

В ПОРЯДКЕ ОБСУЖДЕНИЯ

ОБ ИСПОЛЬЗОВАНИИ КОЭФФИЦИЕНТА ДЖИНИ В ЭКОНОМИКО-СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Г.Л. Громько,
И.Н. Матюхина

Обстоятельство, побудившее авторов статьи высказать свое мнение, – участившиеся случаи некорректного использования коэффициента Джини в экономико-статистических исследованиях, когда данный коэффициент рассчитывается не применительно к распределению суммарных (объемных) показателей по группам, а для средних и относительных (рассчитанных на душу населения) показателей. А между тем исключительную роль в статистическом исследовании играет правильный выбор и корректное применение статистического инструментария для анализа в принципе достоверных и надежных данных. Даже самый простой и широко известный метод обработки статистических данных требует грамотного его использования в зависимости от наличия той или иной информации.

В статье на конкретных примерах продемонстрированы возможности и условия применения популярного среди исследователей коэффициента Джини. Авторы исходят из того факта, что индекс концентрации, чаще именуемый в литературе коэффициентом Джини и первоначально предложенный итальянским статистиком К. Джини для оценки степени неравномерности распределения богатства (доходов) населения, может быть применен для характеристики неравномерности, но уже применительно к другим показателям. Статья содержит изложение основных положений статистической теории измерения такого свойства единиц совокупности, как концентрация, и широко известных формул расчета коэффициента Джини. Авторы акцентируют внимание на важном условии применения этого показателя: исходные данные должны допускать возможность распределения по выделяемым группам совокупности, то есть должны быть представлены обязательно суммарными итогами по группам, а не показателями на душу населения или средними величинами.

Ключевые слова: кривая Лоренца, коэффициент Джини, индекс концентрации, распределение населения по среднедушевым денежным доходам, распределение денежных доходов населения по 20%-ным группам населения, группировки, средние величины, показатели дифференциации.

JEL: C10, C18.

Любое социально-экономическое исследование, начиная с курсовой студенческой работы и кончая крупными диссертационными исследованиями, монографиями, отчетными докладами НИИ, предполагает обязательное использование фактологического материала в виде статистических данных, а также различных приемов и методов, имеющихся в арсенале научного анализа. Качество научных исследований во многом определяется корректным и грамотным обращением со статистическим материалом, методиками расчета статистических показателей и экономической интерпретацией результатов.

В настоящее время, благодаря возможностям современной вычислительной техники и при-

кладным пакетам статистических программ, во многих научных работах широко используются сложные методы комплексного анализа статистических данных, например таких, как факторный анализ, метод главных компонент, кластерный анализ и др. И это, несомненно, следует оценивать как положительный момент, как свидетельство развития и совершенствования статистических методов, расширяющих возможности более глубокого исследования сложных явлений.

Однако увлечение сложными методами анализа, на наш взгляд, не должно сопровождаться некоторым небрежным отношением к более простым методам и показателям, успешно применяющимся в статистике на протяжении многих лет

Громько Галина Леонтьевна (gromyko@econ.msu.ru) – д-р экон. наук, профессор кафедры статистики экономического факультета МГУ им. М.В. Ломоносова.

Матюхина Ирина Николаевна (iri1256@yandex.ru) – канд. экон. наук, доцент кафедры статистики экономического факультета МГУ им. М.В. Ломоносова.

для решения отдельных конкретных задач. К сожалению, приходится констатировать, что в наши дни порой встречаются работы, в том числе и диссертационные, в которых неадекватно или не совсем корректно используются отдельные методы и показатели.

В настоящей статье нам хотелось бы привлечь внимание к некоторым спорным вопросам, связанным с использованием кривой Лоренца и коэффициента Джини (индекса концентрации) в отдельных научных работах. Но прежде чем непосредственно говорить о предмете спора, хотелось бы коротко остановиться на некоторых общих положениях, касающихся вышеупомянутых показателей.

Общетеоретические положения о кривой Лоренца и коэффициенте Джини

Как известно, кривая Лоренца (см. рисунок) - особый вид графика, предложенный американским статистиком Максом Отто Лоренцем (1876-1959) для изображения неравномерного распределения общего объема определенного показателя по группам единиц статистической совокупности. Строится кривая Лоренца в системе прямоугольных координат, в квадрате 100×100 , где на оси абсцисс указаны значения кумулятивных (накопленных нарастающим итогом) долей численности единиц совокупности (p_i), а на оси ординат - значения кумулятивных долей распределяемого суммарного показателя (q_i). Причем если группы единиц совокупности ранжированы в порядке возрастания значений признака (от меньшего к большему), то кривая Лоренца будет расположена ниже диагонали, в форме вогнутости (как на рисунке); а в случае убывания значений признака (от большего к меньшему) - выше диагонали, в форме выпуклости. Диагональ квадрата отражает равномерное распределение (p_i) и (q_i), или полное отсутствие концентрации, то есть когда каждому накопленному проценту числа единиц совокупности соответствует такой же процент распределяемого показателя ($p_i = q_i$).

Чем больше фактическое распределение (p_i) и (q_i) отклоняется от равномерного, тем больше кривая Лоренца удалена от диагонали и тем больше площадь фигуры S_1 , ограниченной на графике линией равномерного распределения и кривой Лоренца.

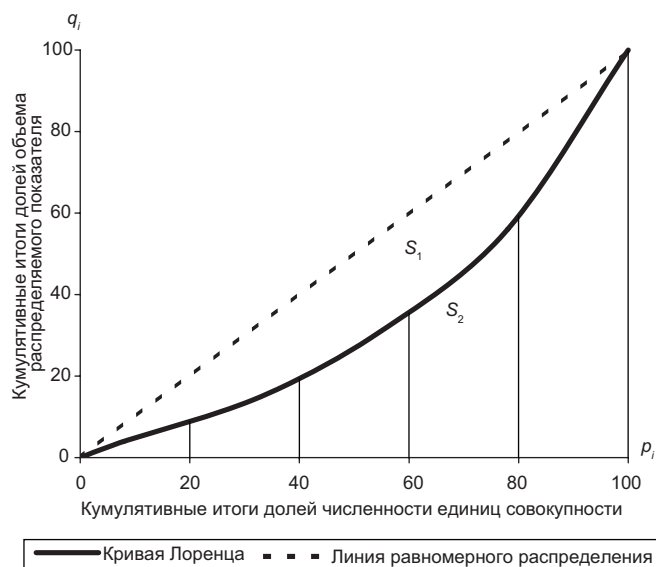


Рисунок. Кривая Лоренца

Современник Лоренца, итальянский статистик Коррадо Джини (1884-1965), предложил использовать график Лоренца для оценки социального неравенства, в частности для оценки степени неравномерности распределения доходов населения, концентрации их в отдельных группах, рассчитав коэффициент концентрации, получивший впоследствии его имя.

В геометрической интерпретации коэффициент Джини представляет собой отношение площади, ограниченной линией равномерного распределения и кривой Лоренца (S_1), к общей площади треугольника ниже диагонали ($S_1 + S_2$), то есть

$$G = \frac{S_1}{S_1 + S_2}.$$

Коэффициент Джини может принимать значения от 0 до 1. Чем ближе его значение к 1, тем выше степень неравномерности распределения, то есть больше концентрация распределяемого показателя в отдельных группах совокупности.

Приведенное выражение, отражая геометрическую сущность показателя, не является рабочей формулой для его расчета. Одну из формул коэффициента Джини можно легко получить исходя из простых рассуждений и математических расчетов. Так, принимая площадь квадрата за единицу, можно записать, что $(S_1 + S_2) = 0,5$.

И тогда $S_1 = 0,5 - S_2$, а коэффициент Джини равен:

$$G = \frac{0,5 - S_2}{0,5} = 1 - 2S_2.$$

Площадь S_2 можно представить как сумму площадей одного треугольника и нескольких трапеций, расположенных ниже кривой Лоренца, с основаниями w_i - доли единиц совокупности по группам и сторонами q_i и q_{i-1} (кумулятивные доли распределяемого суммарного показателя по группам), то есть

$$S_2 = \frac{1}{2} \sum_i^n w_i (q_{i-1} + q_i),$$

где n - число групп; $q_0 = 0$; $q_n = 1$.

Подставляя полученное значение S_2 в предыдущее выражение, получаем следующую формулу коэффициента Джини:

$$G = 1 - \sum_i^n w_i (q_{i-1} + q_i). \quad (1)$$

Наряду с данной формулой, существуют и другие ее модификации. Нередко формула коэффициента Джини приводится в виде:

$$G = 1 - 2 \sum_{i=1}^k d_{x_i} d_{y_i}^n + \sum_{i=1}^k d_{x_i} d_{y_i}, \quad (2)$$

где d_{x_i} - доля i -й группы в общем объеме единиц совокупности; d_{y_i} - доля i -й группы в общем объеме признака; $d_{y_i}^n$ - накопленная доля i -й группы в общем объеме признака¹.

Модификацией выражения (1) является и следующая формула:

$$G = \sum_{i=1}^{n-1} p_i q_{i-1} - \sum_{i=1}^{n-1} p_{i-1} q_i, \quad (3)$$

где p_i - кумулятивная (накопленная) доля i -й группы в общей численности единиц совокупности; q_i - кумулятивная доля i -й группы в общем объеме распределяемого суммарного показателя².

Формула (3) приведена для случая, когда кривая Лоренца расположена ниже диагонали квадрата (в форме вогнутости), то есть когда распределяемый показатель концентрируется в последних группах ранжированных единиц совокупности. В случаях же концентрации исследуемого по-

казателя в первых группах, то есть когда кривая Лоренца расположена выше диагонали (в форме выпуклости), в указанной формуле коэффициент Джини (3) уменьшаемое и вычитаемое меняются местами.

На наш взгляд, формула (3) наиболее проста и удобна для использования. В ней содержатся те же показатели кумулятивных долей числа единиц (p_i) и распределяемого суммарного показателя (q_i), по которым строится кривая Лоренца, а алгоритм действий с ними весьма прост и легко запоминаем, что немаловажно, особенно для студентов. Поэтому мы отдаем предпочтение этой формуле (3), и она будет использована при расчете коэффициента Джини в дальнейшем тексте в приводимых примерах.

Коэффициент Джини первоначально был предложен его автором специально для характеристики степени неравномерности распределения доходов населения. Но со временем границы его использования существенно расширились, что вполне оправданно. Поскольку коэффициент Джини связан с кривой Лоренца, то, естественно, формула, по которой он рассчитывается, может быть успешно использована не только применительно к доходам, но и к другим итоговым показателям, распределение которых по группам можно отразить графически с помощью кривой Лоренца.

Коэффициент Джини как *индекс концентрации* применим ко многим статистическим показателям. Например, для оценки концентрации распределения численности работников, стоимости основных фондов, объемов производства товаров по разным группам предприятий и пр. Но ко «многим» не означает к «любым». Он может рассчитываться для показателей разного содержания, но все они по своей природе (форме) должны являться *суммарными (итоговыми)* показателями, допускающими *распределение* по группам. Мы неслучайно акцентируем внимание на этом, казалось бы, очевидном факте. Сам термин «распределение» применим лишь к объемным (суммарным) показателям, которые действительно могут быть разделены между отдельными группами единиц совокупности, и при этом количественно определена часть этого показателя у отдельных групп (в процентах). По-

¹ Символика приведена по учебнику «Теория статистики» под ред. проф. Р.А. Шмойловой [5, с. 500].

² Символика приведена по учебнику «Теория статистики» под ред. проф. Г.Л. Громько [4, с. 119].

этому и кривая Лоренца, и коэффициент Джини, используемые в анализе при характеристике неравномерности распределения того или иного показателя по группам, строго говоря, могут (и должны) рассчитываться только для суммарного показателя.

Некоторые примеры расчета коэффициента Джини

Рассмотрим несколько примеров расчета коэффициента Джини по различным данным, публикуемым в официальных статистических справочниках РФ.

Наиболее простым является случай, когда совокупность разбита на группы по какому-либо признаку и по каждой группе имеются данные о доле единиц в ней и о доле определенного суммарного (итогового) показателя, концентрация которого исследуется.

Примером такого рода данных является таблица 1 (графы 1 и 2). В этой таблице население России разбито на пять равновеликих по численности групп (тем самым определен удельный вес каждой группы - 20%); и по каждой группе населения показана ее доля в общем объеме денежных доходов (d_{y_i}). Данные таблицы 1 демонстрируют неравномерность распределения денежных доходов при равной численности населения по группам: доля первой группы в общем объеме доходов составляет 5,2%, а пятой (с наибольшими доходами) - 47,6%. Для оценки степени концентрации доходов по квинтильным группам с помощью коэффициента Джини остается только рассчитать кумулятивные (накопленные) итоги упомянутых долей (p_i и q_i), что и приведено в дополнительных графах (3 и 4) таблицы 1.

Таблица 1

Распределение общего объема денежных доходов населения России по 20%-ным группам в 2013 г. (в процентах)

20% -ные группы населения по возрастанию среднедушевого дохода	Доля в общем объеме денежных доходов (d_{y_i})	Кумулятивные доли	
		в численности населения (p_i)	в общем объеме доходов (q_i)
1	2	3	4
Первая (с наименьшими доходами)	5,2	20	5,2
Вторая	9,8	40	15,0
Третья	14,9	60	29,9

Окончание таблицы 1

20% -ные группы населения по возрастанию среднедушевого дохода	Доля в общем объеме денежных доходов (d_{y_i})	Кумулятивные доли	
		в численности населения (p_i)	в общем объеме доходов (q_i)
1	2	3	4
Четвертая	22,5	80	52,4
Пятая (с наибольшими доходами)	47,6	100	100,0
Итого	100,0		

Источник: данные Росстата³; расчеты авторов.

Поскольку коэффициент Джини (G) может принимать значения от 0 до 1, то при его расчете надо учитывать, в каких единицах представлены доли в исходных данных. Если один из показателей выражен в процентах, то результат вычислений делится на 100; если оба показателя выражены в процентах, то на 10000. Так как в таблице 1 оба распределения даны в процентах, то искомый коэффициент Джини по формуле (3) будет равен:

$$G = \frac{(20 \times 15 + 40 \times 29,9 + 60 \times 52,4 + 80 \times 100) - (40 \times 5,2 + 60 \times 15 + 80 \times 29,9 + 100 \times 52,4)}{10000} = (12640 - 8740) : 10000 = 0,390.$$

Для справки отметим, что по данным Росстата, коэффициент Джини по доходам населения РФ в 2013 г. равен 0,419. Это расхождение вполне объяснимо: последнее значение показателя вычислено по более детальным фактическим данным, а не по 20%-ным группам населения.

Коэффициент Джини, характеризующий неравномерность распределения *общих доходов* населения, можно рассчитать и по другим данным, например по данным о *распределении населения* по величине среднедушевых денежных доходов (см. таблицу 2, графы 1 и 2). Здесь представлено лишь распределение *численности населения* России по величине среднедушевых денежных доходов в 2013 г. Распределение же *общего объема доходов* по группам населения не приведено, но легко может быть определено по исходным данным, как показано в дополнительных графах 3 и 5. В конечном счете доля каждой группы населения в общем объеме доходов определена в графе 5 как отношение $x_i w_i / \sum x_i w_i$.

В последующих графах по группам рассчитаны кумулятивные итоги доли численности на-

³ Социальное положение и уровень жизни населения России - 2014 г. Стат. сб. / Росстат. С. 105. URL: http://www.gks.ru/bgd/regl/b14_44/Main.htm.

Распределение населения РФ по величине среднедушевых денежных доходов в 2013 г.

Среднедушевые денежные доходы в месяц, тыс. рублей	Численность населения, в % к итогу (w_i)	Середина интервала душевых доходов (x_i)	$x_i w_i$	Доля объемов доходов по группам, в % $x_i w_i / \sum x_i w_i$	Кумулятивные итоги доли (в %)		Расчеты для коэффициента Джини	
					в общей численности населения, (p_i)	в общем объеме доходов, (q_i)	$p_i q_{i+1}$	$p_{i+1} q_i$
1	2	3	4	5	6	7	8	9
До 5,0	4,2	4	16,8	0,7	4,2	0,7	8,8	-
5,1-7,0	5,6	6	33,6	1,4	9,8	2,1	55,9	6,86
7,1-10,0	10,4	8,5	88,4	3,6	20,2	5,7	254,5	42,4
10,1-14,0	14,2	12,0	170,4	6,9	34,4	12,6	780,9	196,1
14,1-19,0	15,2	16,5	250,8	10,1	49,6	22,7	1934,4	625,0
19,1-27,0	17,5	23,0	402,5	16,3	97,1	39,0	4495,7	1523,2
27,1-45,0	19,3	36,0	694,8	28,0	86,4	67,0	8640,0	3369,6
Свыше 45,0	13,6	60,0	816,0	33,0	100,0	100,0	-	6700,0
	100,0		2473,3	100,0			16170	12456,3

Источник: данные Росстата⁴; расчеты авторов.

селения p_i (графа 6) и доли общего объема доходов q_i (графа 7), а также выполнены расчеты сумм их произведений (в графах 8-9), необходимых для определения коэффициента Джини по формуле (3).

В результате имеем:

$$G = \left(\sum_{i=1}^{n-1} p_i q_{i+1} - \sum_{i=1}^{n-1} p_{i+1} q_i \right) : 10000 = \\ = (16170,2 - 12456,3) : 10000 = 0,371.$$

Полученное значение коэффициента Джини (0,371) близко к результату, рассчитанному по данным таблицы 1.

В таблице 2 приведены данные в целом по России. Аналогичные таблицы (распределения) разрабатываются также по каждому региону (субъекту Российской Федерации). И по каждому региону определяется *среднедушевой размер денежных* доходов населения. Ранжирование регионов по указанному показателю дает представление о дифференциации среднедушевых денежных доходов в региональном разрезе.

Так, в 2013 г. в ранжированном ряду из 80 регионов первое место (с наименьшими доходами) занимала Республика Калмыкия (11311 рублей), затем республики Тыва (13472 рубля), Ингушетия (13821 рубль) и т. д., предпоследнее место - Чукотский автономный округ (52695 рублей) и последнее место - г. Москва (54869 рублей)⁵.

Общее представление о распределении субъектов Российской Федерации по величине среднедушевых денежных доходов населения дает вариационный ряд, приведенный в таблице 3. Нетрудно заметить, что интервалы значений группировочного признака в таблицах 2 и 3 не совпадают. И это вполне объяснимо. В таблице 2 приведено распределение населения по величине среднедушевого денежного дохода, варьиовавшего от значения «меньше 5 тыс. рублей» до «свыше 45 тыс. рублей».

Таблица 3

Распределение регионов России по величине среднедушевых денежных доходов в 2013 г.

Среднедушевые денежные доходы, тыс. рублей в месяц	Число регионов
До 14,0	3
14,1-16,0	7
16,1-18,0	9
18,1-22,0	33
22,1-30,0	17
30,1-40,0	8
Свыше 40,0	3
Итого	80

Примечание: группировка регионов выполнена авторами по данным Росстата.

В таблице 3 группировочным признаком является среднедушевой денежный доход, рассчитан-

⁴ Российский статистический ежегодник. 2014: Стат. сб. / Росстат. - М., 2014. URL: http://www.gks.ru/bgd/regl/b14_13/IssWWW.exe/Stg/d01/06-20.htm.

⁵ Там же. URL: http://www.gks.ru/bgd/regl/b14_13/IssWWW.exe/Stg/d01/06-07.htm.

ный как *средняя величина по региону*, и минимальное значение такого показателя, как было отмечено выше, равно 11311 рублям (для Республики Калмыкия), а максимальное значение - 54869 рублям (для г. Москвы). И в таблице 3 речь идет о совокупности *регионов*, а не о совокупности *населения*.

Таким образом, это две разные по своей сути группировки. И если данные таблицы 2 позволяют рассчитать коэффициент концентрации Джини для показателя общих денежных доходов населения (как это было показано выше), то по данным таблицы 3 корректно этого сделать нельзя.

Следует отметить, что алгоритм расчета коэффициента Джини по данным вариационного ряда, рассмотренный выше на примере таблицы 2, применим лишь в тех случаях, когда *группировочный признак непосредственно относится к единицам, распределение которых представлено в вариационном ряду*. В таблице 2 это условие соблюдено: группировочный признак (среднедушевой денежный доход) рассчитан на одного человека и единицами распределения является совокупность отдельных лиц (население). В таблице 3 отмеченное условие не соблюдено: группировочный признак рассчитан на человека, а распределяются *регионы*, и по этим данным непосредственно определить долю общих доходов населения по выделенным группам регионов невозможно, а следовательно, не может быть рассчитан и коэффициент Джини как показатель неравномерности распределения общих доходов населения по регионам.

Для расчета коэффициента Джини по данным таблицы 3 ее необходимо дополнить сведениями о *численности населения* или об *общей сумме денежных доходов населения* (в абсолютном или относительном выражении, то есть в % к итогу) по выделенным группам регионов.

В статистических публикациях нередко встречаются такие таблицы, в которых, наряду с распределениями единиц той или иной совокупности по определенному признаку, одновременно приводятся и данные о распределении общего суммарного объема исследуемого показателя. В качестве примера может служить таблица 4, в которой показано распределение 75 субъек-

тов РФ по объемам поступлений прямых иностранных инвестиций (ПИИ) в 2010 г. и одновременно приведены данные о фактических объемах поступлений ПИИ (и их доли) по различным регионам.

Данные таблицы 4 наглядно свидетельствуют о неравномерности распределения поступлений ПИИ по отдельным регионам России. По этим данным оценить степень концентрации объемов ПИИ в отдельных регионах с помощью коэффициента Джини можно двояко. Во-первых, используя данные первых двух граф, коэффициент Джини можно рассчитать по алгоритму, примененному выше в примере таблицы 2. Во-вторых, поскольку в таблице 4 имеются и фактические данные об объемах поступления ПИИ по группам субъектов РФ (графа 3) и даже определена их доля в общем объеме (графа 4), то, естественно, проще и надежнее (с точки зрения достоверности) воспользоваться этими данными, рассчитав дополнительно, на основе абсолютных показателей 2-й графы относительные частоты (w_i), то есть доли количеств субъектов по выделенным группам (в %).

Таблица 4

Группировка субъектов Российской Федерации по объемам поступлений прямых иностранных инвестиций в 2010 г.

Объем поступлений ПИИ, тыс. долларов США	Количество субъектов РФ	Объем поступлений ПИИ по группе субъектов РФ, тыс. долларов США (y_i)	Доля объема поступлений ПИИ по группе субъектов РФ в объеме поступлений ПИИ в экономику России, в % (d_i)
1	2	3	4
0,1-10000	17	66380	0,5
10001-50000	26	810151	5,9
50001-100000	12	836705	6,0
100001-150000	3	426710	3,1
150001-500000	12	3283420	23,8
Свыше 500000	5	8386677	60,7
Итого	75	13810043	100,0
В 8 субъектов РФ ПИИ не поступали			

Источник: данные Росстата⁶.

Все последующие расчеты, связанные с определением коэффициента Джини, приведены в расчетной таблице 4а.

⁶ Иностранные инвестиции в 2010 году // Статистический бюллетень. 2011. № 4. URL: http://www.gks.ru/bgd/regl/b11_04/IssWWW.exe/Stg/d04/2-in-invest.htm.

Группы субъектов РФ по объему поступлений ПИИ, тыс. долларов США	Доля (в %)		Кумулятивные итоги доли (в %)		Расчеты для коэффициента Джини	
	числа субъектов РФ по группам (w_i)	объема поступлений ПИИ по группам (d_i)	количества субъектов РФ (p_i)	объема ПИИ (q_i)	$p_i q_{i+1}$	$p_{i+1} q_i$
1	2	3	4	5	6	7
0,1-10000	22,6	0,5	22,6	0,5	144,6	-
10001-50000	34,7	5,9	57,3	6,4	710,5	28,7
50001-100000	16,0	6,0	73,3	12,4	1136,2	469,1
100001-150000	4,0	3,1	77,3	15,5	3037,9	958,5
150001-500000	16,0	23,8	93,3	39,3	9330,0	1446,2
Свыше 500000	6,7	60,7	100,0	100,0	-	3930,0
Итого	100,0	100,0			14359,2	6832,5

Отсюда коэффициент Джини для распределения объема поступлений ПИИ по субъектам РФ равен: $G = (14359,0 - 6832,5) : 10000 = 0,753$.

Полученное значение коэффициента Джини (0,753) характеризует высокую степень концентрации объема поступлений прямых иностранных инвестиций в регионы РФ, что было видно и по исходным данным таблицы 4: в пяти регионах, составляющих 6,7% от общего числа регионов, сосредоточено 60,7% всех поступлений ПИИ в Россию. Разумеется, в конкретном исследовании важно расшифровать (назвать), какие именно субъекты РФ попали в последнюю группу.

Примеры неприемлемого расчета коэффициента Джини для средних и относительных показателей

Все рассмотренные выше примеры были приведены с целью показать, что во всех случаях, независимо от того, в каком виде представлены исходные данные, коэффициент Джини рассчитывается для определенного *суммарного* (итогового) показателя, распределяемого по отдельным группам единиц исследуемой статистической совокупности.

Казалось бы, излишне акцентировать внимание на таком ясном и тривиальном вопросе. Но дело в том, что в последнее время все чаще приходится встречаться с научными работами, в которых коэффициент Джини рассчитывается не для *суммарного показателя*, распределяемого по группам, а для *среднего показателя* или *относительного, рассчитанного на душу населения*, что, на наш взгляд, неприемлемо, так как противоре-

чит содержанию показателя Джини как индекса концентрации.

В качестве примера сошлемся на Ежегодный информационно-аналитический доклад Национального института системных исследований проблем предпринимательства «Мониторинг инвестиционной активности в регионах России. Прямые иностранные инвестиции в 2012 г.» [2].

В этой работе для характеристики неравномерности распределения ПИИ по регионам России в 2012 г. построена кривая Лоренца и рассчитан коэффициент Джини (0,84) для показателя *ПИИ на душу населения* [2, с. 6]. При этом в Приложении 3 к упомянутому докладу [2, с. 27] приведена «Методика построения кривой Лоренца и расчета коэффициента Джини», в которой утверждается: «Кривая Лоренца описывает степень неравномерности распределения значений показателя по различным группам. При построении кривой Лоренца в данном мониторинге по оси X было отложено количество регионов (регионы были предварительно ранжированы по возрастанию ПИИ на душу населения), по оси Y были отложены доли накопленных ПИИ на душу населения в общей сумме ПИИ на душу населения.» (выделено нами - Г.Г. и И.М.).

Далее в методике отмечается, что «коэффициент Джини (G) применяется в статистике для измерения степени неравномерности распределения различных показателей...», и приводится известное выражение, отражающее геометрическую сущность коэффициента Джини, но не указывается рабочая формула, по которой рассчитан коэффициент Джини для показателя ПИИ на душу населения (в 2010 г. - 0,8718; в 2011 г. - 0,8676; в 2012 г. - 0,8436).

Трудно понять логику авторов данной методики и доклада. В методике справедливо отмечает-

ся, что и кривая Лоренца, и коэффициент Джини применяются в статистике для характеристики неравномерности *распределения* различных показателей. Но в докладе при решении конкретной задачи - характеристики неравномерного распределения ПИИ по регионам России - и кривая Лоренца, и коэффициент Джини рассчитываются не для показателя общих объемов ПИИ, а для показателя «ПИИ на душу населения», который к распределению отношения не имеет.

Упомянутый доклад - не единственная работа, рекомендующая использовать коэффициент Джини для показателя, рассчитанного на душу населения. В другом источнике - «OECD Regions at a Glance 2013» - коэффициент Джини определяется как индекс, с помощью которого можно оценить региональные различия таких показателей, как ВВП на душу населения, уровень безработицы и др. (e.g. GDP per capita, unemployment rate, etc.)⁷. При этом коэффициент Джини, используемый применительно к показателям, рассчитанным на душу населения, называется там невзвешенным индексом Джини (an unweighted Gini index).

К сожалению, встречаются и другие работы (диссертации), в которых коэффициент Джини рассчитан, например, для характеристики различий в средней заработной плате по муниципальным образованиям региона и других средних показателей по регионам, где авторы ссылаются на методику вышеупомянутого доклада АНО «НИСИПП». Появление подобного рода работ, где коэффициент Джини использован, на наш взгляд, некорректно, и стало побудительным мотивом написания данной статьи с целью обсуждения затронутого вопроса на страницах журнала.

Случаи использования средних величин как исходных данных для расчета коэффициента Джини

Отрицая возможность расчета коэффициента Джини для *средних показателей*, в то же время следует отметить, что на практике встречаются таблицы, содержащие сведения о *средних* значениях определенного показателя по выделенным группам единиц совокупности, по которым можно рассчитать коэффициент Джини, но *не для самих средних показателей*, а для *суммарного пока-*

зателя, непосредственно связанного с приводимыми средними величинами.

Покажем это на конкретном примере. В таблице 5 (графы 1 и 2) приведены данные выборочного обследования организаций России за апрель 2013 г. о среднемесячной номинальной заработной плате по 10%-ным группам работников (расположенным в порядке возрастания размера заработной платы). Данные второй графы в таблице 5 не являются *распределением* среднемесячной заработной платы по группам (как это иногда ошибочно трактуется), а лишь характеризуют средний размер ее по 10 *равновеликим* по численности работников группам. Возникает вопрос: можно ли по данным таблицы 5 рассчитать коэффициент Джини?

Для показателя *средней заработной платы* по группам работников, как уже неоднократно отмечалось, расчет коэффициента Джини невозможен, вернее неприменим, поскольку *средние* величины не характеризуют *распределение* и *концентрацию* (средоточие) исследуемого показателя по группам. Однако по этим данным (графа 2) можно рассчитать коэффициент Джини для *суммарного* показателя - *общего объема начисленной заработной платы*, предварительно определив долю каждой группы в общем объеме начисленной заработной платы (d_{y_i}) на основании известного положения о том, что в совокупности, разбитой на равновеликие по численности единиц (или в %) группы, соотношения между средними показателями по группам тождественны соотношениям между суммарными (общими) объемами показателей по группам. Так, долю каждой группы в общем объеме начисленной заработной платы можно представить, как

$$d_{y_i} = \frac{\tilde{x}_i w_i}{\sum_1^{10} \tilde{x}_i w_i}.$$

Учитывая, что все группы равновеликие ($w_i = 10\%$), приведенное выше выражение можно сократить на $w_i = const$, и тогда долю каждой группы в общем объеме начисленной заработной платы можно определить как отношение групповых средних показателей заработной платы (\tilde{x}_i) к их сумме ($\sum_1^{10} \tilde{x}_i$), то есть

$$d_{y_i} = \frac{\tilde{x}_i w_i}{\sum_1^{10} \tilde{x}_i w_i} = \frac{\tilde{x}_i}{\sum_1^{10} \tilde{x}_i}.$$

⁷ OECD Regions at a Glance 2013. URL: http://www.oecd-ilibrary.org/content/book/reg_glance-2013-en.

Средняя начисленная заработная плата по 10%-ным группам работников за апрель 2013 г.

10%-ные группы работников	Среднемесячная заработная плата, рублей (\bar{x}_i)	Доля, в %		Кумулятивные доли, в %		Расчеты для коэффициента Джини	
		численности работников (w_i)	общей суммы начисленной заработной платы (d_{y_i})	численности работников (p_i)	общей суммы начисленной заработной платы (q_i)	$p_i q_{i+1}$	$p_{i+1} q_i$
1	2	3	4	5	6	7	8
Первая (с наименьшим размером)	6138	10	2,1	10	2,1	-	42
Вторая	9715	10	3,3	20	5,4	54	162
Третья	12893	10	4,4	30	9,8	196	392
Четвертая	16075	10	5,5	40	15,3	459	765
Пятая	19456	10	6,6	50	21,9	876	1314
Шестая	23344	10	7,9	60	29,8	1490	2086
Седьмая	28262	10	9,6	70	39,4	2364	3152
Восьмая	35090	10	11,9	80	51,3	3591	4617
Девятая	45934	10	15,6	90	66,9	5352	6690
Десятая (с наибольшим размером)	97576	10	33,1	100	100,0	9000	-
Итого	(294528)	100	100,0			23382	19220

Источник: данные Росстата⁸; расчеты авторов.

По данным таблицы 5, сумма групповых средних равна: $\sum_1^{10} \bar{x}_i = 294528$ рублей. Сопоставляя с этой суммой показатель *средней* заработной платы в каждой группе, определяем удельный вес каждой группы в *общем объеме* начисленной заработной платы. Так, для первой группы: $d_{y_i} = (6138 : 294528)100 = 2,1\%$; для второй группы $d_{y_i} = (9715 : 294528)100 = 3,3\%$ и т. д. Результаты расчетов d_{y_i} для всех групп приведены в таблице 5 (графа 4); там же в дополнительных графах (5-8) представлены результаты подсчетов кумулятивных долей (p_i и q_i), а также произведения последних, необходимых для расчета коэффициента Джини по формуле (3). В итоге коэффициент Джини равен:

$$G = \left(\sum_{i=1}^{n-1} p_i q_{i+1} - \sum_{i=1}^{n-1} p_{i+1} q_i \right) : 10000 = \\ = (23382 - 19220) : 10000 = 0,416.$$

Данный коэффициент Джини (0,416) характеризует степень неравномерности распределения *общей суммы (фонда) месячной заработной платы* по 10%-ным равновеликим группам работников, но рассчитан он по показателям *среднемесячной*

заработной платы (на человека) по группам. И поскольку все расчеты в данном примере осуществлялись с показателями *среднемесячной заработной платы* (то есть *со средними величинами*), то при поверхностном подходе к решению задачи может сложиться мнение, что и коэффициент Джини рассчитан для этого *среднего показателя*, то есть *для средней заработной платы*.

Не исключено, что именно такого рода случаи, когда коэффициент Джини реально может быть рассчитан по данным о средних показателях в равновеликих по численности единиц группах (как в таблице 5), послужили поводом для появления вышеупомянутых методик, рекомендующих использование коэффициента Джини для *любых* показателей, в том числе и для средней заработной платы, для ПИИ на душу населения по регионам и пр. С последним, на наш взгляд, согласиться нельзя.

Выводы

Коэффициент Джини как *индекс концентрации*, связанный непосредственно с кри-

⁸ Распределение численности работников по размерам начисленной заработной платы (по результатам выборочного обследования организаций за апрель 2013 г.). URL: http://www.gks.ru/free_doc/new_site/population/trud/obsled/trud2013.htm.

вой Лоренца и предназначенный для характеристики неравномерности распределения исследуемого показателя по группам (населения, предприятий, регионов и т. п.), может (и должен) использоваться *лишь для суммарных (итоговых) показателей*, которые могут быть реально подвергнуты *разбиению (распределению) по группам* с определением доли последних в общем итоге. При этом доля каждой группы в общем итоге суммарного показателя может быть определена по разным исходным данным, как было показано в приводимых примерах.

Что же касается средних показателей, то непосредственно для них, на наш взгляд, коэффициент Джини неприемлем, поскольку средние, приведенные для отдельных групп совокупности, не являются *распределением*. Но по ним (по значениям средних в отдельных группах) в некоторых случаях, как это было показано в таблице 5, может быть рассчитана доля суммарного показателя, связанного со средним показателем в каждой группе, и по значениям последних уже может рассчитываться коэффициент Джини для *суммарного показателя*. Такой способ расчета коэффициента Джини - *для суммарного показателя*, но по данным о *средних величинах* - возможен только в тех случаях, когда выделенные группы *равновелики по численности* и при этом *средние* рассчитаны именно на единицу группировки, как это имеет место в рассматриваемом примере. В другом случае, если бы, например, среднемесячная заработная плата была приведена по группам регионов,

то по таким данным расчет коэффициента Джини был бы неприемлем.

Таким образом, при расчете коэффициента Джини всегда важно четко определить (и учитывать), *для какого суммарного показателя* он рассчитывается и *по каким* данным.

Принимая во внимание наличие в публикуемых работах различных точек зрения по расчету и использованию коэффициента Джини, представляется целесообразным и необходимым обсудить этот вопрос в кругу специалистов-статистиков и других заинтересованных лиц. Заранее благодарим всех, кто откликнется на нашу статью.

Литература

1. Громько Г.Л. Теория статистики: Практикум. - 5-изд., испр. и доп. - М.: ИНФРА-М, 2013. - (Высшее образование: Бакалавриат).
2. Краснов С.М., Сайдуллаев Ф.С. Мониторинг инвестиционной активности в регионах России. Прямые иностранные инвестиции в 2012 г. Ежегодный информационно-аналитический доклад. - М.: АНО «НИСИПП», 2013. URL: <http://nisse.ru/upload/iblock/2fe/invest2012.pdf>.
3. Популярный экономико-статистический словарь-справочник / под редакцией И.И. Елисейевой. - М.: Финансы и статистика, 1993.
4. Теория статистики: учебник / под ред. Г.Л. Громько. - 3-е изд., перераб. и доп. - М.: ИНФРА-М, 2013.
5. Шмойлова Р.А., Минашкин В.Г., Садовникова Н.А., Шувалова Е.Б. Теория статистики : учебник / под ред. Р.А. Шмойловой. - 4-е изд., перераб. и доп. - М.: Финансы и статистика, 2006.
6. Энциклопедия статистических терминов в 8 т. Т. 1. Методологические основы статистики. - М., 2011. URL: http://www.gks.ru/free_doc/new_site/rosstat/st-book11/tom1.pdf.

ON THE USE OF THE GINI INDEX IN ECONOMIC AND STATISTICAL STUDIES

Galina Gromyko

Author affiliation: Lomonosov Moscow State University (Moscow, Russia). E-mail: gromyko@econ.msu.ru.

Irina Matyukhina

Author affiliation: Lomonosov Moscow State University (Moscow, Russia). E-mail: iri1256@yandex.ru.

Frequent cases of incorrect use of the Gini index in economic and statistical studies have forced the authors to express their opinion. This refers to those calculations that consider average and relative (per capita) rates rather than divide total rates into groups. A right choice and a correct use of statistical tools play a crucial role in statistical research for an analysis of valid and reliable statistical data. Even the simplest and most widely known method of statistical data processing requires competent use depending on the original data.

The data provided in the article demonstrates possibilities and conditions of the Gini index use, which is popular among researchers. The concentration index (often referred to as the Gini index in the literature) was originally proposed by the Italian statistician K. Gini to assess the degree of uneven distribution of wealth (income) within a population. The authors proceed from the fact that the concentration index can be used for characterizing inequalities of many other indicators. The article contains description of key provisions of the statistical theory of measuring such properties units collectively as «concentration» and well-known formulas for calculating the Gini

index. The authors emphasize the important condition of this indicator application: raw data must consider the possibility of distribution allocated to the group together, i. e. it must be necessarily presented as group totals, not as indicators per capita or average values.

Keywords: Lorenz curve, Gini coefficient, concentration index, distribution of the population by average per capita money income, distribution of money income by 20% groups of population, statistical grouping, averages, indicators of differentiation.

JEL: C10; C18.

References

1. **Gromyko G.L.** *Teoriya statistiki: Praktikum* [Statistics theory: Tutorial]. Moscow, INFRA-M Publ., 2013. (In Russ.).
2. **Krasnov S.M., Saidullaev F.S.** *Monitoring investitsionnoy aktivnosti v regionah Rossii. Prjamyje inostrannye investitsii v 2012 g. Ezhegodnyj informatsionno-analiticheskiy doklad* [Monitoring of investment activity in the regions of Russia. Foreign direct investments in 2012. The annual informative and analytical report]. Moscow, ANO «NISIPP» Publ., 2013. (In Russ.). Available at: <http://nisse.ru/upload/iblock/2fe/invest2012.pdf>.
3. **Eliseeva I.I.** (ed.) *Populyarnyj ekonomiko-statisticheskiy slovar'-spravochnik* [Popular economic and statistical dictionary-handbook]. Moscow, Finansy i Statistika Publ., 1993. (In Russ.).
4. **Gromyko G.L.** (ed.) *Teoriya statistiki: uchebnik. 3-e izdanie, pererab. i. dop.* [Statistics theory: Handbook. 3rd edition, revised and complemented]. Moscow, INFRA-M Publ., 2013. (In Russ.).
5. **Shmoilova R.A., Minashkin V.G., Sadovnikova N.A., Shuvalova E.B.** *Teoriya statistiki, 4-e izdanie, pererab. i. dop.* [Statistics theory, 4th edition, revised and complemented]. Moscow, Finansy i Statistika Publ., 2006. (In Russ.).
6. *Entsiclopediya statisticheskikh terminov v 8 t. T 1. Metodologicheskie osnovy statistiki* [The encyclopediya of statistical terms in 8 vol. Vol. 1: Methodological basics of statistics]. Moscow, 2011. (In Russ.). Available at: http://www.gks.ru/free_doc/new_site/rosstat/stbook11/tom1.pdf.

ПОДПИСКА - 2015

Продолжается подписка на журнал «Вопросы статистики» на 2-е полугодие 2015 г., которую можно оформить во всех отделениях почтовой связи ФГУП «Почта России» и в альтернативных предприятиях России, стран СНГ и Балтии по каталогу агентства Роспечать «Газеты. Журналы» (подписные индексы 70127, 71807) или по объединенному каталогу «Пресса России» (подписной индекс Т71807), а также через АНО ИИЦ «Статистика России».

С 2003 г. выпускается электронная версия журнала. Вы можете оформить годовую подписку на электронную версию журнала или заказать отдельные номера, отправив на адрес редакции письмо-заявку.

Контактный телефон: **+7 (495) 607 42 52**

E-mail: **shop@infostat.ru**

Сайт: **<http://www.infostat.ru>**

Адрес редакции: **107450 Москва, ул. Мясницкая, 39, стр. 1.**