлении процессами в системах различной природы. Например, в бетонировании [1], обогащении полезных ископаемых [2], металлургии [3], электросварке [4], технологии пищевых продуктов [5], экономической статистике [6; 7]. В зависимости от содержательного смысла задачи целью исследования могут быть поиски способов получения максимального приближения плотности (функции) распределения свойств s_j к равномерному распределению либо достижению противоположной цели. Профессиональная компетенция авторов позволяет затронуть только статистические и вычислительные аспекты этой задачи.

Анализ литературы. Разработка методов оценки отклонения исходного распределения случайной величины от равномерного распределения стало одной из основных задач раздела количественной социологии, изучающего бедность и неравенство [7]. История изучения бедности как социального явления подробно рассмотрена в работе [8].

Современные знания об этой проблеме приведены в работах [9–14]. Важность этого направления наглядно показана в работе [13]. В ней приведены данные об уровне бедности в Украине, определённые по различным критериям и их значения по состоянию на 2013 г. Эти результаты показаны в табл. 1.

Таблица 1 – Процент граждан Украины, которых следует считать бедными по состоянию на 2013 г.

Критерии ООН	Критерий прожиточного минимума	Относительный критерий бедности
2,9	9,9	26

Различие в величинах сравниваемых показателей (табл. 1) подчёркивает важность выбранного направления исследований. Среди показателей характеризующих бедность следует различать абсолютные показатели и относительные. Абсолютные показатели, как правило, именованные и характеризуют уровень потребления блага, измеренный в каких-либо денежных единицах (de). Их размерность, чаще всего, имеет вид $[de / (чел \cdot день)]$. Эти данные приведены, например, в работе [10]. Относительные показатели, как правило, основаны на использовании характеристик, полученных в результате статистического анализа рядов распределения, характеризующих уровень потребления данного блага. Такими показателями, как правило, могут быть среднее значение ряда распределения, его среднеквадратическое отклонение, значения верхних и нижних квартилей, децилей и квинтилей. Подробно способы определения этих характеристик изложены в работе [15]. Детальный ана-

лиз сходства и различия в этих показателях выходит за рамки данного сообщения. Одной из первых работ, в которой были рассмотрены проблемы измерения уровня бедности, стала работа [16]. Её библиографическое описание, выполненное по действующим правилам, приведено в списке литературы. Для облегчения поиска этой работы на рис. 1 показан её титульный лист оригинала.

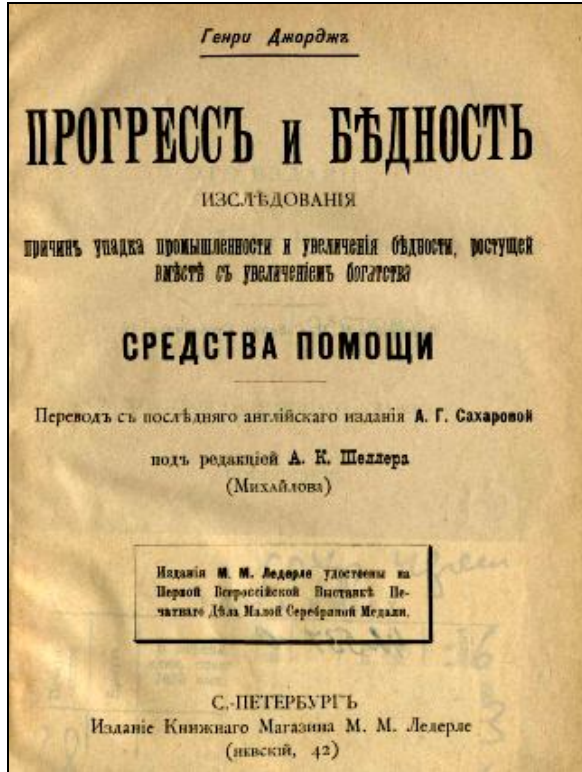


Рис. 1. Титульный лист книги Генри Джордж “Прогресс и бедность. Исследование упадка промышленности и увеличения бедности, растущей с увеличением богатства”

В силу сделанных предположений выражение

$$x = \frac{s - s_{\min}}{s_{\max} - s_{\min}} \quad (1)$$

преобразовывает исходную случайную величину s_j в случайную величину $x \in [0,1]$. В этом случае одним из способов определения меры отклонения полученной таким образом функции распределения случайной величины X от равномерного распределения используют индекс Джини (IG).

На рис. 2 показан полигон накопленных частот для одного из возможных вариантов распределения значений случайной величины X . На рис. 3 показана функция равномерного распределения (линия, соединяющая точки с координатами $(0,0)$ и $(1,1)$) и полученная функция распределения линия $L(x)$.

Если $S1$ и $S2$ – площади соответствующих элементов рис. 2, то величину индекса Джини определяют по условию:

$$IG = \frac{S1}{S + S2} = \frac{S1}{1/2} = 2A = 2 \left(\frac{1}{2} - S \right) = 1 - 2 \int_0^1 L(x) \cdot (2) dx$$

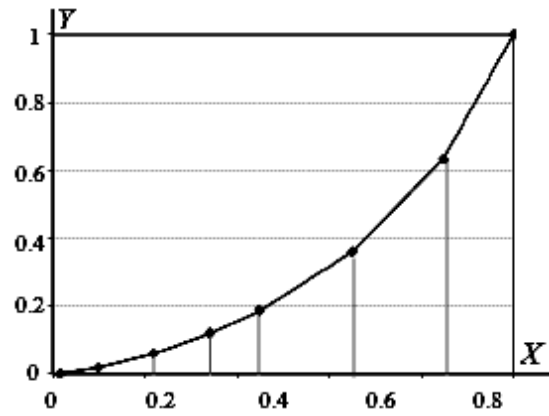


Рис. 2. Полигон накопленных частот распределения значений случайной величины X

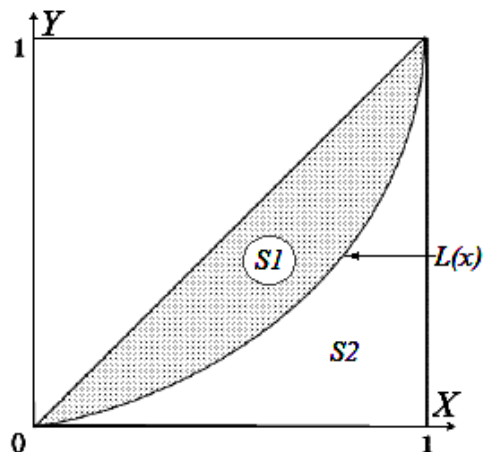


Рис. 3. Графическое отображение индекс Джини

Методические особенности вычисления этого индекса и его численные значения для различных государств, в том числе и для Украины, рассмотрены в работе [18]. Кривая $L(x)$ получила название кривой Лоренца, предложившего её для применения в задачах социологии и политической экономики [19]. В работе [20] справедливо отмечено, что для вычисления указанного в условии (2) интеграла используют данные, представленные в табличном виде. Для интегрирования предложено использовать формулу трапеций в виде, представленном в работе [20]. Следуя работе [20, с. 195] назовём таблицей упорядоченную пару чисел (x_i, y_i) для которой $x_1 < x_2 < \dots < x_n$. Эта таблица задаёт точные значения (то есть такие, которые определены без учета погрешностей их измерения) функции $y_i = f(x_i)$. Используя обобщённую формулу трапеций, получим:

$$I = \int_{x_1}^{x_n} f(x) dx \approx \frac{1}{2} \sum_{i=1}^{n-1} (x_{i+1} - x_i) (y_i + y_{i+1}) \quad (3)$$

Тогда индекс Джини будет равен величине

$$IG = 1 - \sum_{i=1}^{n-1} (x_{i+1} - x_i) \cdot (y_i + y_{i+1}), \quad (4)$$

в этом виде он приведен в работе [21].

В работе [20] для определения величины погрешности вычислений интеграла вида (3) использовано предположение о том, что допустима кусочно–полиномиальная аппроксимация полигона накопленных частот, показанного на рис. 2. Далее авторы работы [20] используют общепринятую процедуру вычисления погрешности метода трапеций при аналитически заданной функции. Её описание дано в работе [21].

Независимо от предметной области при выполнении описанных вычислений важным элементом, обеспечивающим их корректное выполнение, служит исчерпывающее изложение процедуры подготовки данных. Например, принятая в Польше методика подготовки данных для вычисления индекса Джини при определении отклонения фактической функции распределения доходов от равномерного распределения подробно описана в работе [22]. Для подготовки данных использован метод выборочного исследования домохозяйств. Каждый месяц исследование охватывает 3132 домашних хозяйств. Таким образом, годовые результаты касаются 37584 домохозяйств. Выборки домохозяйств, привлекаемых к исследованию, определяются на основе двухэтапной схемы с учетом административно-территориального деления страны. Первичные статистические регионы (ПСР) минимально включают 250 городских и 150 сельских домохозяйств. Всего используется около 30 000 первичных статистических регионов, из которых на втором этапе производится случайная выборка домашних хозяйств. В процессе такого обследования ротация (замена) выборок производится один раз в месяц; ежемесячно в каждом первичном статистическом регионе (ПСР) случайным образом определяют два домохозяйства; исследование в этих домохозяйствах проводят в выбранный месяц года в течение последующих двух лет.

Следует отметить, что в работе [22] и других, аналогичных ей работах, методика оценки влияния погрешности определения исходных данных на результат вычисления индекса Джини не описана.

Таким образом, **целью статьи** является разработка предложений по учету погрешностей наблюдений при определении индекса Джини и построении кривой Лоренца.

Изложение основного материала

Способы оценки вычислительных погрешностей при проведении выборочных исследований в зависимости от принятого плана их проведения подробно описаны в работе [23]. Для выполнения вычислений, необходимых при определении индекса Джини с учетом погрешностей, обусловленных использованием выборочных данных, авторами выбрана технология интервальных вычислений с представлением чисел в системе центр – радиус. Теоретические основы этого метода описаны в работе [24]. Специализированный программный калькулятор, реализующий этот метод, представлен в работе [25]. Следует заметить, что ранее авторы данного сообщения также применяли интервальные вычисления для решения социологических задач [26].

В соответствии с этими работами на множестве действительных чисел R , определим интервальное число A в виде замкнутого интервала:

$$A = \langle \underline{a}, \bar{a} \rangle = (a_1, a_2), \quad \underline{a} \leq \bar{a}; \quad a_1 \leq a_2, \quad (5)$$

и представим в виде:

$$A = \langle a, r_a \rangle, \quad (6)$$

где

$$a = \frac{a_1 + a_2}{2}, \quad r_a = \frac{a_2 - a_1}{2}, \quad a, r_a \in R. \quad (7)$$

При применении системы центр-радиус действия сложения и вычитания с интервальными числами выполняют по следующим правилам:

$$A + B = \langle a + b, r_a + r_b \rangle; \quad (8)$$

$$A - B = \langle a - b, r_a + r_b \rangle. \quad (9)$$

Примем, что границы интервалов, которые ограничивают рассматриваемые числа, образованы вычислительными ошибками, погрешностями измерений или неполным знанием области изменения некоторой физической величины. Поэтому в условии (6) должны быть выполнены неравенства:

$$a \geq r_a \geq 0, \quad b \geq r_b \geq 0, \quad (10)$$

иначе будем считать, что задача, в рамках наших представлений об исследуемом объекте, физического смысла не имеет. Операцию умножения в системе центр-радиус выполняют по такому правилу:

$$\langle a, r_a \rangle \langle b, r_b \rangle = \langle ab + r_a r_b, ar_b + br_a \rangle. \quad (11)$$

Рассмотрим основные арифметические операции в том случае, когда один из операндов - постоянное число. В системе центр-радиус постоянное число C представим в виде $C = \langle c, 0 \rangle$. Примем, что $A = \langle a, r_a \rangle$ и $B = \langle b, 0 \rangle$. Тогда операции сложения и вычитания представим в виде:

$$A + B = \langle a + b, r_a \rangle; \quad (12)$$

$$A - B = \langle a - b, r_a \rangle. \quad (13)$$

Для умножения интервального числа, представленного в системе центр-радиус, на постоянную величину примем, что:

$$AB = \begin{cases} \langle a, 0 \rangle \langle b, r_b \rangle, & A = const, B \neq const; \\ \langle a, r_a \rangle \langle b, 0 \rangle, & A \neq const, B = const. \end{cases} \quad (14)$$

Используя условия (5)...(14) получим условия, необходимые для вычисления выражения (4) в интервальном виде:

$$\begin{aligned} x_i &\rightarrow [X](i) = (\hat{x}(i); r_x(i)); \\ y_i &\rightarrow [Y](i) = (\hat{y}(i); r_y(i)); \end{aligned} \quad (15)$$

$$\begin{aligned} x_{i+1} - x_i &\rightarrow ((\hat{x}(i+1); r_x(i+1)) - (\hat{x}(i); r_x(i))) = \\ &= (\hat{x}(i+1) - \hat{x}(i); r_x(i+1) + r_x(i)) = (u_i; r_u(i)); \end{aligned} \quad (16)$$

$$y_{i+1} + y_i \rightarrow \left((\hat{y}(i+1); r_y(i+1)) + (\hat{y}(i); r_y(i)) \right) = \quad (17)$$

$$= (\hat{y}(i+1) + \hat{y}(i); r_y(i+1) + r_y(i)) = (v_i; r_v(i));$$

$$(x_{i+1} - x_i) \cdot (y_i + y_{i+1}) \rightarrow (u(i); r_u(i)) \cdot (v(i); r_v(i)) = (u(i)v(i) + r_u(i) \cdot r_v(i); u(i)r_v(i) + v(i)r_u(i)). \quad (18)$$

В работе [18] приведены сведения о еженедельных семейных доходах в Великобритании в 1992 г., полученные при анализе выборки объемом 7418 семей. Результаты обработки этих данных показаны в табл. 2.

Результаты вычисления индекса Джини для этих данных, полученные с использованием условий (4), (16) ... (18) приведены в табл. 3.

Из приведенных в этой таблице данных следует, что учёт погрешностей выборочных наблюдений при сравнении индекса Джини для различных стран может привести к неожиданным результатам. По данным, приведенным в [28] индекс Джини в 2016 г. составил для: США – 0,411; Греции – 0,367; Боливии – 0,484.

Следовательно, публикация индекса Джини без приведения данных о погрешности его определения может сказаться на результатах работ по страноведческому анализу.

Из табл. 2 и графика, приведенного на рис. 4 хорошо видно, что, в данном случае, на верхние 6,2 % домохозяйств приходится 18,7 % дохода.

Эта же доля дохода приходится на 36,1 % домохозяйств имеющих небольшие доходы. Для дальнейшего анализа по данным, приведенным в табл. 3 и используя процедуру сплайн – аппроксимации построим кривую Лоренца, показанную на рис. 4.

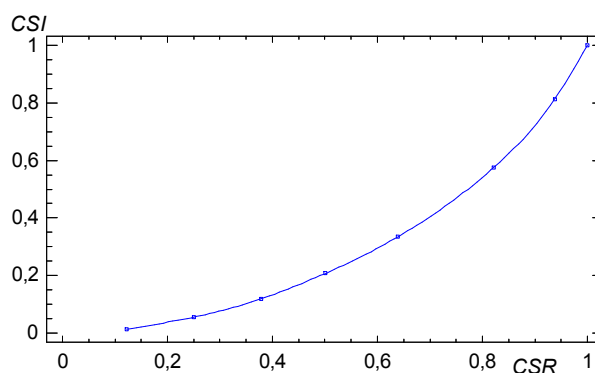


Рис. 4. Кумулятивные доли объема выборки (CSR) и дохода (CSI)

Аппроксимируем эти данные регрессионной моделью вида:

$$CSI = \exp(a + b \ln(CSR)) = \quad (19)$$

$$= \exp(-0,13024 + 1,97962 \ln(CSR)).$$

Критерии качества полученной модели приведены в табл. 4. График, соответствующий этой модели показан на рис. 5.

Таблица 2 – Кумулятивные доли объема выборки и дохода

Кумулятивная доля:	1	2	3	4	5	6	7	8
– объема выборки (x)	0,123	0,250	0,379	0,501	0,639	0,822	0,938	1,000
– объема дохода (y)	0,015	0,056	0,120	0,208	0,335	0,576	0,813	1,000

Таблица 3 – Результаты вычисления индекса Джини с использованием интервальных чисел в системе центр – радиус

Величина радиуса (% определяемой величины)					
2,5%			5 %		
НГ	Ц	ВГ	НГ	Ц	ВГ
0,36	0,41	0,45	0,34	0,41	0,48

Примечание: НГ – нижняя граница величины индекса Джини;

Ц – центральное значение индекса Джини;

ВГ – верхняя граница величины индекса Джини.

Таблица 4 – Критерии качества аппроксимации кривой Лоренца

Коэффициенты модели	Среднеквадратическая ошибка	Величина P _v	Величина F - отношения	Величина R ² (adj), %
a = - 0,13024	0,044	0,0249	–	–
b = 1,9796	0,044	< 1 · 10 ⁻⁴	–	–
F - критерий	–	< 1 · 10 ⁻⁴	2034	–
Коэффициент детерминации скорректированный R ²	–	–	–	99,67

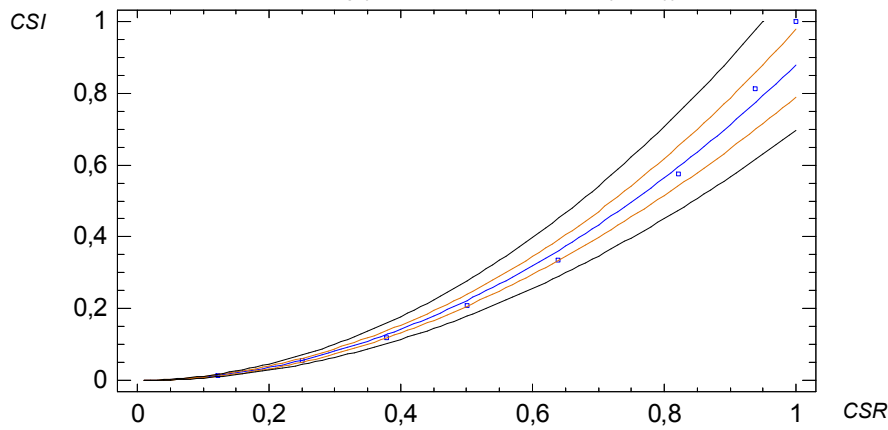


Рис. 5. Уравнение, аппроксимирующее кривую Лоренца для полученных данных

Из табл. 4 следует, что приведенное в условии (19) выражение удовлетворительно аппроксимирует полученные результаты и выражение для кривой Лоренца можно представить в нашем случае в виде:

$$L(CSR) = q(CSR)^b = 0.8778(CSR)^{1,9762}. \quad (20)$$

Для выполнения последующего анализа потребуется вычисление значений условия (19) в системе центр – радиус. Эти вычисления связаны с необходимостью вычисления в интервальном виде значений логарифмической и показательной функций.

Следуя работам [25; 26] представим логарифмическую функцию в виде:

$$\ln \langle x, r_x \rangle = \sum_{i=1}^6 \langle a_i, 0 \rangle \left[\langle -1; 0 \rangle^{i-1} + \frac{\langle 1; 0 \rangle}{\langle x, r_x \rangle^i} \right] \frac{(\langle x, r_x \rangle - \langle 1; 0 \rangle)^i}{\langle i; 0 \rangle}. \quad (21)$$

Далее при описании вычислительных алгоритмов, во избежание недоразумений, связанных с использованием десятичных дробей, вместо символа $\langle a, r_a \rangle$ будем использовать символ $\langle a; r_a \rangle$.

Коэффициенты a_i , необходимые для вычисления величины $\ln \langle x, r_x \rangle$ приведены в табл. 5.

Таблица 5 – Значение коэффициентов для приближения функции $\ln(x)$

a_1	0,500000	a_4	0,030303
a_2	0,227273	a_5	0,007576
a_3	0,090909	a_6	0,0001082

Произвольную показательную функцию представим в виде:

$$a^x = \sum_{k=0}^{\infty} \frac{(x \ln a)^k}{k!}. \quad (22)$$

Тогда, её интервальным расширением будет функция вида:

$$\langle a; r_a \rangle^{\langle x; r_x \rangle} = \sum_{k=0}^6 \frac{(\langle x; r_x \rangle \ln \langle a; r_a \rangle)^k}{k!}. \quad (23)$$

Экспоненту с отрицательным показателем, точнее её рациональное приближение, представим в виде

$$e^{-x} = \left[\sum_{k=0}^6 a_k x^k \right]^{-4}, \text{ при } 0 \leq x \leq 16. \quad (24)$$

Значения коэффициентов a_k , используемых для приближения величины e^{-x} , приведены в табл. 6.

Таблица 6 – Значение интерполяционных коэффициентов для приближения величины e^{-x}

a_0	1	a_4	0,0001715620
a_1	0,2499986842	a_5	0,0000054302
a_2	0,0312575832	a_6	0,0000006906
a_3	0,00259137121		

Интервальное расширение функции (24) примет вид:

$$e^{-\langle x; r_x \rangle} = \left[\sum_{k=0}^6 \langle a; r_a \rangle_k \langle x; r_x \rangle^k \right]^{-4}, \text{ } 0 \leq x \leq 16. \quad (25)$$

Экспоненту с положительным показателем представим в виде:

$$e^{\langle x; r_x \rangle} = 1 / \left[\sum_{k=0}^6 \langle a; r_a \rangle_k \langle x; r_x \rangle^k \right]^{-4}. \quad (26)$$

Это позволяет осуществлять действия с числами в диапазоне $[1, 12 \cdot 10^{-7}; 8, 88 \cdot 10^6]$.

В социологической литературе утвердилось мнение о том, что наличие широкой прослойки, именуемой средним классом, обеспечивает социальную общественную стабильность. В работах [29] и [30] приведены условия существования среднего класса, сформулированные с использованием индекса Джини и кривой Лоренца. В работе [29] в качестве критериальной величины принято значение индекса Джини, в работе [30] в качестве критериальной величины принята разность третьей и первой

квартили, соответствующая распределению, приведенному в табл. 2.

Способ вычисления квартилей изложен в работе [15].

Проверим выполнение этих условий для наших данных с учётом ошибок их определения. При вы-

числении радиусов интервалов коэффициентов уравнения (19) использовались данные о среднеквадратической погрешности определения коэффициентов регрессии. Эти значения умножали на коэффициент $k = 1,64$.

Результаты вычислений показаны в табл. 7.

Таблица 7 – Условия существования среднего класса

Литературный источник	Условия критерия	Значения критерия		
		НГ	Ц	ВГ
Работа [29]	$IG \leq \frac{1}{2}$	0,34	0,41	0,48
Работа [30]	$L\left(\frac{3}{4}\right) - L\left(\frac{1}{4}\right) \geq \frac{1}{2}$	0,38	0,44	0,51

Вывод

Полученные результаты показывают, что пре-

процедурами получения исходных данных и их последующей обработкой может привести к серьёзным ошибкам в оценке содержательного смысла результатов исследования.

СПИСОК ЛИТЕРАТУРЫ

- Ramachandran V. S., Feldman R. F., Beaudoin J. J. *Handbook of Analytical Techniques in Concrete Science and Technology*. London : Heyden, 1981. 278 p.
- Білецький В. С., Голіков О. С. *Методичні вказівки до виконання лабораторних робіт по курсу "Переробка, збагачення і комплексне використання корисних копалин"*. Покровськ : ДонНТУ, 2014. 49 с.
- Кравченко В. П. Анализ способов грануляции шлаковых расплавов и факторов, влияющих на качество граншлака. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2015. Вип. 30, Т. 1, С. 51–58.
- Конспект лекцій з дисципліни "Виробництво матеріалів для зварювання, наплавлення та напилення", ч. I «Виробництво покритих електродів» для студентів спеціальності 131 „Прикладна механіка” усіх форм навчання / Укл.: М.І. Андрущенко, О.Є. Капустян. – Запоріжжя: ЗНТУ, 2017 – 70 с.
- Аналіз дисперсного стану рослинної сировини зернового походження / В. М. Чорний, Ю. Ю. Прищепа, Н. В. Лапіна, Г. В. Ляшко, А. В. Рибачок, Т. Г. Мисюра, Н. В. Попова, Ю. В. Запорожець // Одеська національна академія харчових технологій, Наукові праці. – 2016. – Вип. 47, Т. 2. – С. 70–75.
- Рождественська Л. Г. Статистика ринку товарів і послуг: Навч. пос. / Л. Г. Рождественська. – К.: КНЕУ, 2005. – 419 с.
- Добреньков В. И. Социология: Учебник / В. И. Добреньков, А. И. Кравченко. – Москва: ИНФРА, 2001. – 624 с.
- Сычѐва В. С. Измерение уровня бедности: история вопроса от XVIII века до наших дней / В. С. Сычѐва // Социология: методология, методы, математическое моделирование (4М). – 2002. – № 14. – С. 176–186.
- Бродська С. С. Світовий досвід вимірювання межі бідності / С. С. Бродська // Наукові записки. – Національний університет "Києво-Могилянська академія". – Т. 21. Соціологічні науки. – С. 44–49.
- Мартин Рэвеллон. Сравнительные оценки бедности [Электронный ресурс] / Мартин Рэвеллон. – Режим доступа к материалу: <https://docplayer.ru/49981763-Sravnitelnye-ocenki-bednosti-martin-revellon.html>
- Білоус І. Методика оцінки бідності в Україні та проблеми фінансового забезпечення її подолання / І. Білоус // Світ фінансів. – 2012. – № 4. – С. 120–126.
- Мельник С. Критерії оцінки бідності працюючого населення в Україні / С. Мельник, Н. Гаєвська // Україна: аспекти праці. – 2012. – № 2. – С. 14–19.
- Хомяк М. Бідність в Україні у показниках соціальної статистики / М. Хомяк // Соціологічні студії. – 2013. – № 2(3). – С. 70–76.
- Чуприна О. О. Соціальна нерівність і бідність в Україні / О. О. Чуприна // Вісник Національної юридичної академії України імені Ярослава Мудрого. – 2011. – № 2(5). – С. 57–76.
- Ежов А. И. Выравнивание и вычисление рядов распределения / А. И. Ежов. – Москва : Госстатиздат, 1961. – 335 с.
- Генри Джордж. Прогресс и бедность. Исследование упадка промышленности и увеличения бедности, растущей с увеличением богатства / Джордж Генри. – С.Пб.: Издание книжного магазина М. М. Ледерле, 1896. – 678 с.
- Примостко О.О. Методичні підходи до оцінки індексу якості життя населення України / О. О. Примостко // Регіональна економіка. – 2015. – № 2. – С. 80–88.
- Lorenz M. O. Methods for measuring the concentration of wealth / M. O. Lorenz. – Pub. Amer. Stat. Assoc, 1905. – Vol. 9. – 209 p.
- Костробій П. Математичне моделювання індексу суспільної нерівності / П. Костробій, І. Ковалець, Л. Гнатів // Фізико-математичне моделювання та інформаційні технології. – 2013. – Вип. 17. – С. 81–89.
- Каханер Д. Численные методы и программное обеспечение / Д. Каханер, К. Моулер, С. Нэш. – Москва: МИР, 1998. – 575 с.
- Згуровський М. З. Сталий розвиток у глобальному і регіональному вимірах / М. З. Згуровський. – Київ: Політехніка, НТУУ «КПІ», 2006. – 84 с.

22. Główny Urząd Statystyczny. Zeszyt metodologiczny. Badanie budżetów gospodarstw domowych. – Warszawa, 2018. – 86 s.
23. Кокрен У. Методы выборочного исследования / У. Кокрен – Москва: СТАТИСТИКА, 1976. – 440 с.
24. Жуковська О.А. Основи інтервального аналізу / О.А. Жуковська. – К.: Освіта України, 2009. – 136 с.
25. Дубницький В. Ю. Вычисление значений элементарных и специальных функций с интервально заданным аргументом, определённым в системе центр-радиус / В. Ю. Дубницький, А. М. Кобылин, О. А. Кобылин // Прикладная радиоэлектроника. – 2017. – Т. 16, № 3-4. – С. 147–154.
26. Дубницький В. Ю. Інтервальне оцінювання кількості учасників масових протестних акцій / В. Ю. Дубницький, Г. Г. Зубрицька, А. М. Кобылин // Сучасні інформаційні системи. – 2018. – Т. 2, № 4. – С. 11–20. – DOI: <https://doi.org/10.20998/2522-9052.2018.4.02>
27. Barrov M. M. Statistics for economics accounting and business studies, London and New York : Longman, 1996, 321 p.
28. Коэффициент Джини по странам [Электронный ресурс] / Режим доступа: <https://theworldonly.org/koeffitsient-dzhini-po-stranam>
29. Геворкян П. С. Распределение богатства в обществе и средний класс / П. С. Геворкян, В. И. Малыхин // Труд и социальные отношения. – 2010. – № 12. – С. 90–97.
30. Павлов О. И. Кривая Лоренца и математическое определение среднего класса / О.И. Павлов, О.Ю. Павлова // Управление экономическими системами : электронный журнал. – 2017. – № 9 (103). – Режим доступа: <https://cyberleninka.ru/article/n/krivaya-lorentsa-i-matematicheskoe-opredelenie-srednego-klassa>

REFERENCES

1. Ramachandran, V.S., Feldman, R.F. and Beaudoin, J.J. (1981), *Handbook of Analytical Techniques in Concrete Science and Technology*, Heyden, London, 278 p.
2. Beletsky, V.S. and Golikov, O.S. (2014), *Methodical instructions for laboratory work on the course "Processing, enrichment and complex use of minerals"*, DonNTU, Pokrovsk, 49 p.
3. Kravchenko, V.P. (2015), "Analysis of methods for granulation of slag melts and factors affecting the quality of the granular slag", *Bulletin of the Azov State Technical University*, Series: Technical Sciences, No. 30, Vol. 1, pp. 51–58.
4. Andrushchenko, M.I. and Kapustian, O.Ye. (2017), *Production of materials for welding, surfacing and spraying, Production of coated electrodes*, ZNTU, Zaporizhzhya, 70 p.
5. Chorniy, V.M., Pryshepa, Yu.Yu., Lapina, N.V., Lyashko, G.V., Rybachok, A.V., Misyura, T.G., Popova, N.V. and Zaporozhets Yu.V. (2016), "Analysis of Disperse Condition of Vegetable Raw Materials of Grain Origin", *Scientific Works*, Odessa National Academy of Food Technologies, Issue 47, Vol. 2, P. 70–75.
6. Rozhdyestvenska, L.H. (2005), *Market Statistics of Goods and Services*, KNEU, Kyiv, 419 p.
7. Dobrenkov, V.I. and Kravchenko, A.I. (2001), *Sociology*, INFRA, Moscow, 624 p.
8. Sychev, V.S. (2002), "Measuring Poverty Level: Background from the 18th Century to the Present Day", *Sociology: Methodology, Methods, Mathematical Modeling (4M)*, No. 14, pp. 176–186.
9. Brodska, S.S. (2003), World Experience in Measuring the Poverty Line, *Scientific Notes*, Vol. 21. Sociological Sciences, NUKMA, Kyiv, pp. 44–49.
10. Martin, Ravellon (2019), *Comparative Poverty Estimates*, available to: <https://docplayer.ru/49981763-Sravnitelnye-ocenki-bednosti-martin-revellon.html>
11. Bilous I. (2012), "The Methodology of Assessment of Poverty in Ukraine and Problems of Financial Support to Overcome it", *World Finance*, No. 4, pp. 120–126.
12. Melnyk, S. and Gaevskaya, N. (2012), "Criteria for assessing the poverty of the working population in Ukraine", *Ukraine: aspects of labor*, No. 2, pp. 14–19.
13. Khomyak, M. (2013), "Poverty in Ukraine in the indicators of social statistics", *Sociological Studies*, No. 2 (3), pp. 70–76.
14. Chuprina, O.O. (2011), "Social Inequality and Poverty in Ukraine", *Bulletin of the National Law Academy of Ukraine*, No. 2 (5), pp. 57–76.
15. Yezhov, A.I. (1961), *Alignment and Calculation of Distribution Series*, Gosstatizdat, Moscow, 335 p.
16. Henry, George (1896), *Progress and Poverty. A study of the decline of industry and the increase in poverty growing with wealth*, Publishing of a bookstore M.M. Lederle, S.-Petersburg, 678 p.
17. Primostko, O.O. (2015), "Methodological Approaches to the Assessment of the Quality of Life of the Population of Ukraine", *Regional Economics*, No. 2, pp. 80–88.
18. Lorenz, M.O. (1905), *Methods for measuring the concentration of wealth*, Pub.Amer.Stat.Assoc, Vol. 9, 209 p.
19. Kostrobii, P., Kovalec, I. and Gnatov, L. (2013), "Mathematical modeling of the index of social inequality", *Physical-mathematical modeling and information technologies*, No. 17, pp. 81–89.
20. Kahaner, D., Moler, K. and Nash, S. (1998), *Computational Methods and Software*, MIR, Moscow, 575 p.
21. Zgurovsky, M.Z. (2006), *Sustainable development in the global and regional dimensions*, Politehnika, NTUU "KPI", Kyiv, 84 p.
22. Central Statistical Office (2018), Methodological book, *Study of household budgets*, Warsaw, 86 p.
23. Kokren, U. (1976), *Methods of selective research*, Statistics, Moscow, 440 p.
24. Zhukovska O.A. (2009), *Fundamentals of Interval Analysis*, Education of Ukraine, Kyiv, 136 p.
25. Dubnitsky, V.Yu., Kobylin, A.M. and Kobylin O.A. (2017), "Calculating the values of elementary and special functions with an interval specified argument defined in the center-radius system", *Applied Radioelectronics*, Vol. 16, No. 3-4, pp. 147-154.
26. Dubnitsky, V.Yu., Zubritska, G.G. and Kobylin, A.M. (2018), "Interval estimation of the number of participants in mass protest actions", *Advanced information systems*, Vol. 2, No. 4, pp. 11–20, DOI: <https://doi.org/10.20998/2522-9052.2018.4.02>
27. Barrov, M.M., (1996), *Statistics for economics accounting and business studies*, Longman, London and New York, 321 p.
28. TheWorldOnly (2019), *Gini coefficient by country*, available to: <https://theworldonly.org/koeffitsient-dzhini-po-stranam>.

29. Gevorkyan P.S. and Malykhin V.I. (2010), "Distribution of wealth in society and the middle class", *Labor and Social Relations*, No. 12, pp. 90–97.
30. Pavlov, O.I. and Pavlova, O. Yu. (2017), "The Lorenz curve and the mathematical definition of the middle class", *Management of Economic Systems*, No. 9 (103), available to: <https://cyberleninka.ru/article/n/krivaya-lorentsa-i-matematicheskoe-opredelenie-srednego-klassa>.

Received (Надійшла) 21.02.2019

Accepted for publication (Прийнята до друку) 10.04.2019

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Дубницький Валерій Юрійович – кандидат технічних наук, старший науковий співробітник, старший науковий співробітник Харківського навчально-наукового інституту ДВНЗ "Університет банківської справи", Харків, Україна;

Valeriy Dubnitskiy – Candidate of Technical Sciences, Senior Research, Senior Research Associate of Kharkiv Educational Scientific Institute SHEI "University of Banking", Kharkiv, Ukraine;

e-mail: dubnitskiy@gmail.com; ORCID ID: <http://orcid.org/0000-0003-1924-4104>

Зубрицька Галина Григорівна – бакалавр, студентка магістратури Ягеллонського університету, Краків, Польща;

Halyna Zubrytska – Bachelor, Postgraduate Student of Jagiellonian University, Krakow, Poland;

e-mail: halyna.zubrytska@student.uj.edu.pl; ORCID ID: <http://orcid.org/0000-0002-9668-5491>

Ходирєв Олександр Іванович – старший викладач Харківського навчально-наукового інституту ДВНЗ «Університет банківської справи», Харків, Україна;

Alexander Khodyrev – Senior Instructor of Kharkiv Educational Scientific Institute SHEI "University of Banking", Kharkiv, Ukraine;

e-mail: khodyrev@gmail.com; ORCID ID: <https://orcid.org/0000-0001-9871-9440>

Визначення індексу Джині з урахуванням похибок вибірових спостережень

В. Ю. Дубницький, Г. Г. Зубрицька, О. І. Ходирєв

Анотація. У статті розглядається випадкова величина, отримана в результаті вимірювання величини конкретної фізичної властивості. Досліджуються статистичні та обчислювальні аспекти завдання оцінки відхилення розподілу даної випадкової величини від рівномірного розподілу. Вже згадана випадкова величина нормується. Одним із способів визначення міри відхилення отриманої таким чином функції розподілу від рівномірного розподілу використовують індекс Джині. **Метою статті** є розробка пропозицій щодо врахування похибок спостережень при визначенні індексу Джині і побудові кривої Лоренца. **Результати.** Розглянуто задачу обчислення індексу Джині та параметрів кривої Лоренца з урахуванням похибок вибірових спостережень. Показано, що ця задача виникає в різних предметних областях, в тому числі і соціології. Для обліку помилок вибірових спостережень використані інтервальні обчислення в системі центрального радіуса. Для обчислення індексу Джині застосовано чисельне інтегрування за формулою трапецій з використанням інтервальних чисел. **Висновок.** Показано, що неврахування помилок вибірових спостережень може привести до помилкових висновків про рівень соціальної нерівності суспільства та оцінці його середнього класу.

Ключові слова: індекс Джині; крива Лоренца; середній клас; інтервальні обчислення; чисельне інтегрування інтервально визначеної функції.

Determination of Gini index with considering the error of selected observations

V. Dubnitskiy, H. Zubrytska, A. Khodyrev

Abstract. In this work the following problem is considered. For a set of physically realizable objects, to each of them the value of its physically measurable property is assigned. The value of this property for a particular object is assumed to be random. The error in determining the numerical value of this property is negligible compared to the value of the property being measured. We assume that the value of the measured property is positive and finite. It is necessary to estimate the deviation of the distribution of this property from the uniform distribution. Such problems arise in the study and management of processes in the various systems. For example, in concrete science, enrichment of the fields. Depending on the substantive meaning of the task, the goal of the research may be to find ways to get the maximum approximation of the density (function) of the property distribution to a uniform distribution or to achieve the opposite goal. The Gini index is chosen as a measure of the deviation of the distribution function of the measured property values from the uniform one, and the distribution itself is approximated by the Lorenz curve. The value of the Gini index is included in the system of indicators which characterize the level of prosperity of the various states or units of territorial - administrative division. It is shown that statistical sampling methods are the main way of obtaining the necessary data for the construction of the Gini index. Was described the methodology of conducting sampling studies used in Poland. To perform the necessary calculations in determining the Gini index and building a Lorenz curve, taking into account errors caused by using the sample data, the authors selected interval computing technology with the representation of numbers in the center - radius system. To calculate the Gini index, one of the methods of numerical integration is used - the trapezoid method. The implementation of this method is proposed in the interval form. The parameters of the expression approximating the Lorenz curve are determined. The calculation of the Gini index and the construction of the Lorenz curve is performed for information about weekly family income in the UK in 1992. It is shown that the refusal to take into account sampling errors can lead to erroneous conclusions when performing a comparative analysis of the uneven distribution of income between countries or their territories. A run test was made for checking the existence of a middle class in the surveyed population using two criteria. The first is based on comparing the Gini index value with the criterial index, the second one is based on comparing the difference between the third and first quartile of the sample described by the Lorenz curve with the criterial value. The results of numerical analysis for each of the criteria are given, which were performed using interval numbers determined in the center - radius system.

Keywords: Gini index; Lorenz curve; middle class; interval calculations; numerical integration of an interval defined function.