Л. А. Золотухина
докт. физ.-мат. наук, профессор кафедры прикладной математики и математического моделирования Санкт-Петербургского государственного морского технического университета

Д. В. Тихоненко
аспирант кафедры экономической кибернетики и математических методов в экономике Санкт-Петербургского государственного экономического университета

Г. М. Фридман
докт. техн. наук, зав. кафедрой экономической кибернетики и математических методов в экономике Санкт-Петербургского государственного экономического университета

ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ МЕЖДУ ПОКАЗАТЕЛЯМИ ДИФФЕРЕНЦИАЦИИ НАСЕЛЕНИЯ ПО ДОХОДАМ

Введение

Основной характеристикой достигнутого обществом уровня благосостояния являются размеры доходов населения и степень их дифференциации.

Изучение дифференциации населения составляет одну из актуальных задач экономической статистики. Распределение населения по доходам дает полную информацию для оценки степени его расслоения. Однако сравнивать информацию в таком виде сложно и неудобно. Поэтому на основе распределения обычно рассчитываются следующие статистические характеристики:

– обобщающие показатели распределения: среднее, модальное и медианное значения дохода;

– показатели структуры распределения: децильный, квинтильный и доходы других уровней;

– показатели дифференциации доходов населения.

Экономисты чаще всего в качестве показателя дифференциации используют коэффициент Джини G (индекс концентрации доходов), который представляет собой меру отклонения кривой Лоренца от кривой абсолютного равенства (Микроэкономическая статистика, 2004; Лопатников, 2003). При этом чем больше величина G, тем выше неравенство в распределении дохода. Если коэффициент Джини приближается к нулю, это означает состояние абсолютного равенства; если коэффициент Джини стремится к единице, это означает состояние нищего большинства и сверхбогатого меньшинства.

За все время использования коэффициента Джини было разработано множество методов его расчета. Самый широко используемый метод подразумевает кусочно-линейную интерполяцию кривой Лоренца при группированных данных. Алгоритм расчета можно найти в (Венецкий, Венецкая, 1974; Микроэкономическая статистика, 2004). Другие алгоритмы можно найти в (Ogwang, 2000; Yitzhaki, 1991). При подобных методах расчета результаты вычислений коэффициента Джини зависят от способа группировки данных и различаются даже для одной исследуемой совокупности. Существует ряд методов вычисления коэффициента

Джини и по негруппированным данным. Подробный обзор этих методов и литературы можно найти в (Kuan Xu, 2004).

Следует отметить, что коэффициент Джини недостаточно показателен для характеристики дифференциации населения по доходам, поскольку основной вклад в его величину вносят доходы населения в средней части распределения. Дополнительную критику коэффициента Джини можно найти, например, в (Atkinson, 1973; Greselin et al., 2013; Duro, 2008).

Более наглядную картину неравенства населения по доходам дают децильный и квинтильный коэффициенты. Определения этих коэффициентов существуют в различных вариантах. Следуя определению в работах (Курс экономической теории, 2001; Экономическая теория, 2004), децильный (квинтильный) коэффициент — это отношение среднего дохода 10% (20%) населения с самыми высокими доходами к среднему доходу 10% (20%) населения с самыми низкими доходами. В работах (Курс социально-экономической статистики, 2000; Большой экономический словарь, 2007) децильный коэффициент определяется как отношение 9-го дециля распределения населения по доходам к 1-му, а квинтильный — как отношение 8-го дециля ко 2-му. Заметим, что децильный коэффициент в смысле первого определения в отечественной литературе называется также коэффициентом фондов. Разница при расчетах коэффициентов по первому и по второму определениям существенна — их величины могут различаться более чем вдвое. Так как значения коэффициентов, рассчитанные по первому определению, встречаются чаще, его мы и будем придерживаться в дальнейшем.

Еще один показатель дифференциации, который исследуется в работе, — коэффициент Пальма. Он был предложен относительно недавно (Palma, 2006) и базируется на наблюдении, что доходы 50% населения с 4-го по 9-й дециль являются устойчивыми и мало изменяются по времени и по странам. Коэффициент Пальма рассчитывается как отношение средних доходов 10% самого богатого к средним доходам 40% самого бедного населения. Дальнейшее исследование коэффициента Пальма проведено в (Cobham, Sumner, 2013).

Область применения этих показателей достаточно обширна. Они могут быть использованы для сравнения распределения доходов между разными странам, при этом нет зависимости от масштаба экономики сравниваемых стран. Они могут быть использованы для сравнения распределения доходов по разным группам населения одной страны, например для сельского и городского населения. Они также позволяют отслеживать динамику неравномерности распределения дохода на разных этапах. Например, коэффициент Джини для России в 1991 г. был равен 0,24, в 2008 г. — 0,42. В странах северной Европы он находится в диапазоне от 0,2 до 0,3. Децильный коэффициент в СССР равнялся 4,5, в России в 2002 г. составил 14, а в 2013 г. составил 16,2 (Росстат, 2014).

Хорошо известно, что распределение доходов описывается логнормальным распределением. Теоретическое обоснование и эмпирическое подтверждение использования логнормального закона в качестве распределения населения по уровню доходов можно найти в (Aitchison, Brown, 1957; Айвазян, 1970; Айвазян, Мхитарян, 1998). На логнормальном распределении основана и используемая Росстатом методика расчета показателей дифференциации населения по уровню доходов (Госкомстат, 1998).

Поэтому в настоящей работе аналитические зависимости между коэффициентами дифференциации доходов (Джини, децильный, квинтильный, Пальма) найдены также в предположении логнормального закона распределения доходов. Полученные формулы для определения коэффициентов достаточно просты и зависят лишь от одного параметра, который легко оценивается по двум числовым характеристикам распределения, например по среднему и медиане, или по моде и медиане, или по среднему и среднеквадратическому отклонениям. Таким образом, для их вычисления уже не требуется знания всего статистического ряда, а достаточно знания только двух его числовых характеристик.

Найденные в работе функциональные зависимости между коэффициентами дифференциации доходов позволяют находить по значению одного из коэффициентов значения остальных. Ранее эти зависимости подбирались эмпирически, путем построения регрессионных кривых. Например, в работе (Cobham, Sumner, 2013) для аппроксимации была выбрана экспоненциальная функция. В отличие от такого метода в настоящей работе аналитические зависимости между коэффициентами выведены из теоретических соображений, поэтому носят универсальный характер и не зависят от эмпирических данных. Проведенный статистический анализ показал высокую степень адекватности полученных теоретических зависимостей эмпирическим данным с коэффициентом детерминации $R^2 = 0{,}9 \div 0{,}99$.

1. Коэффициенты дифференциации доходов

Предположим, что доходы населения ζ распределены по логнормальному закону с плотностью:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln x - a)^2}{2\sigma^2}}, \; x > 0 \tag{1}$$

где $a, \sigma > 0$ — параметры распределения. Тогда функция распределения доходов $F(x) = P(\zeta < x)$, которую можно интерпретировать как долю населения, имеющего доход меньший x, равна:

$$F(x) = \Phi\left(\frac{\ln x - a}{\sigma}\right),$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Для логнормального закона коэффициент Джини может быть вычислен по формуле (Aitchison, Brown, 1957):

$$G = \operatorname{erf}\frac{\sigma}{2} = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1 , \tag{2}$$

где $\operatorname{erf} x = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dt$ — функция ошибок.

Формулы для вычисления коэффициентов децильного K_D, квинтильного K_Q и Пальма K_P получены в настоящей работе и имеют следующий вид:

$$K_D = \frac{1 - \Phi(u_{0,9} - \sigma)}{\Phi(u_{0,1} - \sigma)} = \frac{1 - \Phi(u_{0,9} - \sigma)}{1 - \Phi(u_{0,9} + \sigma)}; \tag{3}$$

$$K_Q = \frac{1 - \Phi(u_{0,8} - \sigma)}{\Phi(u_{0,2} - \sigma)} = \frac{1 - \Phi(u_{0,8} - \sigma)}{1 - \Phi(u_{0,8} + \sigma)}; \tag{4}$$

$$K_P = \frac{1 - \Phi(u_{0,9} - \sigma)}{\Phi(u_{0,4} - \sigma)}. \tag{5}$$

В формулах (3)–(5) через u_p ($p = 0,1; 0,2; 0,4; 0,8; 0,9$) обозначена p-квантиль стандартного нормального закона $\Phi(u_p) = p$. Вывод этих формул приведен в Приложении 1.

Отметим, что все показатели в формулах (2)–(5) зависят только от одного параметра σ. Оценка этого параметра может быть произведена по двум числовым характеристикам логнормального распределения, например по медиане и моде, или по дисперсии и математическому ожиданию, или по математическому ожиданию и медиане, или по математическому ожиданию и моде (соответствующие формулы см. в Приложении 2).

2. Функциональные зависимости между коэффициентами

Рассматривая попарно уравнения (2) и (3), (2) и (4), (2) и (5), как параметрические формы задания кривой, мы получим следующие функциональные зависимости между коэффициентом Джини и тремя другими коэффициентами дифференциации:

$$\begin{cases} K_D = \dfrac{1-\Phi(1,282-\sigma)}{1-\Phi(1,282+\sigma)} \\ G = \operatorname{erf}\dfrac{\sigma}{2} \end{cases}, \tag{6}$$

$$\begin{cases} K_Q = \dfrac{1-\Phi(0,842-\sigma)}{1-\Phi(0,842+\sigma)} \\ G = \operatorname{erf}\dfrac{\sigma}{2} \end{cases}, \tag{7}$$

$$\begin{cases} K_P = \dfrac{1-\Phi(1,282-\sigma)}{\Phi(-0,253+\sigma)} \\ G = \operatorname{erf}\dfrac{\sigma}{2} \end{cases}. \tag{8}$$

Найденные функциональные зависимости (6)–(8) позволяют находить по любому из коэффициентов значения остальных, например по коэффициенту Джини — значения коэффициентов K_D, K_Q, K_p.

Графики всех полученных зависимостей G от K_D, K_Q, K_p предоставлены на рис. 1, при этом сплошная линия — зависимость $G - K_D$, пунктирная — зависимость $G - K_Q$, штрихпунктирная — зависимость $G - K_p$.

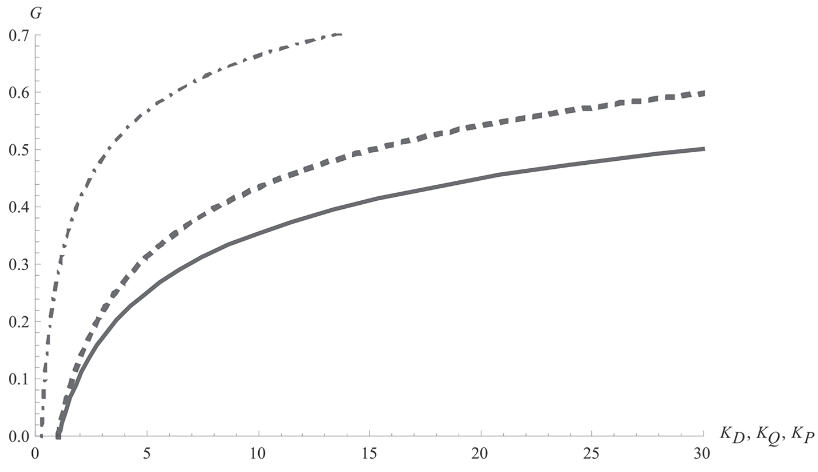

Рис. 1. Графики функциональных зависимостей между коэффициентом Джини и коэффициентами Пальма, децильным и квинтильным

Из графиков видно, что сравнительно небольшие изменения коэффициента Джини соответствуют существенно большим изменениям других коэффициентов, особенно децильного. Таким образом, эти коэффициенты нагляднее характеризуют степень расслоения населения по доходам, чем коэффициент Джини. Например, увеличение коэффициента G на 10% от 0,4 до 0,44 ведет к увеличению коэффициентов K_p, K_Q, K_D соответственно на 25%, 27%, 34%.

3. Анализ эмпирических данных

Были проведены многочисленные расчеты по сравнению эмпирических данных с полученными аналитическими закономерностями. Некоторые из них приведены в этом разделе.

Прежде всего была проведена проверка гипотезы о логарифмической нормальности распределения доходов населения России. Анализировались группированные данные о доходах населения России в 2012 г. (Росстат, «Распределение населения по величине среднедушевых доходов» от 18 апреля 2013 г.). В качестве параметров логнормального распределения брались оценки, полученные методом максимального правдоподобия: $\overline{a} = 9{,}6625$, $\overline{\sigma} = 0{,}8105$. Согласие проверялось по критерию χ^2 Пирсона. Гипотеза принимается на уровнях значимости α = 0,1 0,05: 0 ($\chi^2_{0,1} = 9{,}24$, $\chi^2_{0,05} = 11{,}07$, $\chi^2_{0,01} = 15{,}09$ при наблюдаемом значении $\chi^2 = 1{,}52$), что подтверждает обоснованность использования логнормального закона для аппроксимации эмпирического распределения доходов населения России.

Ниже приведены зависимости между коэффициентом Джини и другими коэффициентами по эмпирическим данным из трех независимых источников. Каждая аналитическая кривая зависимости была рассмотрена как регрессия, при этом был рассчитан коэффициент детерминации R^2 — показатель адекватности модели.

Данные из источника *World Bank* (World Development Indicators). Была использована информация о коэффициенте Джини, а также о доле доходов богатейших и беднейших 10% населения в 35 европейских странах и в России. Информация доступна в сети Интернет (World Bank).

График аналитической зависимости (6) и эмпирические данные приведены на рис. 3. Данные по различным странам на этом и последующем рисунках представлены точками. Вычисленный коэффициент детерминации $R^2 = 0{,}9826$ свидетельствует об очень высокой (very high) степени адекватности модели по шкале Чеддока.

Данные из источника *United Nations Development Programme* (*UNDP*), включающие в себя информацию о коэффициентах Джини и децильном для 139 стран мира, по состоянию на 2009 г. Информация была предоставлена в отчете *UNDP* (UNDP, 2009).

График аналитической зависимости (6) и эмпирические данные представлены на рис. 4. Полученный коэффициент детерминации $R^2 = 0{,}897$, что по шкале Чеддока соответствует высокой (high) степени адекватности модели.

На рис. 4 показана зависимость между коэффициентами Джини и Пальма по данным из источника (Cobham, Sumner, 2013) со ссылкой на *World Bank PovcalNet*; коэффициент Джини брался из данных *World Bank* (World Bank) для тех же стран. Полученный коэффициент детерминации $R^2 = 0{,}993$, что соответствует очень высокой (very high) степени адекватности аналитической зависимости (8) экспериментальным данным.

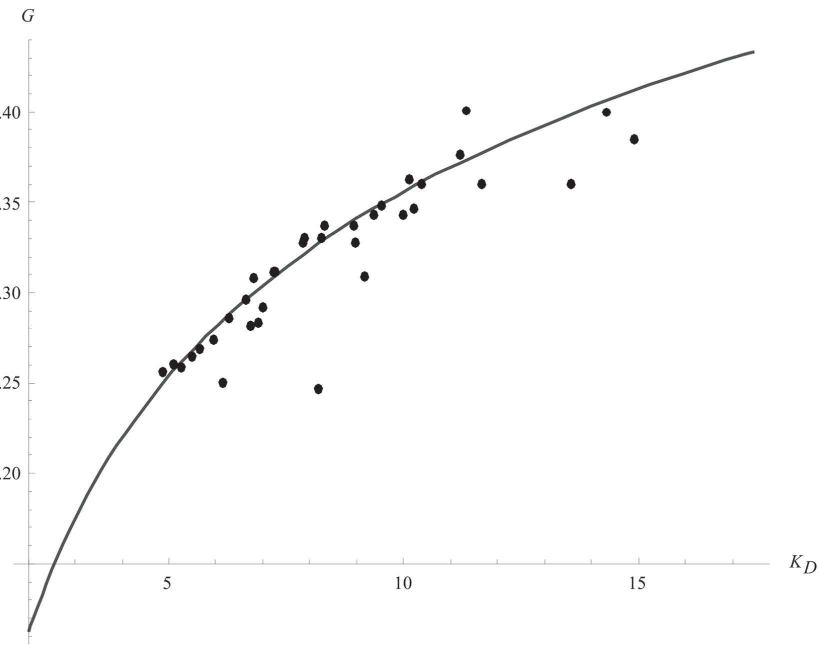

Рис. 2. Сравнение зависимости *G* от K_D с данными World Bank

Приведенные численные результаты свидетельствуют о том, что во всех случаях эмпирические данные хорошо согласуются с полученными теоретическими зависимостями.

Содержащиеся в статье визуализации и расчеты проводились в программном пакете *Wolfram Mathematica* 9.

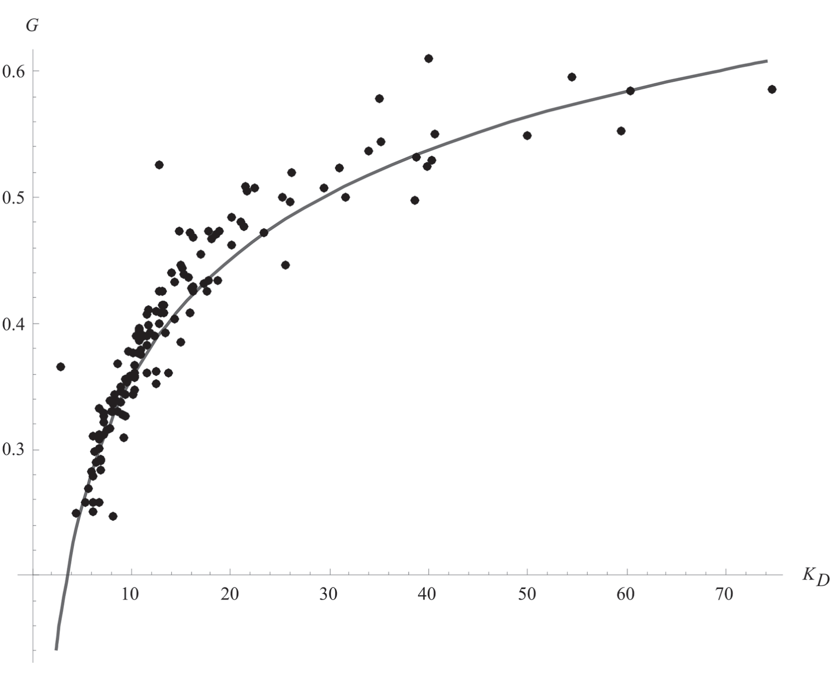

Рис. 3. Сравнение зависимости *G* от K_D с данными *UNDP*

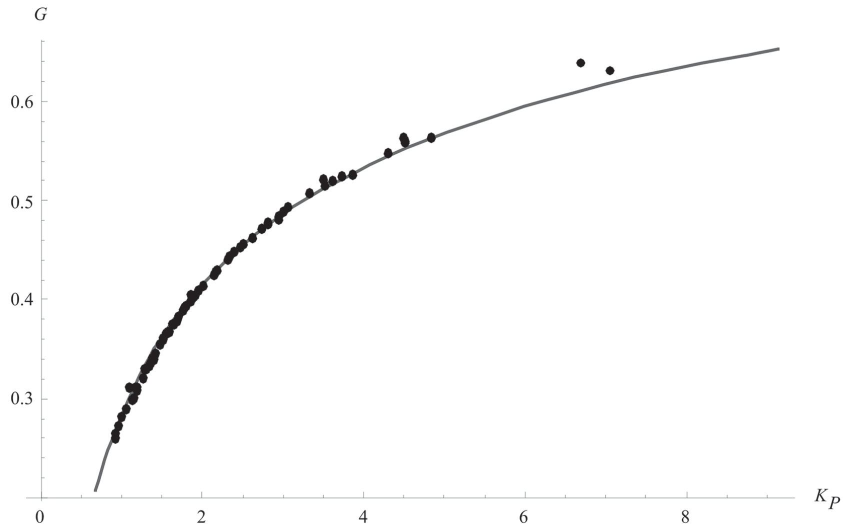

Рис. 4. Сравнение зависимости *G* от K_p с данными World Bank PovcalNet

Заключение

В предположении логнормального распределения разработан аналитический метод определения зависимостей между различными коэффициентами дифференциации населения по доходам (Джини, Пальма, децильный, квинтильный).

Каждый из рассмотренных коэффициентов зависит только от одного параметра, который может быть вычислен по двум характеристикам распределения, а функциональные зависимости между любой парой коэффициентов выведены из теоретических соображений и поэтому носят универсальный характер и не зависят от конкретных эмпирических данных.

Проведенное сопоставление аналитических результатов с опытными данными по большому числу стран, включая Россию, показало высокую степень их адекватности, что позволяет рекомендовать полученные результаты к использованию в практических расчетах.

Приложение 1

При выводе формул (3)–(5) нам понадобятся следующие числовые характеристики логнормального распределения (Aitchison, Brown, 1957), (Айвазян и др., 1983):

математическое ожидание: $$E\xi = \mu_1 = e^{a+\frac{\sigma^2}{2}}; \tag{9}$$

p-квантиль: $$x_p = e^{a+\sigma u_p}. \tag{10}$$

Из следует:

$$u_p = \frac{\ln x_p - a}{\sigma}, \tag{11}$$

где $u_p - p$ — квантиль стандартного нормального закона $\Phi(u_p) = p$.

Децильный коэффициент, квинтильный коэффициент и коэффициент Пальма вычисляются сходным между собой образом как отношение средних доходов крайних групп населения, так что значение каждого из них может быть вычислено по формуле:

$$K = \frac{\int_{x_p}^{\infty} x f(x) dx}{\int_0^{x_q} x f(x) dx} = \frac{\mu_1 - \Psi(x_p)}{\Psi(x_q)},$$

где $\Psi(x)$ — функция неполного математического ожидания (Кендалл, Стьюарт, 1966).

Заметим, что в англоязычной литературе для таких отношений существует специальный термин inter-decile ratios (междецильные коэффициенты) (Cobham, Sumner, 2013).

Для функции неполного математического ожидания (средний доход группы населения, имеющей доход меньше x) при логнормальном распределении получаем следующее выражение:

$$\begin{aligned}\Psi(x) &= \int_0^x x\, f(x)\, dx = \int_0^x x \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - a)^2}{2\sigma^2}} dx = \\ &= e^{a+\frac{\sigma^2}{2}} \int_{-\infty}^{\frac{\ln x - a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\sigma)^2}{2}} dy = \mu_1 \Phi\left(\frac{\ln x - a}{\sigma} - \sigma\right).\end{aligned} \tag{12}$$

Взяв значения p и q для децильного коэффициента равными 0,9 и 0,1 соответственно, для квинтильного коэффициента 0,8 и 0,2, а для коэффициента Пальма 0,9 и 0,4 и используя (11) и (12), получим аналитические формулы (3)–(5) для этих трех коэффициентов.

Приложение 2

Выражения для дисперсии, медианы и моды логнормального закона имеют вид:

дисперсия: $$D\xi = e^{2a+\sigma^2}\left(e^{\sigma^2}-1\right); \tag{13}$$

медиана: $$x_{med} = e^{a}; \tag{14}$$

мода: $$x_{mod} = e^{a-\sigma^2}; \tag{15}$$

Из формул (9), (13)—(15) путем несложных преобразований нетрудно получить формулы для оценок $\bar{\sigma}$ параметра σ при известных значениях числовых характеристик распределения ζ:

– при известных значениях выборочного среднего $\bar{x}$ и выборочного среднего квадратического отклонения S:

$$\bar{\sigma} = (2\ln\frac{s}{\bar{x}} + 1)^{1/2};$$

– при известных значениях выборочного среднего $\bar{x}$ и выборочной медианы $\bar{x}_{med}$:

$$\bar{\sigma} = (2\ln\frac{\bar{x}}{\bar{x}_{med}})^{1/2};$$

– при известных значениях выборочного среднего $\bar{x}$ и выборочной моды $\bar{x}_{mod}$:

$$\bar{\sigma} = (\frac{2}{3}\ln\frac{\bar{x}}{\bar{x}_{mod}})^{1/2};$$

– при известных значениях выборочной медианы $\bar{x}_{med}$ и выборочной моды $\bar{x}_{mod}$:

$$\bar{\sigma} = (\ln\frac{\bar{x}_{mod}}{\bar{x}_{med}})^{1/2}.$$

Источники

Айвазян С. А. Моделирование семейных доходов // Экономика и математические методы. 1970. Т. 6. № 2.

Айвазян С. А., *Енюков И. С.*, *Мешалкин Л. Д.* Прикладная статистика. Основы моделирования и первичная обработка данных. М., 1983.

Айвазян С. А., *Мхитарян В. С.* Прикладная статистика и основы эконометрики. М., 1998.

Большой экономический словарь / под ред. А. Н. Азрилияна. М., 2007.

Венецкий И. Г., *Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе. М., 1974.

Госкомстат. Об утверждении методик расчета баланса денежных доходов и расходов населения и основных социально-экономических индикаторов уровня жизни населения. Постановление Госкомстата РФ от 16 июля 1996 г. № 61.

Кендалл М., *Стьюарт А.* Теория распределений. М., 1966.

Курс социально-экономической статистики / под ред. М. Г. Назарова. М., 2000.

Курс экономической теории: Общие основы экономической теории. Микроэкономика. Макроэкономика. Основы национальной экономики: Учеб. пособие / под ред. А. В. Сидоровича. М., 2001.

Лопатников Л. И. Энциклопедический экономико-математический словарь. М., 2003.

Микроэкономическая статистика / под ред. С. Д. Ильенковой. М., 2004.

Росстат. Распределение общего объема денежных доходов по 20-процентным группам населения и основные показатели социально-экономической дифференциации» от 30 января 2014 г. [Электронный ресурс]. — Режим доступа: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/level/.

Экономическая теория / под ред. А. И. Добрынина, Л. С. Тарасевича. СПб., 2004.

Aitchison J., *Brown J. A.* The lognormal distribution. Cambridge, UK, 1957.

Atkinson A. On the Measurement of Inequality', Reprinted From // Wealth, Income and Inequality / Ed. by *A.* Atkinson. Harmondsworth (UK), 1973.

Cobham A., *Sumner A.* Putting the Gini Back in the Bottle? 'The Palma' as a Policy-Relevant Measure of Inequality / King's International Development Institute. London, 2013.

Duro J. Cross-country Inequalities in Welfare and Its Decomposition by Sen factors: The Virtues of the Theil Index // Applied Economics Letters. 2008. Vol. 15. N 13. P. 1041–1045.

Greselin F., *Pasquazzi L.*, *Zitikis R.* Contrasting the Gini and Zenga Indices of Economic Inequality // Journal of Applied Statistics. 2013. Vol. 40. N 2. P. 282–297.

Ogwang T. A Convenient Method of Computing the Gini Index and Its Standard Error // Oxford Bulletin of Economics and Statistics. 2000. Vol. 62. N 1. P. 123–129.

Palma J. G. Globalizing Inequality: 'Centrifugal' and 'centripetal' Forces at Work / DESA Working Paper № 35. N. Y.: UN Department of Economic and Social Affairs, 2006.

UNDP, Human Development Report 2009. [Электронный ресурс]. — Режим доступа: http://hdr.undp.org/en/media/HDR_2009_EN_Complete.pdf

World Bank. URL: http://data.worldbank.org/indicator/SI.POV.GINI, http://data.worldbank.org/indicator/SI.DST.10TH.10/countries

Xu K. How has the Literature on Gini Index Evolved in the Past 80 Years. Dalhousie University, Department of Economics, 2004.

Yitzhaki S. Calculating Jackknife Variance Estimators for Parameters of the Gini Method // Journal of business and economic statistics. 1991. Vol. 9. N 2.