

Вероятностные смеси в измерениях межтерриториальной дифференциации

Владимир Васильевич Глинский,

Юлия Николаевна Исмайловая

Новосибирский государственный университет экономики и управления «НИХ»,
г. Новосибирск, Россия

В статье в обобщенном виде представлены результаты исследования в области проблемы оценок уровня дифференциации социально-экономического развития территориальных образований Российской Федерации. Аргументируется подход, в соответствии с которым измерение межрегиональной дифференциации предполагает использование смесей вероятностных распределений.

Методика, разработанная авторами и апробированная на реальных данных, позволяет объективно судить о наличии или отсутствии межрегиональной дифференциации. Теоретической платформой данной методики послужила исследовательская гипотеза, в соответствии с которой предполагается, что межтерриториальная дифференциация оценивается по определенному статистическому показателю, отобранному на основе содержательного, качественного анализа. Дифференциация практически отсутствует, если вся статистическая совокупность описывается одним законом распределения вероятностей. В том случае, если статистическая совокупность описывается смесью вероятностных распределений, то следует ожидать наличия значимого уровня дифференциации по рассматриваемому показателю.

В математической статистике традиционно задача разделения смеси распределения вероятностей (оценка параметров плотностей распределений и весовых коэффициентов) решается несколькими схожими технологиями, например, EM-алгоритм, медианные модификации EM, SEM-алгоритм, с учетом специфики выбранного объекта (субъекты РФ – небольшая по объему выборка). В работе для решения этой задачи применен SEM-алгоритм. В качестве информационной базы эмпирического исследования использовалась официальная статистика (открытые данные Федеральной службы государственной статистики).

Выявлена типология субъектов Российской Федерации исходя из двух характеристик во временном интервале 2005–2017 гг.: по уровню преступности (использован показатель, названный «уровень насилия» – число убийств и покушений на убийство в расчете на 100000 человек) и по среднедушевому доходу, что позволило, помимо всего прочего, дополнительно проверить гипотезу о целесообразности традиционного использования тенденций дифференциации уровня преступности в качестве индикатора социально-экономического неравенства. Результаты выполненного исследования, по мнению авторов, могут быть использованы в качестве инструментальной и информационной базы для принятия управленческих решений по государственному регулированию процессов, обеспечивающих выравнивание регионов России по социально-экономическому развитию в процессе реализации мероприятий, направленных на сокращение преступности и экономического неравенства.

Ключевые слова: межтерриториальная дифференциация, социальная статистика, статистика преступности, региональная статистика, математико-статистические методы, смеси вероятностных распределений, разделение смесей распределений, EM-алгоритм, SEM-алгоритм, уровень насилия, среднедушевой доход.

JEL: C02, C12, C13, C38, C40, C60.

doi: <https://doi.org/10.34023/2313-6383-2020-27-3-53-64>.

Для цитирования: Глинский В.В., Исмайловова Ю.Н. Вероятностные смеси в измерениях межтерриториальной дифференциации. Вопросы статистики. 2020;27(3):53-64.

Probabilistic Mixtures in Measurements of Interterritorial Differentiation

Vladimir V. Glinsky,

Yuliya N. Ismailylova

Novosibirsk State University of Economics and Management (NSUEM), Novosibirsk, Russia

The article summarizes research results of the study on the problem of assessing the differentiation level of socio-economic development of territorial units of the Russian Federation. The authors propose an approach to measuring differentiation using mixtures of probability distributions.

This technique was developed and tested on real data that allows one to determine the presence or absence of interregional differentiation. The research hypothesis, that interterritorial differentiation is estimated by a specific statistical indicator selected based on a content, qualitative analysis, served as a theoretical platform of this methodology. Differentiation is practically absent if the entire statistical population is described by a single law of probability distribution. If the statistical population is described by a mixture of probability distributions, then one should expect the presence of a significant level of differentiation by the considered indicator.

In mathematical statistics, the problem of separating a mixture of probability distributions (estimating parameters of distribution densities and weighting coefficients) is traditionally solved using several similar methods. For example, the expectation-maximization (EM) algorithm, median modifications of the EM-algorithm, SEM-algorithm, taking into account the specifics of the selected object (constituent entities of the Russian Federation - a small sample). To solve this problem, the authors used the SEM algorithm. As the information base of the empirical study, official statistics were used (open data from the Federal State Statistics Service).

The typologies of the constituent entities of the Russian Federation were identified based on two characteristics within 2005-2017-time interval. The first one being the level of violence (using the "homicide rate" indicator - the number of homicides and attempted murders per 100000 population). And second - average per capita income, which made it possible, among other things, to additionally test the hypothesis of the traditional use of differentiation trends in the level of violence as an indicator of economic inequality. According to the authors, the results of this study can be used as instrumental and informational support for managerial decisions aimed at regulating the differentiation of Russian regions by the level of violence and economic inequality.

Keywords: interterritorial differentiation, social statistics, crime statistics, regional statistics, mathematical and statistical methods, mixtures of probability distributions, mixtures of distributions separation, EM-algorithm, SEM-algorithm, violence level, average per capita income.

JEL: C02, C12, C13, C38, C40, C60.

doi: <https://doi.org/10.34023/2313-6383-2020-27-3-53-64>.

For citation: Glinsky V.V., Ismailylova Yu.N. Probabilistic Mixtures in Measurements of Interterritorial Differentiation. *Voprosy Statistiki*. 2020;27(3): 53-64. (In Russ.)

Введение

Насилие в различных его проявлениях всегда являлось неотъемлемой частью существования человеческого общества. Но с течением времени насилие становится все более неприемлемым, и необходимо понимать, чем именно обусловлен тот или иной уровень насилия и отражением каких именно социально-экономических реалий он является. Подходы к трактовке понятия «насилие», его классификационные разрезы в содержательном, субъектном, видовом, объектном и т.п. контекстах достаточно подробно рассмотрены в ряде работ юристов, социологов, философов, например, в [1]. Мы рассматриваем насилие в данном исследовании с позиций статистических измерений, соответственно, в конкретной узкой трактовке

- как причинение физического вреда в одной из крайних форм (умышленные убийства).

В 2015 г. ООН были разработаны цели устойчивого развития (ЦУР), официальное название «Преобразование нашего мира: повестка дня в области устойчивого развития на период до 2030 года», которые 25 сентября 2015 г. были приняты 193 странами, в том числе и Россией. ЦУР включают в себя семнадцать целей, шестнадцатая из которых звучит следующим образом: «содействие построению миролюбивых и открытых обществ в интересах устойчивого развития, обеспечение доступа к правосудию для всех и создание эффективных, подотчетных и основанных на широком участии учреждений на всех уровнях».

Важной задачей для выполнения этой цели является существенное снижение распространен-

ности форм насилия и снижение коэффициентов смертности¹. В качестве индикатора выполнения задачи выбран показатель «число жертв умышленных убийств на 100000 человек». Показатель определяется как общее число жертв умышленных убийств, деленное на общее количество населения, в расчете на 100000 человек. Умышленным убийством согласно Международной классификации Преступлений для целей Статистики признается незаконная смерть, причиненная человеку с намерением вызвать ее или нанести серьезныеувечья².

Заметим, что в России существуют некоторые особенности статистического учета числа убийств. Эту статистику ведут правоохранительные органы и органы здравоохранения. Согласно Уголовному кодексу Российской Федерации, убийство - умышленное причинение смерти другому человеку³. Но в случае отсутствия орудия преступления убийство может быть квалифицировано по ст. 111 ч. 4 УК РФ как «умышленное причинение тяжкого вреда здоровью, повлекшее по неосторожности смерть потерпевшего». Еще более усложняет ситуацию то, что по ст. 105 ч. 1 учитываются также и покушения на убийство, а если погибло два или более человек, то в статистике это будет отражено как одно убийство. Органы здравоохранения для квалификации «смерти в результате убийства» используют МКБ-10 - Международную классификацию болезней 10-го пересмотра. На территории Российской Федерации действие МКБ-10 как единого нормативного документа для формирования системы учета и отчетности в системе здравоохранения было введено с 1 января 1999 года приказом Министерства здравоохранения № 170 от 1997 г. С 2006 по 2010 гг. в РФ действовала номенклатура 1997 г., модифицированная в 2005 г., основанная на МКБ-10, которая включала в себя 21 рубрику классификации смертей по внешним причинам. С 2011 г. действует модифицированная номенклатура 2010 г., так же

основанная на МКБ-10, но уже включающая в себя 50 рубрик для классификации смертей по внешним причинам⁴.

Несмотря на то, что сведения об убийствах, которые дают органы здравоохранения, лучше отвечают определению «умышленное убийство» согласно Международной классификации Преступлений для целей Статистики, в разрезе регионов РФ эти данные не являются открытыми. Федеральной службой государственной статистики публикуется информация о числе убийств и покушений на убийство от Министерства внутренних дел.

Обзор исследований

В работах [15-19] нами были рассмотрены ряд принципиальных теоретических и прикладных моментов, касающихся типологии данных с использованием совокупностной и вариационной концепций, оценок территориальной дифференциации, в том числе, на основе решения задачи разделения вероятностных смесей. Исследования об уровне насилия отражены в работах Гилинского Я.И. [1 и 2], Олькова С.Г. [3 и 4], Щербаковой Е.М.⁵, Бадова А.Д. [5 и 6], Лунеева В.В. [7], Лысовой А.В. [8], Land K., McCall P.L., Cohen L.E. [9], Tcherni M. [10], Buonanno P., Vargas J.F. [11], Choe J. [12], Dahlberg M., Gustavsson M [13]. В ряде публикаций отмечается взаимосвязь между уровнем насилия и глубиной социально-экономического неравенства [1-5, 12-13]; приводятся количественные оценки зависимости, в частности, в Глобальном исследовании по проблеме убийств (2011), проведенном профильным Управлением ООН (по наркотикам и преступности) по странам мира отмечена высокая обратная связь между уровнем человеческого и экономического развития и показателями убийств; в странах, отличающихся высокой степенью имущественного расслоения, коэффициенты убийств почти в четыре раза превышают аналогичные показатели

¹ Цели в области устойчивого развития. URL: <https://www.un.org/sustainabledevelopment/ru/sustainable-development-goals/>.

² Международная классификация Преступлений для целей Статистики, ICCS 2015. URL: https://www.unodc.org/documents/data-and-analysis/statistics/crime/ICCS/Russian_iccs_2016_web.pdf.

³ «Уголовный кодекс Российской Федерации» от 13.06.1996 № 63-ФЗ (ред. от 27.12.2019), ст. 105.

⁴ Юмагузин В.В., Кваша Е.А. Смертность от внешних причин смерти в России во второй половине 20 - начале 21 века. Демоскоп Weekly. 2012. № 535-536. URL: <http://www.demoscope.ru/weekly/2012/0535/analit07.php>.

⁵ Щербакова Е.М. Преступность в России, 2018 // Демоскоп Weekly. 2019. № 809-810. URL: <http://demoscope.ru/weekly/2019/0809/barom01.php>.

Щербакова Е.М. Преступность в России, 2017 // Демоскоп Weekly. 2018. № 769-770. URL: <http://demoscope.ru/weekly/2018/0769/barom01.php>.

для стран с более равномерным распределением доходов⁶. Исследование уровня насилия как самостоятельного показателя и как индикатора обострения социально-экономического неравенства достаточно актуально.

Постановка проблемы

В работах, посвященных анализу числа убийств, зачастую рассматривается лишь динамика этого показателя, факторы ее формирующие, определяются описательные статистики, характеризующие этот показатель. С позиций принятия адекватных управлеченческих решений важно понимать следующее:

во-первых, вариация совокупности числа убийств находится в рамках одного закона распределения, или совокупность имеет более сложную структуру, ее части (типы) описываются разными законами распределения, что можно трактовать как наличие дифференциации по этому показателю;

во-вторых, связана ли дифференциация регионов по числу убийств с социально-экономической дифференциацией в стране.

Основной информационной базой послужили сборники Федеральной службы государственной статистики (Росстата)⁷.

Вычисления базируются на использовании пакетов программ: Microsoft Office Excel, StatSoft STATISTICA 12, а также на применении языка программирования Python.

Методика исследования

В данном исследовании предполагается, что дифференциация территорий по некоторому показателю отсутствует, если вся совокупность описывается одним законом распределения; если же совокупность описывается смесью распределений, то предполагается наличие дифференциации по указанному показателю.

Совокупность описывается смесью распределений, если закон распределения представляется в виде взвешенной суммы законов распределения:

$$F(x) = \sum_{i=1}^k p_i F_i(x), \quad (1)$$

где $F(x)$ - функция распределения показателя; $F_i(x)$ - функция распределения i -ой компоненты смеси; p_i - смешивающая вероятность, причем $\sum_{i=1}^k p_i = 1$.

Для смеси нормальных распределений формулу (1) удобнее представить в виде взвешенной суммы плотностей:

$$p(x) = \sum_{i=1}^k p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-a_i)^2}{2\sigma_i^2}}, \quad (2)$$

где $a_i \in \mathbb{R}$ и $\sigma_i^2 > 0$ - математическое ожидание и дисперсия i -ой компоненты смеси; p_i - смешивающая вероятность, причем $\sum_{i=1}^k p_i = 1$.

Задача оценки параметров смеси (оценка параметров плотностей распределений и весовых коэффициентов) называется задачей разделения смеси. Для решения подобных задач обычно применяется ЕМ-алгоритм. Он представляет собой итерационную процедуру, состоящую из двух этапов: на первом этапе (Е-этап) вычисляется условное математическое ожидание логарифма функции правдоподобия при известных значениях наблюдаемых переменных; на втором этапе (М-этап) определяется оценка максимального правдоподобия, которая используется для Е-этапа на следующей итерации.

Однако существенным недостатком ЕМ-алгоритма является неустойчивость относительно начальных данных: изменение одного наблюдения или первоначальных оценок на Е-этапе может привести к резкому изменению итоговых оценок. Для устранения этого недостатка можно использовать медианные модификации ЕМ-алгоритма или SEM-алгоритм (Stochastic EM-algorithm). Но на выборках небольшого объема медианные модификации не дают достаточной точности, поэтому предлагается использовать SEM-алгоритм. Отличие SEM-алгоритма от классического ЕМ-алгоритма заключается в случайном «встряхивании» выборки на каждой итерации (S-этап) перед осуществлением Е- и М-этапов [14].

⁶ Управление Организации Объединенных Наций по наркотикам и преступности. Глобальное исследование по проблеме убийств, 2011 год. Динамика, обстоятельства, данные. URL: https://www.unodc.org/documents/data-and-analysis/statistics/Homicide_Global_homicide_2011_Russian.pdf.

⁷ Регионы России. Социально-экономические показатели. 2010: стат. сб. / Росстат. М., 2010.

Регионы России. Социально-экономические показатели. 2013: стат. сб. / Росстат. М., 2013.

Регионы России. Социально-экономические показатели. 2018: стат. сб. / Росстат. М., 2018.

Оценку параметров смеси k нормальных распределений при помощи SEM-алгоритма можно представить в следующем виде:

1. Задается начальное приближение неизвестных параметров $(p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k)$.

2. Пусть известно значение вектора параметров на m -ой итерации SEM-алгоритма: $(p_1^{(m)}, \dots, p_k^{(m)}, a_1^{(m)}, \dots, a_k^{(m)}, \sigma_1^{(m)}, \dots, \sigma_k^{(m)})$. Вычисляются значения

$$g_{ij}^{(m)} = \frac{\frac{p_i^{(m)}}{\sigma_i^{(m)}} \exp\left\{-\frac{1}{2}\left(\frac{x_j - a_i^{(m)}}{\sigma_i^{(m)}}\right)^2\right\}}{\sum_{r=1}^k \frac{p_r^{(m)}}{\sigma_r^{(m)}} \exp\left\{-\frac{1}{2}\left(\frac{x_j - a_r^{(m)}}{\sigma_r^{(m)}}\right)^2\right\}}.$$

3. Генерируются векторы $\vec{y}_j^{(m+1)} = (y_{1j}^{(m+1)}, y_{2j}^{(m+1)}, \dots, y_{nj}^{(m+1)})$ как реализации случайных векторов с полиномиальным распределением с параметрами 1 и $g_{1j}^{(m)}, \dots, g_{kj}^{(m)}$ ($g_{ij}^{(m)}$ – вероятность того, что $y_{ij}^{(m+1)} = 1$) для $j = 1, \dots, n$.

4. Для каждого $i = 1, \dots, k$ вычисляются

$$v_i^{(m+1)} = \sum_{j=1}^n y_{ij}^{(m+1)}.$$

5. Определяются значения параметров на $(m + 1)$ -ой итерации SEM-алгоритма:

$$\begin{aligned} p_i^{(m+1)} &= \frac{v_i^{(m+1)}}{n}; \\ a_i^{(m+1)} &= \frac{1}{v_i^{(m+1)}} \sum_{j=1}^n y_{ij}^{(m+1)} \cdot x; \\ \sigma_i^{(m+1)} &= \sqrt{\frac{1}{v_i^{(m+1)}} \sum_{j=1}^n y_{ij}^{(m+1)} \cdot (x - a_i^{(m+1)})^2}. \end{aligned}$$

6. В качестве критерия выхода из алгоритма предлагаются следующие условия: $|p_1^{(m+1)} - p_i^{(m)}| < \varepsilon$, $|a_1^{(m+1)} - a_k^{(m)}| < \varepsilon$, $|\sigma_1^{(m+1)} - \sigma_k^{(m)}| < \varepsilon$, для всех $i = 1, \dots, k$ ε можно выбирать сколь угодно малым.

Более подробно этот алгоритм описан в [14].

В случае неизвестного количества компонент смеси их оптимальное количество предлагаются определять при помощи критерия Акаике:

$$k_{omn} = \arg \min_k \left\{ -\ln \prod_{j=1}^n \left(\sum_{i=1}^{i=k} p_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_j - a_i)^2}{2\sigma_i^2}} \right) + 3k - 1 \right\}$$

Для дальнейшей типологии необходимо задать правило, которое бы позволило определять принадлежность объектов к тому или иному типу. Для этого нужно задать пороговые значения, определяющие границы типов. В большинстве случаев визуально границы типов найти невозможно, так как дополнительные компоненты смеси могут лишь утяжелить хвосты распределения, не добавив локальных максимумов плотности распределения. Поэтому для типологизации объектов предлагается рассчитать вероятности их принадлежности к каждой компоненте смеси и считать объект принадлежащим той компоненте, вероятность принадлежности к которой наибольшая. Иными словами, объект со значением показателя x считается принадлежащим i -ой компоненте

смеси, если $\frac{p_i p_i(x)}{\sum_{l=1}^k p_l \cdot p_l(x)} > \frac{p_j p_j(x)}{\sum_{l=1}^k p_l \cdot p_l(x)}$ для всех

$i \neq j$, то есть если i -ая компонента дает наибольший вклад в плотность распределения в точке x . Это условие выполняется тогда и только тогда, когда $p_i p_i(x) > p_j p_j(x)$ для всех $i \neq j$. Таким образом, решив систему из $(k - 1)$ неравенства типа $p_i p_i(x) > p_j p_j(x)$, можно получить граничные или пороговые значения показателя, то есть если значение показателя объекта находится между ними, то объект принадлежит i -ой компоненте смеси. В случае двухкомпонентной смеси пороговые значения можно задать явно.

Предлагаемый подход для определения наличия дифференциации и дальнейшей типологии можно отразить в следующей схеме:

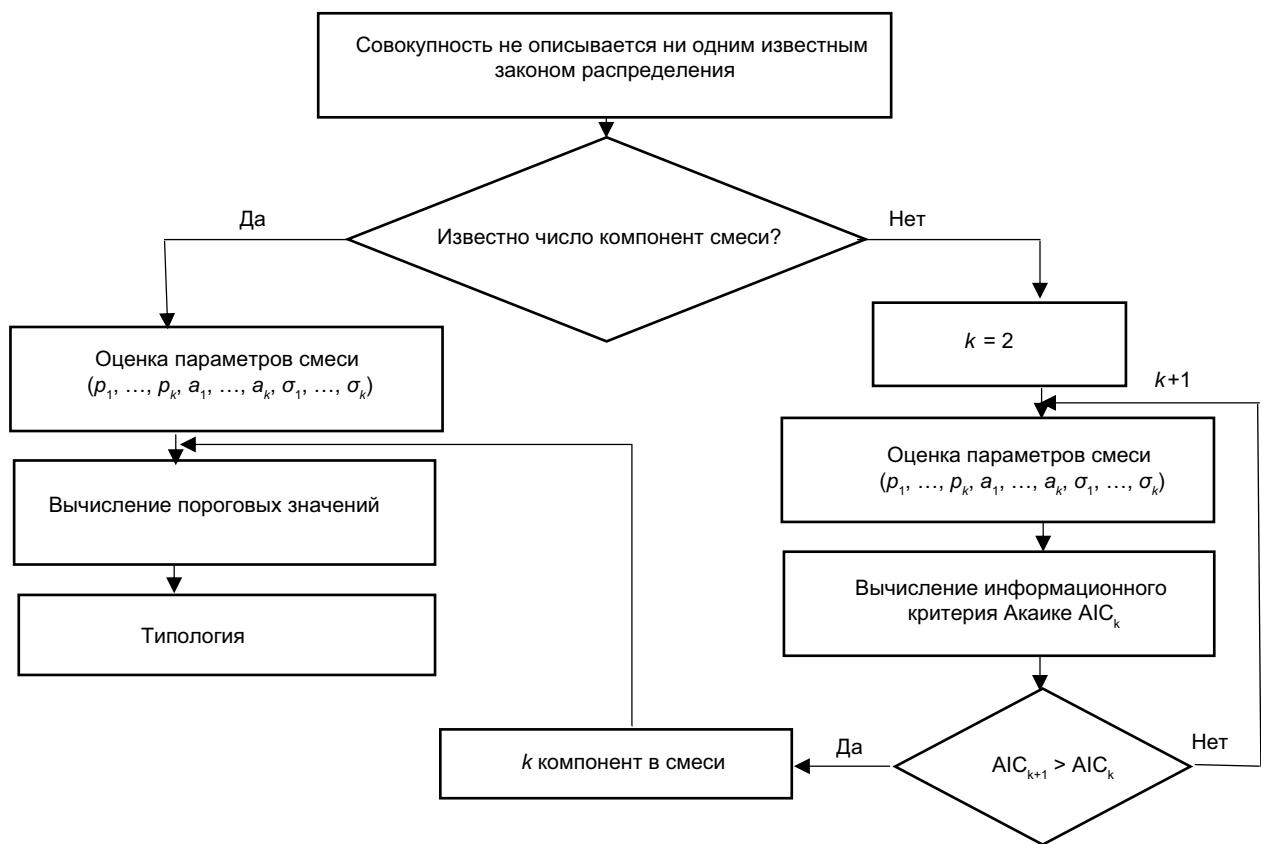


Рис. 1. Методика типологизации на основе смеси вероятностных распределений

Предлагаемую методику можно использовать для территориальных образований различных уровней.

Дифференциация регионов по уровню насилия

Подход к определению дифференциации территорий был применен к субъектам Российской Федерации.

Согласно опубликованным Федеральной службой государственной статистики данным, за исследуемый период число убийств по Российской Федерации снижается. Динамику этого показателя можно увидеть на рис. 2.

В последние годы наблюдается положительная тенденция - снижение числа убийств. Необходимо ответить на вопрос, результатом чего является такое снижение - результатом сдвига распределения и, как следствие, изменения его параметров (в частности, среднего значения) или этот результат обусловлен более сложными изменениями? Были проверены гипотезы о соответствии реальных данных логнормальному или нормальному распределению при помощи критерия согласия Пирсона: на уровне значимости 0,05 гипотезы не

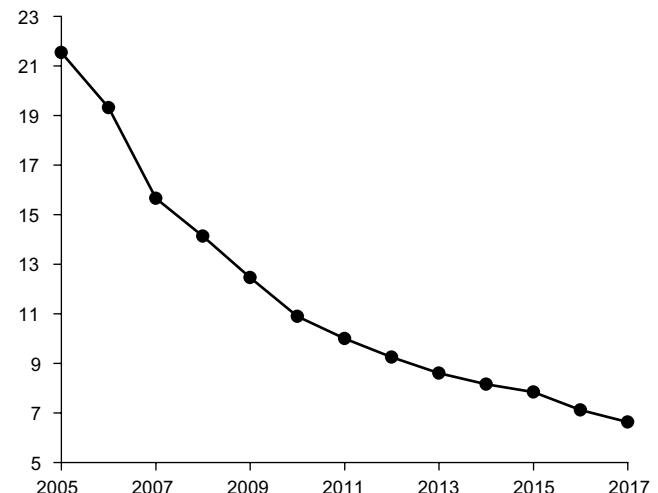


Рис. 2. Динамика числа убийств и покушений на убийство на 100 000 человек в РФ за 2005-2017 гг.

подтверждаются, кроме 2005 и 2006 гг. Это означает, что в 2005 и 2006 гг. число убийств описывалось нормальным распределением, а затем, начиная с 2007 г. число убийств не описывается ни одним из предполагаемых теоретических законов распределения. И, следовательно, снижение среднего значения нельзя объяснить простым изменением числовых характеристик распределения.

На основе имеющихся выводов было выдвинуто предположение, что количество убийств описывается смесью двух или более нормальных распределений, соответственно есть дифференциация (качественные отличия) территорий по уровню насилия.

В результате применения SEM-алгоритма и использования информационного критерия Акаике для оценки количества компонент смеси для всех исследуемых лет были сделаны следующие выводы: начиная с 2007 г., число убийств описывается двухкомпонентной смесью нормальных распределений, причем вторая компонента характеризуется более высокими показателями числовых характеристик - среднего значения и стандартного отклонения.

Необходимо отметить, что для двухкомпонентной смеси нормальных распределений формула (2) перепишется следующим образом:

$$p(x) = p \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}} + (1-p) \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-a_2)^2}{2\sigma_2^2}},$$

где $a_1, a_2 \in \mathbb{R}$ и $\sigma_1^2, \sigma_2^2 > 0$ и - математические ожидания и дисперсии компонент смеси, p - смешивающая вероятность.

Отметим также, что в случае двухкомпонентной смеси уровень дифференциации можно определить экспресс-оценкой - через дисперсию альтернативного признака (максимум дифференциации регионов по числу убийств достигается в случае, когда $p = 0,5$, то есть все объекты, подлежащие типологизации делятся поровну на два типа).

В таблице 1 представлены оценки параметров компонент смеси и смешивающей вероятности.

Таблица 1

Оценки параметров смесей распределений

Год	p	a_1	σ_1	a_2	σ_2
2007	0,73	15,03	4,67	34,51	10,9
2008	0,98	13,94	4,94	37,62	12,17
2009	0,86	12,39	3,9	31,99	11,94
2010	0,79	11,27	3,75	30,27	13,41
2011	0,77	10,06	3,33	25,72	7,95
2012	0,78	9,32	3,26	26,12	9,84
2013	0,69	7,55	1,90	18,14	6,41
2014	0,79	8,26	3,36	23,54	8,94
2015	0,79	8,17	3,25	23,37	7,88
2016	0,77	6,68	2,41	18,81	6,75
2017	0,74	6,09	2,06	16,00	5,26

В связи с тем, что рассматриваемый показатель описывается смесью двух нормальных распределений, представляется естественным разделить все регионы РФ на два типа: «безопасные» регионы - регионы с низким средним числом убийств, распределение которых с большей вероятностью описывается первой компонентой смеси, и «криминальные» регионы - регионы с высоким числом убийств, распределение которых с большей вероятностью описывается второй компонентой смеси. Для такого разделения предлагается

$$\text{вычислить } p_{1i} = \frac{pp(x_i; a_1; \sigma_1)}{pp(x_i; a_1; \sigma_1) + (1-p)p(x_i; a_2; \sigma_2)}$$

$$p_{2i} = \frac{(1-p)p(x_i; a_2; \sigma_2)}{pp(x_i; a_1; \sigma_1) + (1-p)p(x_i; a_2; \sigma_2)} \text{ - вероятности}$$

принадлежности субъекта РФ, для которого исследуемый показатель составляет x_i , к первому и

второму классу соответственно. Субъект считается принадлежащим первому классу в случае, когда $p_{1i} > p_{2i}$, и ко второму - в противоположном случае. Это неравенство можно переписать следующим образом:

$$p \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}} > (1-p) \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-a_2)^2}{2\sigma_2^2}}.$$

Его решение сводится к решению квадратного уравнения:

$$x^2 \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) - 2x \left(\frac{a_1}{\sigma_1^2} - \frac{a_2}{\sigma_2^2} \right) + \frac{a_1^2}{\sigma_1^2} - \frac{a_2^2}{\sigma_2^2} + 2 \ln \frac{1-p}{p} - 2 \ln \frac{\sigma_2}{\sigma_1} = 0.$$

При $\sigma_1 < \sigma_2$ условие принадлежности субъекта РФ, для которого исследуемый показатель состав-

ляет x_i , к первому классу в итоге можно записать следующим образом:

$$\tilde{x}_1 \leq x_i \leq \tilde{x}_2$$

где \tilde{x}_1 и \tilde{x}_2 – корни квадратного уравнения, указанного выше.

Одно из решений квадратного уравнения оказывается отрицательным, поэтому принадлежность к первому и второму классам (типам) определяется при помощи единственного порогового

значения: в случае, когда число убийств на 100000 населения для данного региона больше порогового значения, то субъект относится к «кriminalным», если же этот показатель меньше порогового значения – то к «безопасным».

В таблице 2 за исследуемый период представлены «кriminalные» субъекты Российской Федерации, а также пороговое значение для каждого года. Курсивом выделены устойчивые субъекты РФ, которые всегда попадали во вторую группу регионов.

Таблица 2

«Кriminalные» субъекты Российской Федерации

Год	Пороговое значение (x_b)	Субъект Российской Федерации
2007	24,87	Ненецкий автономный округ, Республика Бурятия, Республика Тыва, Забайкальский край, Иркутская область, Республика Саха (Якутия), Камчатский край, Приморский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2008	30,31	Ненецкий автономный округ, Республика Бурятия, Республика Тыва, Забайкальский край, Магаданская область
2009	22,3	Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Иркутская область, Магаданская область, Сахалинская область, Чукотский автономный округ
2010	20,23	Ненецкий автономный округ, Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Иркутская область
2011	17,67	Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Иркутская область, Кемеровская область, Республика Саха (Якутия), Амурская область, Магаданская область, Еврейская автономная область, Чукотский автономный округ
2012	17,03	Республика Бурятия, Республика Тыва, Забайкальский край, Иркутская область, Амурская область, Республика Саха (Якутия), Хабаровский край, Сахалинская область, Еврейская автономная область, Чукотский автономный округ
2013	11,79	Республика Коми, Ненецкий автономный округ, Новгородская область, Карачаево-Черкесская Республика, Пермский край, Курганская область, Курганская область, Ямало-Ненецкая автономная область, Республика Алтай, Республика Бурятия, Республика Тыва, Республика Хакасия, Забайкальский край, Иркутская область, Кемеровская область, Республика Саха (Якутия), Приморский край, Хабаровский край, Амурская область, Магаданская область, Сахалинская область, Еврейская автономная область, Чукотский автономный округ
2014	16	Курганская область, Республика Алтай, Республика Бурятия, Республика Тыва, Республика Хакасия, Забайкальский край, Иркутская область, Приморский край, Амурская область, Магаданская область, Чукотский автономный округ
2015	15,74	Ненецкий автономный округ, Курганская область, Республика Алтай, Республика Бурятия, Республика Тыва, Забайкальский край, Амурская область, Магаданская область, Сахалинская область
2016	12,28	Республика Коми, Ненецкий автономный округ, Республика Алтай, Республика Бурятия, Республика Тыва, Республика Хакасия, Забайкальский край, Иркутская область, Республика Саха (Якутия), Хабаровский край, Амурская область, Магаданская область, Еврейская автономная область, Чукотский автономный округ
2017	10,68	Архангельская область, Ненецкий автономный округ, Курганская область, Республика Алтай, Республика Бурятия, Республика Тыва, Республика Хакасия, Забайкальский край, Иркутская область, Кемеровская область, Республика Саха (Якутия), Хабаровский край, Амурская область, Магаданская область, Сахалинская область, Еврейская автономная область, Чукотский автономный округ

Кроме того, в качестве еще одного показателя дифференциации для Российской Федерации были рассчитаны коэффициенты Джини (G) по

числу убийств и покушений на убийства за исследуемый период.

Таблица 3

Динамика коэффициента Джини по числу убийств и покушений на убийство за 2007-2017 гг.

Год	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
G	0,434	0,433	0,416	0,423	0,421	0,424	0,419	0,469	0,463	0,463	0,465

Стоит отметить, что определенной тенденции в поведении коэффициента Джини и в числе «кriminalных» регионов РФ не наблюдается. И несмотря на значительное снижение числа убийств по стране в целом, нельзя сделать достоверный вывод о снижении дифференциации регионов по этому показателю.

Дифференциация регионов по уровню доходов

Для проверки гипотезы о том, что число убийств связано с глубиной социально-экономического неравенства, предложенная методика была применена к данным о среднедушевых доходах по субъектам РФ.

За период 2005-2017 гг. гипотеза о том, что среднедушевые доходы описываются нормальным законом, была отклонена на уровне значимости 0,05, что также позволяло предположить, что среднедушевые доходы описываются смесью двух или более нормальных распределений. И в результате применения предложенной выше методики были получены следующие выводы: за исследуемый период среднедушевые доходы описываются смесью двух нормальных распределений, что позволяет разделить все субъекты РФ на два типа - «бедные» регионы и «обеспеченные» регионы. Результаты типологизации представлены в таблице 4.

Таблица 4

«Обеспеченные» субъекты Российской Федерации

Год	Пороговое значение (x_b)	Субъект Российской Федерации
2005	8755	г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Самарская область, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2006	8739	г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Самарская область, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2007	11327	г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Самарская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2008	13087	г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Самарская область, Пермский край, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2009	16878	Московская область, г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2010	18912	Московская область, г. Москва, Республика Коми, Архангельская область, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Самарская область, Пермский край, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2011	21071	Московская область, г. Москва, Республика Коми, Архангельская область, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Пермский край, Самарская область, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2012	25178	Московская область, г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2013	28023	Московская область, г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2014	30515	Московская область, г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ

Окончание таблицы 4

Год	Пороговое значение (x_b)	Субъект Российской Федерации
2015	34520	Московская область, г. Москва, Республика Коми, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2016	34706	Московская область, г. Москва, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ
2017	35099	Московская область, г. Москва, Ненецкий автономный округ, Мурманская область, г. Санкт-Петербург, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Ямало-Ненецкий автономный округ, Республика Саха (Якутия), Камчатский край, Хабаровский край, Магаданская область, Сахалинская область, Чукотский автономный округ

Курсивом выделены устойчивые субъекты, попадающие в «обеспеченные» регионы за весь исследуемый период. Также необходимо отметить, что в случае среднедушевых доходов, устойчивых субъектов значительно больше. Поведение удельного веса «обеспеченных» регионов среди всех субъектов РФ не носит определенной тенденции, поэтому сделать достоверный вывод

о снижении или увеличении экономической дифференциации за исследуемый промежуток времени нельзя.

Для еще одной характеристики экономической дифференциации в РФ из статистических сборников Росстата⁸ были собраны данные о коэффициенте Джини по доходам, результаты представлены в таблице 5.

Таблица 5

Динамика коэффициента Джини по доходам за 2005-2017 гг.

Год	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
G	0,405	0,410	0,422	0,423	0,422	0,421	0,417	0,42	0,419	0,416	0,413	0,412	0,410

Динамика коэффициента Джини также не демонстрирует определенной тенденции, что не позволяет сделать вывод об уменьшении или увеличении экономического неравенства в России, и не опровергает гипотезу о связи числа убийств с социально-экономической дифференциацией.

Кроме того, в «криминальные» субъекты РФ за исследуемый период попадали регионы как с высокими собственными коэффициентами Джини [такие как, например, Ненецкий автономный округ, Чукотский автономный округ, Республика Саха (Якутия)], так и с низкими собственными коэффициентами Джини (например, Республика Алтай, Республика Хакасия, Республика Тыва). Таким образом, полученные результаты не противоречат выдвинутой гипотезе.

Заключение

В статье предложен подход к определению дифференциации, основанный на смесях ве-

роятностных распределений. Разработанная на основе этого подхода методика позволяет выявить наличие или отсутствие дифференциации территориальных образований по заданному показателю, а в случае наличия дифференциации позволяет выполнить корректную типологию территорий.

Территориальная дифференциация по уровню насилия сложное социально-экономическое явление, поэтому не исключено, что для территорий различного уровня этот процесс может детерминироваться различными факторами. Поэтому продолжением работы видится подобное исследование на страновом уровне.

Прикладное значение исследования заключается в статистической поддержке управленческих решений, направленных на регулирование дифференциации территорий по уровню насилия, что позволит приблизиться к достижению одной из целей устойчивого развития.

⁸ Регионы России. Социально-экономические показатели. 2010: стат. сб. / Росстат. М., 2010.

Регионы России. Социально-экономические показатели. 2013: стат. сб. / Росстат. М., 2013.

Регионы России. Социально-экономические показатели. 2018: стат. сб. / Росстат. М., 2018.

Литература

1. Гилинский Я.И. Социальное насилие: Монография / Я.И. Гилинский. ООО Издательский Дом «Алеф-Пресс», 2013. СПб. 185 с.
2. Gilinskiy Y. Crime in Contemporary Russia // European Journal of Criminology. 2006. Vol. 3. Iss. 3. P. 259-292.
3. Ольков С.Г. Влияние степени неравенства в распределении доходов народонаселения на уровень умышленных убийств на планете. Исследование законов распределения, концентрации и дифференциации неравенства и умышленных убийств на земле в начале XXI столетия // Публичное и частное право. 2010. № 3 (7). С. 7-24.
4. Ольков С.Г. Корреляционный анализ структуры преступности в ее объяснении и прогнозировании, изучение влияния безработицы на различные структурные составляющие преступности в России // Вестник научных трудов юридического факультета «Юристъ» Общая редакция И.Ш. Мухаметзянов, С.Л. Алексеев, научный редактор А.Ю. Епихин. 2015. С. 181.
5. Бадов А.Д. Геокриминогенное положение как фактор преступности // Известия Российской академии наук. Серия географическая. 2009. № 2. С. 48-51.
6. Бадов А.Д. География преступности в России: изменения за постсоветский период // Вестник Московского университета. Серия 5: География. 2009. № 2. С. 64-70.
7. Лунеев В.В. Преступность XX века: мировые, региональные и российские тенденции. Изд. 2-е, перераб. и доп. М.: Волтерс Клювер, 2005. 912 с.
8. Lysova A.V. Homicide in Russia, Ukraine and Belarus / A.V. Lysova, N.G. Shchitov, W.A. Pridemore // Handbook of European Homicide Research: Patterns, Explanations and Country Studies / M.C.A. Liem, W.A. Pridemore (eds). N.Y.: Springer-Verlag, 2012. P. 251-469.
9. Land K., McCall P.L., Cohen L.E., Structural Covariates of Homicide Rates: Are There Any Invariances Across Time and Social Space? // The American Journal of Sociology. Vol. 95. No. 4. (Jan., 1990). P. 922-963.
10. Tcherni M. Structural Determinants of Homicide: The Big Three // Journal of Quantitative Criminology. 2011. 27(4). P. 475-496.
11. Buonanno P., Vargas J.F. Inequality, crime, and the long run legacy of slavery // Journal of Economic Behavior and Organization. 2019. No. 159 (C) P. 539-552.
12. Choe J. Income inequality and crime in the United States // Economics Letters. 2008. Vol. 101. Iss. 1. P. 31-33.
13. Dahlberg M., Gustavsson M. Inequality and crime: separating the effects of permanent and transitory income // Oxford Bulletin of Economics and Statistics. 2008. Vol. 70. Iss. 2. P. 129-153.
14. Королёв В.Ю. ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор. М.: ИПИ РАН, 2007. 94 с.
15. Глинский В.В., Третьякова О.В., Скрипкина Т.Б. О типологии регионов России по уровню эффективности здравоохранения // Вопросы статистики. 2013. № 1. С. 57-68.
16. Глинский В.В., Серга Л.К., Пуляевская В.Л. Статистический инструментарий в решении задач управления развитием территорий // Вопросы статистики. 2014. № 10. С. 14-20.
17. Глинский В.В. Мицеская статистика малого бизнеса. Проблемы статистического изучения турбулентных совокупностей // ЭКО. 2008. № 9 (411). С. 51-62.
18. Исмайилова Ю.Н. Метод моментов как способ декомпозиции смесей вероятностных распределений // Статистика - язык цифровой цивилизации. Сборник докладов II Открытого российского статистического конгресса. 2018. С. 135-142.
19. Исмайилова Ю.Н. О разделении смесей вероятностных распределений при помощи метода моментов // Учет и статистика. 2018. № 4 (52). С. 45-51.

Информация об авторах

Глинский Владимир Васильевич - д-р экон. наук, заведующий кафедрой статистики, Новосибирский государственный университет экономики и управления «НИХ». 630099, г. Новосибирск, ул. Каменская, 56. E-mail: v.v.glinskij@nsuem.ru. ORCID: <https://orcid.org/0000-0002-7149-3020>.

Исмайилова Юлия Николаевна – старший преподаватель, кафедра статистики, Новосибирский государственный университет экономики и управления «НИХ». 630099, г. Новосибирск, ул. Каменская, 56. E-mail: ismaiyllova@gmail.com.

Финансирование

Статья подготовлена в рамках выполнения гранта РФФИ 20-010-00560.

References

1. **Gilinskiy Ya.I.** *Social Violence*: Monograph. Saint-Petersburg: Ltd Publishing House «Alef-Press»; 2013. P. 185. (In Russ.)
2. **Gilinskiy Y.** Crime in Contemporary Russia. *European Journal of Criminology*. 2006;3(3):259-292.
3. **Olkov S.** The Effect of the Inequality Degree in the Distribution of Population Incomes on the Intentional Killings Level on the Planet. Study of the Laws of Distribution, Concentration and Differentiation of Inequality and Intentional Killings on Earth at the Beginning of the XXI Century. *Public and Private Law*. 2010;3(7):7-24. (In Russ.)
4. **Olkov S., Mukhametzyanov I.Sh., Alekseev S.L., Epikhin A.Yu.** (eds.) Correlation analysis of the crime structure in its explanation and forecasting, the study of the unemployment impact on the various structural components of crime in Russia. *Scientific Works Bulletin of the Law Faculty «Jurist»*; 2015. 181 p. (In Russ.)
5. **Badov A.** Geocriminogenic Situation as a Crime Factor. *Proceedings of the Russian Academy of Sciences. Geographical series*. 2009;(2):48-51. (In Russ.)
6. **Badov A.D.** The Geography of Crime in Russia: Changes for the Post-Soviet Period. *Vestnik Moskovskogo universiteta. Seriya 5, Geografiya*. 2009;(2):64-70. (In Russ.)
7. **Luneev V.V.** *Crime in the XX century: Global, Regional and Russian Trends*. 2nd Ed., Rev. Moscow: Wolters Kluwer Publ.; 2005. 912 p. (In Russ.)
8. **Lysova A.V.** Homicide in Russia, Ukraine and Belarus. In: Liem M.C.A., Pridemore W.A. (eds). *Handbook of European Homicide Research: Patterns, Explanations and Country Studies*. New York: Springer-Verlag; 2012. P. 251-469.
9. **Land K., McCall P.L., Cohen L.E.** Structural Covariates of Homicide Rates: Are There Any Invariances Across Time and Social Space? *The American Journal of Sociology*. 1990;95(4):922-963.
10. **Tcherni M.** Structural Determinants of Homicide: The Big Three. *Journal of Quantitative Criminology*. 2011;27(4):475-496.
11. **Buonanno P., Vargas J.F.** Inequality, Crime, and the Long Run Legacy of Slavery. *Journal of Economic Behavior and Organization*. 2019;159(C):539-552.
12. **Choe J.** Income Inequality and Crime in the United States. *Economics Letters*. 2008;101(1):31-33.
13. **Dahlberg M., Gustavsson M.** Inequality and Crime: Separating the Effects of Permanent and Transitory Income. *Oxford Bulletin of Economics and Statistics*. 2008;70(2):129-153.
14. **Korolev V.** *The EM Algorithm, Its Modifications, and Their Application to the Problem of Separating Probability Distributions Mixtures. Theoretical Review*. Moscow: IPI RAN; 2007, 94 p. (In Russ.)
15. **Glinskiy V.V., Tret'yakova O.V., Skripkina T.B.** Typology of Regions of the Russian Federation by Health Care Effectiveness Level. *Voprosy Statistiki*. 2013;(1):57-68. (In Russ.)
16. **Glinskiy V.V., Serga L.K., Pulyaevskaya V.L.** Statistical Tools in Solving the Problems of Managing the Development of the Territories. *Voprosy Statistiki*. 2014;(10):14-20. (In Russ.)
17. **Glinskiy V.V.** Small Business Mythological Statistic. Problems of Turbulent Sets Study. *ECO*. 2008;3(9):51-62. (In Russ.)
18. **Ismaiylova Yu.** The Method of Moments as a Way of Probability Distributions Mixtures' Decomposition. Statistics - the Language of Digital Civilization. In: *Collection of Reports of the II Open Russian Statistical Congress*, 2018. P. 135-142. (In Russ.)
19. **Ismaiylova Yu.N.** On the Separating Mixtures of Probability Distributions by Means of the Method of Moments. *Accounting and Statistics*. 2018;4(52):45-51. (In Russ.)

About the authors

Vladimir V. Glinskiy - Dr. Sci. (Econ.), Head, Department of Statistics, Novosibirsk State University of Economics and Management. 56, Kamenskaya Str., Novosibirsk, 630099, Russia. E-mail: v.v.glinskij@nsuem.ru. ORCID: <https://orcid.org/0000-0002-7149-3020>.

Yuliya N. Ismaiylava - Senior Lecturer, Department of Statistics, Novosibirsk State University of Economics and Management. 56, Kamenskaya Str., Novosibirsk, 630099, Russia. E-mail: ismaiylava@gmail.com.

Funding

This work was financed by the Russian Foundation for Basic Research (RFBR) (Project №20-010-00560).