
Применение метрики Минковского первого порядка в социальных исследованиях

И.Н. Мощенко

Южный федеральный университет, Ростов-на-Дону

Аннотация: Приведен симметричный критерий выбора используемой при интерпретации результатов измерений метрики для шкал различного уровня. На конкретном примере численно исследованы следствия несоответствия в паре шкала - метрика для технологии семантического дифференциала. Анализ проведен на конкретном примере интерпретации результатов анкетирования эмоционального восприятия политических порядков, выполненного в мае 2014 г. на базе РГСУ. Получено, что для изучения групповых статических характеристик допустимо использовать при интерпретации общепринятый подход с Евклидовой метрикой. При исследовании же индивидуальных характеристик и последующем их применении для моделирования необходимо брать метрику, симметрично соответствующую шкале измерения. Манхэттенское расстояние в данном случае. Получены функции распределения эмоционального восприятия местной и центральной властей с использованием этой метрики. Показано, что исследуемая аудитория соответствует принципу Парето в области формирования конфликтогенных процессов.

Ключевые слова: шкала измерений, метрика Евклида, Манхэттенское расстояние, инвариантность, симметричный критерий, семантический дифференциал, эмоциональное восприятие, политический порядок, функция распределения, латентная конфликтогенность.

С теоретической точки зрения любое социологическое измерение представляет собой отображение эмпирической системы (ЭС) с отношениями (соотношениями) в математическую систему (МС) с отношениями (соотношениями) [1]. При этом мерой близости (различия) двух эмпирических объектов является расстояние между их образами в пространстве математической системы. И оно играет одну из ведущих ролей при интерпретации результатов.

Чаще всего в качестве математической системы выступает векторное пространство и для расчетов расстояния d_{12} между двумя объектами 1 и 2 используют Эвклидову метрику:

$$d_{12} = \sqrt{\sum_i (x_{1i} - x_{2i})^2} \quad , \quad (1)$$

где x_{1i} и x_{2i} – i -е координаты объектов, суммирование проводится по всем координатам. Но встречаются отдельные работы, где используются иные

метрики. К примеру, такая широко распространенная среди социологов программа компьютерной обработки результатов как SPSS (статистика для социологов), предлагает пользователям следующие шесть мер различия [2-4]:

- Евклидово расстояние. Для интервальных данных оно используется по умолчанию.

- Квадрат расстояния Евклида.

$$d_{12} = \sum_i (x_{1i} - x_{2i})^2. \quad (2)$$

- Чебышева. Максимальная абсолютная разность между соответствующими координатами векторов.

$$d_{12} = \max_i |x_{1i} - x_{2i}|. \quad (3)$$

- Минковского. Корень p -той степени из суммы p -тых степеней абсолютных разностей между соответствующими координатами векторов.

$$d_{12} = \sqrt[p]{\sum_i |x_{1i} - x_{2i}|^p}. \quad (4)$$

- Блок. Сумма абсолютных разностей между соответствующими координатами векторов. Является частным случаем метрики Минковского первого порядка ($p=1$). Это расстояние также известно как Манхэттенское.

$$d_{12} = \sum_i |x_{1i} - x_{2i}|. \quad (5)$$

- Настроенная. Корень r -той степени из суммы r -тых степеней абсолютных разностей между соответствующими координатами векторов.

$$d_{12} = \sqrt[r]{\sum_i |x_{1i} - x_{2i}|^r}. \quad (6)$$

Следует отметить, что в программе эти меры предлагаются для интервальных шкал. Для этих шкал использование метрики ничем не ограничено, и их выбор зависит от задач интерпретации данных и от предпочтения исследователя. В частности, Евклидово расстояние хорошо отражает различия в среднем, согласовано с такими статистическими процедурами как линейная аппроксимация методом наименьших квадратов и нахождение линейной корреляции. Метрика Чебышева (3) используется, если важны максимальные отклонения между двумя объектами.

Для низких шкал (номинальная и порядковая) это уже не так. Сам тип шкалы накладывает ограничения на метрику. Все шкалы характеризуются определенными допустимыми операциями симметрии, оставляющими инвариантным результат. И это должно отражаться в метрике. Она должна либо оставаться инвариантной, либо преобразовываться ковариантным образом при таких операциях. Интервальные шкалы допускают масштабные преобразования, произвольный выбор начала координат и любой поворот в многомерном пространстве вокруг начала (группа O_n). Инвариантность при изменении масштаба достигается использованием относительных расстояний, а для групп сдвига начала координат и O_n инвариантность выполняется автоматически. В принципе допустимы еще преобразования Декартовой системы координат в любую косоугольную, но тогда и метрика должна трансформироваться соответствующим образом.

Все это справедливо и для номинальных и порядковых шкал. Но они допускают и другие симметричные операции. Чем ниже тип шкалы, тем шире круг таких преобразований. Номинальная шкала определяется не во всем векторном пространстве, а на его дискретной многомерной решетке. Для такой шкалы имеет смысл только операции «равно» или «не равно» [1]. При этом допустимы любые сдвиги узлов решетки, исключая их склеивание. Естественной мерой различия для такого случая является дискретная метрика [5]:

$$d_{12} = \begin{cases} 1, & \text{если } \bar{x}_1 \neq \bar{x}_2 \\ 0, & \text{если } \bar{x}_1 = \bar{x}_2 \end{cases} \quad (7)$$

Нетрудно проверить, что такое расстояние не изменяется при вышеуказанных операциях. Другое дело, что оно малоинформативно, позволяет только делать выводы, совпадают или нет исследуемые свойства.

Более содержательным является покоординатная дискретная метрика, которая вводится аналогичным (7) образом, но по каждой координате отдельно. Общее расстояние между двумя точками определяется суммой

покоординатных расстояний. Такой подход также согласован с допустимыми преобразованиями шкалы, но позволяет уже оценивать степень различия свойств объектов.

Порядковая шкала также задана на дискретной решетке. Здесь допустимы любые топологические преобразования решетки, не нарушающие покоординатных порядков узлов. Вышеупомянутая дискретная метрика при этом сохраняется, но ее использование приводит к понижению шкалы до номинальной и потере информации. Более подходящим в этом случае является Манхэттенское расстояние (5). С одной стороны, оно инвариантно при указанных симметричных преобразованиях дискретной решетки, а с другой - позволяет определять порядок расположения объектов.

Использование при интерпретации результатов измерений метрики, не соответствующей по симметричным свойствам шкале, приводит к завышению уровня последней, что может привести к искажениям. Это не означает обязательного появления ошибок, все зависит от структуры исходных данных и решаемых задач.

Целью данной работы является исследование влияние выбора метрики на результаты обработки измерений, выполненных методом семантического дифференциала. Анализ проведен на конкретном примере интерпретации результатов анкетирования, проведенного в мае 2014 г. на базе РГСУ [6,7]. При этом было опрошено 136 студентов младших курсов, возраста от 17 до 22 лет. Специального анализа по представительности выборки хотя бы на уровне РГСУ не проводилось. Работа носила пилотажный характер, хотя в первом приближении по таким характеристикам как национальный и гендерный состав и социально-экономическое обеспечение рассматриваемую аудиторию можно считать типичной для технического вуза Ростова.

Нас интересовали политические установки студенчества, как когнитивные, так и аффективные составляющие. Для выявления когнитивных компонент использовалась методика прямого опроса.

Полученные при этом результаты приведены в [6,7]. Аффективная составляющая определялась по технологии семантического дифференциала [8]. Ранее на базе этой технологии нами была разработана методика мониторинга и моделирования эмоционального восприятия социальных процессов. Хотя она широко описана в наших предыдущих работах [9-11], здесь мы повторимся, так как ее ключевые моменты необходимы для понимания результатов, представленных в этой работе.

В анкете использовалось 20 биполярных семибальных (от -3 до +3) шкал, образованных парами прилагательных с противоположным смыслом, имеющим эмоциональную окраску. Респондентов просили оценить по этим шкалам четыре типа политических порядка: два реальных, по области и по России в целом. И два идеальных, один положительный, максимально привлекательный для респондента, а второй – отрицательный, вызывающий полное отторжение. В соответствии с технологией семантического дифференциала, при таком опросе происходит проекция эмоционального восприятия исследуемого объекта в абстрактное семантическое пространство используемых шкал [8]. Такая проекция носит чисто субъективный характер, и вопросы об отношении к идеальным конструктам нужны именно для того, что бы задать в каждом индивидуальном семантическом пространстве масштаб и направление оценок.

Для характеристики эмоционального восприятия реальных местного и центрального политических порядков нами использовался уровень восприятия r_x , нормированный от -1 до +1. Крайние точки при этом соответствуют восприятию, совпадающему с эмоциональной реакцией на идеальные конструкты. Интерпретация результатов измерений проводилась на основе теории порождения данных респондентами, известной как модели идеальной точки Кумбса [1], в соответствии с которой относительное расстояние $d_{+l,x}$ в семантическом пространстве между образами реального объекта и идеального положительного конструкта монотонно зависит от

уровня эмоционального восприятия. Чем выше уровень восприятия, тем оно меньше. Для расстояния $d_{-l,x}$ до идеального отрицательного политического порядка эта зависимость имеет противоположный знак.

Отсюда следует, что вышеупомянутые относительные расстояния позволяют оценить искомый уровень восприятия. В линейном приближении с учетом используемой нормировки получается:

$$r_x = (d_{-l,x} - d_{+l,x})/2. \quad (8)$$

Подход Кумбса, основанный на близостях до «идеальной» точки, не ограничивается линейной зависимостью. В дальнейшем в социологии была разработана технология многомерного развертывания, позволяющая определить нужную степень нелинейности в этой зависимости [1]. Однако эта методика имеет статистический характер и работает только на групповом уровне. Ввиду сугубой субъективности семантических пространств расчеты же в нашем случае нужно производить на индивидуальном уровне. Поэтому нами использована другая нелинейная модель, разработанная исходя из концепции типичности в рамках теории катастроф [9-11].

Скорость изменения уровня эмоционального восприятия при отклонении от стационарного состояния в этой модели описывается градиентным образом:

$$\frac{dr_x}{dt} = -\frac{dF}{dr_x}, \quad (9)$$

где F – некая гладкая функция, типичным образом зависящая от уровня восприятия (названная нами потенциалом восприятия). В соответствии с результатами теории катастроф в качестве типичного потенциала можно взять полином четвертой степени:

$$F = \frac{1}{4}r_x^4 + \frac{1}{2}ar_x^2 + br_x, \quad (10)$$

где феноменологические коэффициенты a и b определяются вышеупомянутыми относительными расстояниями:

$$a = (d_{lx} + d_{-lx})/2; \quad b = -(d_{-lx} - d_{lx})/2. \quad (11)$$

Отметим, что потенциал восприятия (10) глобально минимален. И при любых фиксированных значениях феноменологических параметров a и b и произвольном начальном r_x уравнение (9) с течением времени выходит на стационарное значение уровня восприятия. Расчет этих стационарных решений на индивидуальном уровне проводился нами в программе многоагентного моделирования AnyLogic. Для каждого агента (моделирующего респондента) задавалось начальное (нулевое) значение степени восприятия и по уравнениям (9-11) вычислялась динамика его поведения в зависимости от модельного времени, вплоть до выхода всей системы на стационарное состояние. А уже по этим стационарным значениям рассчитывалось распределение эмоционального восприятия по респондентам.

Таким образом, в предлагаемой методике оценки уровня восприятия по исходным данным рассчитываются на индивидуальном уровне относительные расстояния в семантическом пространстве $d_{-l,x}$ и $d_{+l,x}$, по (2) либо (9-11) вычисляются стационарные уровни, а по ним – функции распределения эмоционального восприятия по респондентам. При этом в соответствии с общепринятой методикой во всех предыдущих исследованиях расстояния мы считали в Евклидовой метрике.

Как указано выше, такая метрика совместима с интервальными шкалами. И обычно при интерпретации исходных данных в технологии семантического дифференциала предполагают (явно или неявно, по используемым методам обработки) именно интервальный тип, хотя это не так. Здесь используются биполярные дискретные шкалы. Другими словами, образ в семантическом пространстве задается на дискретной многомерной решетке. Такой подход эквивалентен шкалам порядкового типа. Процедуру завышения типа шкал зачастую обосновывают большой плотностью возможных значений для образа каждого объекта (число значений равно произведению количества шкал на число градаций) и возможным

перекрытием соседних значений из-за известной нестабильности мнений респондентов. При интерпретации полученных результатов по средним характеристикам (сравнение усредненных семантических портретов, факторизация и выявление скрытых латентных факторов, к примеру) с этим можно согласиться. Но при анализе по индивидуальным относительным расстояниям такое завышение типа шкал может привести к погрешностям.

Для иллюстрации возможных искажений на рис. 1 приведена часть квадратной решетки, на которой синей линией показано геометрическое место точек (четверть круга), расположенных на расстоянии 4 от начала координат, определенное в Евклидовой метрике. Узлам решетки соответствуют возможные значения семантических координат. Для дискретных порядковых решеток, как мы говорили ранее, следует использовать Манхэттенское расстояние (5). Красной линией изображено то же самое геометрическое место, но посчитанное уже в правильной метрике (5).

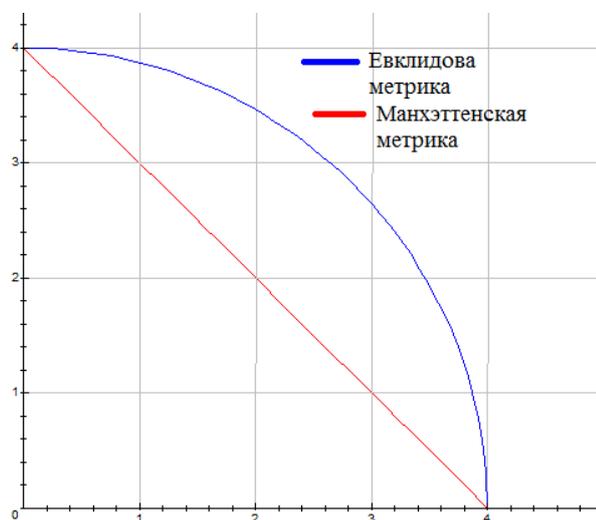


Рис.1 Часть семантической квадратной решетки с изолиниями эмоционального восприятия, посчитанными в разных метриках.

В соответствии с предлагаемой методикой, индивидуальные уровни эмоционального восприятия политических порядков определяются только по относительным расстояниям. И линии, изображенные на рис. 1 являются

изолиниями уровня восприятия. Как видно из рисунка, в зависимости от используемой метрики они проходят через разные узлы, что и приведет к различию в функциях распределения восприятия по респондентам.

Для оценки этого эффекта по исходным данным нами были посчитаны в двух метриках функции распределения эмоционального восприятия местного и центрального политических порядков в линейном приближении (8), представленные на рис. 2.

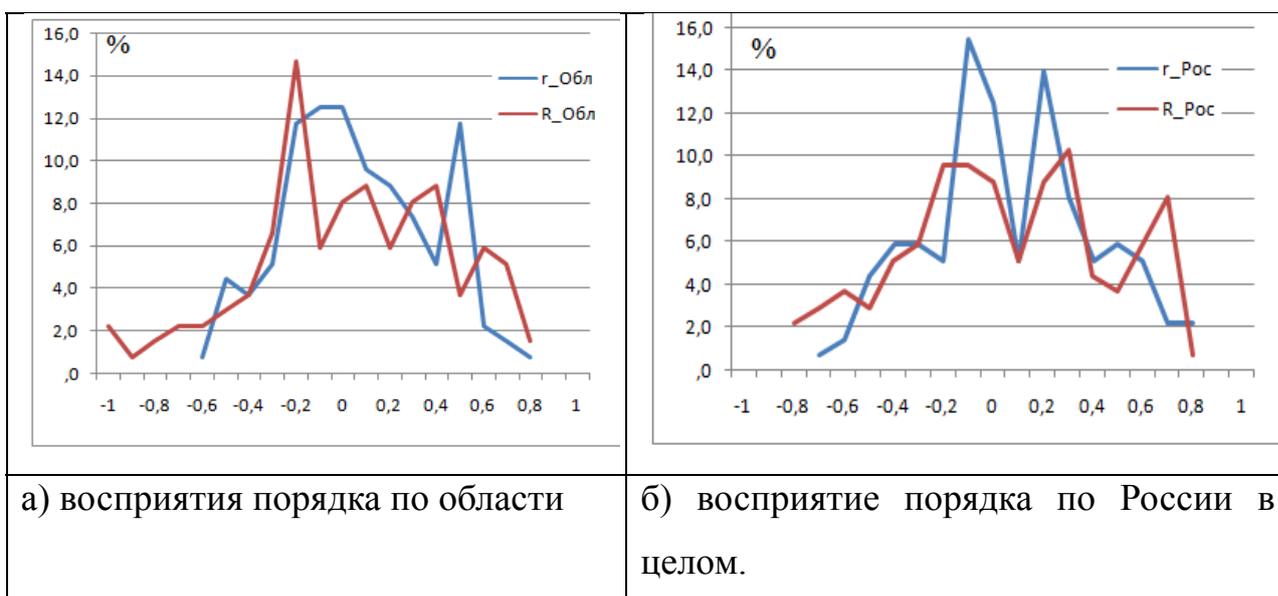


Рис. 2. Функции распределения уровня эмоционального восприятия политических порядков, линейное приближение. г - посчитанные с использованием Евклидовой метрики. R - посчитанные с использованием Манхэттенской метрики.

Рисунок демонстрирует сильную зависимость вида функции распределения от используемой метрики, как для восприятия местного порядка, так и центрального. При смене метрики из Евклидовой в Манхэттенскую, с функцией распределения происходят следующие изменения. Для восприятия местного порядка в нейтральной и средней положительной (до 0,5) области она уменьшается, а в отрицательной увеличивается. В сильно положительной части также наблюдается увеличение. За счет средней части происходит расширение функции по

краям. Для восприятия центрального порядка уменьшается средняя нейтральная область, при возрастании отрицательных и положительных краев, т.е. опять наблюдается расширение функции распределения за счет центральной части. Отметим, что линейное приближение дает приблизительную оценку по сравнению с нелинейной моделью. Здесь мы привели результаты изменений этих функций только лишь потому, что они напрямую свидетельствуют об изменении относительных расстояний при смене метрик (см. (8)).

Функции распределения, полученные по более адекватной нелинейной модели, приведены на рис. 3. Опять же, для двух типов метрик.

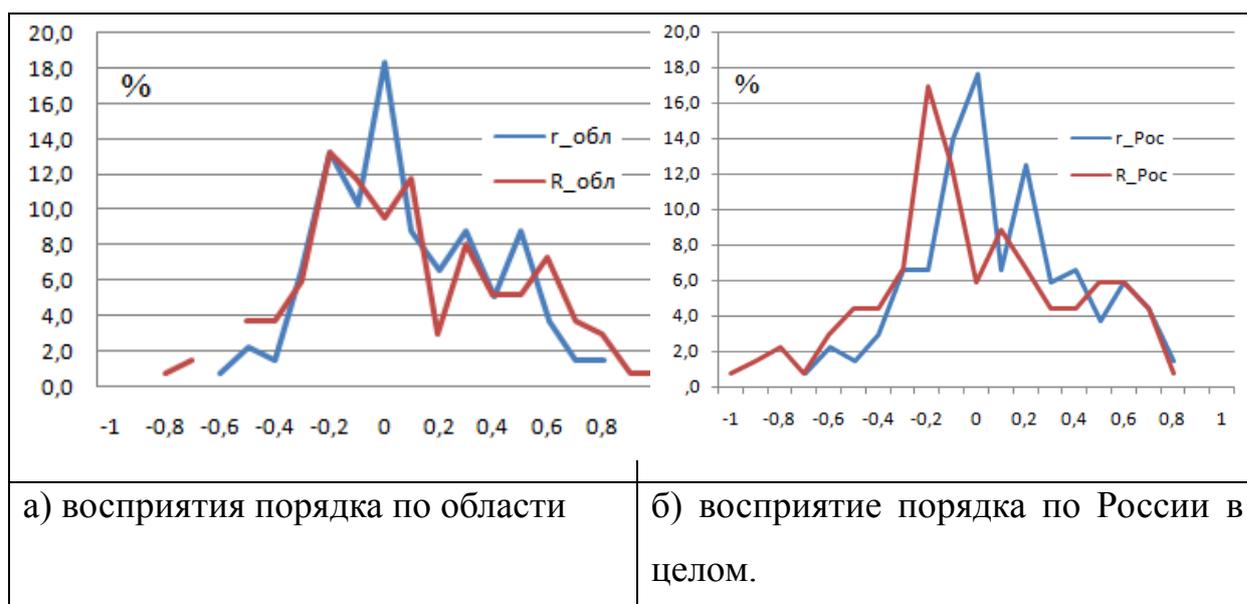


Рис. 3. Функции распределение уровня эмоционального восприятия политических порядков, нелинейная модель. r - посчитанные с использованием Евклидовой метрики. R - посчитанные с использованием Манхэттенской метрики.

Здесь изменения, обусловленные сменой метрик не такие значительные, как в предыдущем случае. Для восприятия местного порядка также наблюдается расползание функции по краям за счет центра. Среднее значение при этом практически не меняется. Для восприятия центрального

порядка отрицательная область расширяется, немного за счет уменьшения слабо положительной части, а в основном происходит сдвиг отрицательной области влево, в более отрицательную часть. Максимальный пик немного смещается из нейтрального положения в отрицательное, хотя среднее значение и в этом случае практически не сдвигается.

Большую наглядность изменений в функциях распределений, вызванных сменой метрик, дает рис. 4., где изображена разность этих функций Δ :

$$\Delta = f_{\text{Матх}} - f_{\text{Евкл}} \quad (12)$$

$f_{\text{Матх}}$ – функция распределения, посчитанная с использованием Евклидовой метрики, $f_{\text{Евкл}}$ – с использованием Манхэттенской метрики.

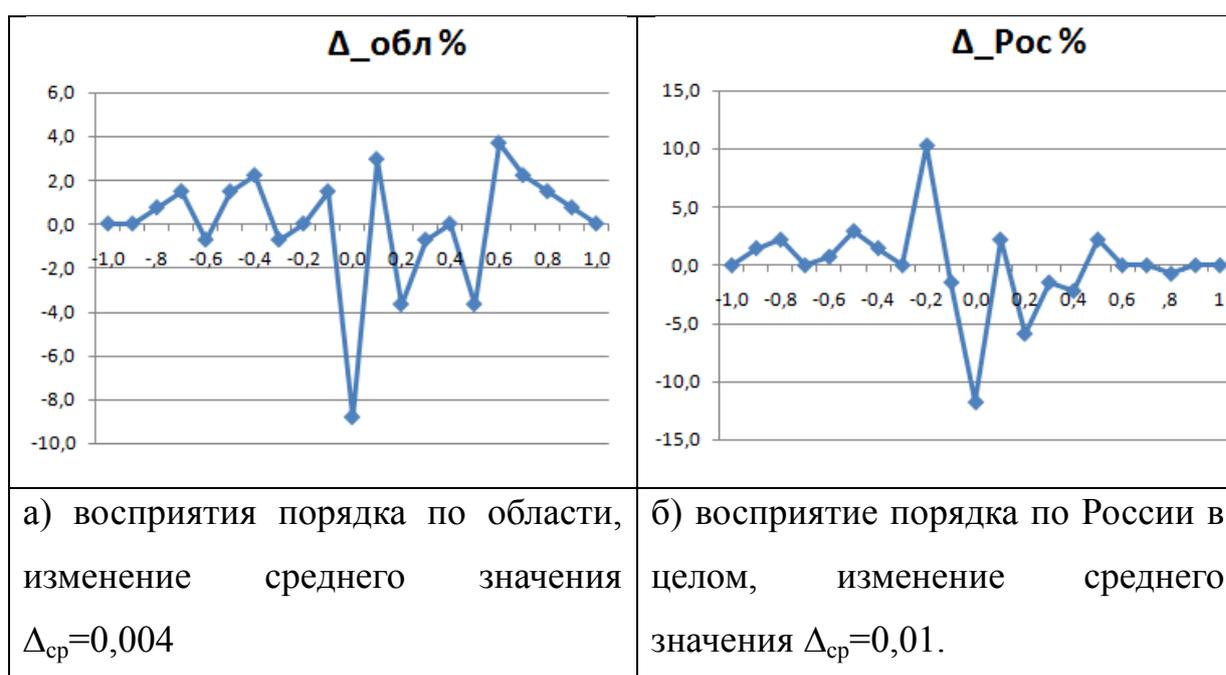


Рис. 4. Разность функций распределение уровня эмоционального восприятия политических порядков, нелинейная модель.

Рисунки демонстрируют, что только в центре распределения изменения достигают 10%. Для восприятия местного порядка это связано с расползанием центрального пика по ширине, а для центральной власти – с

небольшим сдвигом всей функции влево. Для остальной части изменения не превышают 4%. Они носят нерегулярный характер, особых закономерностей не видно, хотя можно говорить об увеличении доли респондентов как с сильно отрицательным восприятием ($\leq -0,4$) порядков, так и с сильно положительным ($\geq 0,5$). Средние значения при этом, для обоих способов расчета практически одинаковы (разница не более 0,01).

Небольшая величина и нерегулярность изменений, связанной со сменой метрик, изрезанность самих функций позволяют сделать вывод, что на групповом статистическом уровне использование не соответствующей шкале метрики не сказывается. Все полученные нами в Евклидовой метрике ранее результаты оценки группового восприятия социальных процессов (см. ссылки в [9-11]) не требуют пересчета. Хотя этот вывод сделан на основании конкретного численного эксперимента, по нашему мнению он может быть обобщен на любое измерение методом семантического дифференциала. Для анализа групповых параметров этим методом допустимо использовать при интерпретации Евклидову метрику.

Однако это не так, если получаемые результаты используются далее для моделирования на индивидуальном уровне. К примеру, полученные нами функции распределения эмоционального восприятия местной и центральной политической власти в дальнейшем планируется применить для оценок рисков протестных выступлений [12]. В соответствии с принципом Парето (в широком смысле), за формирование латентных конфликтогенных процессов (как, впрочем, и всяких других) в любой социальной группе ответственна лишь небольшая ее часть. В данном случае это будет подгруппа, характеризующаяся сильным отрицательным восприятием политических порядков. А как мы видели, именно на отрицательном конце функции распределения, полученные в Евклидовой метрике, занижены. И при индивидуальных расчетах следует пользоваться метрикой, совместимой со шкалой. В нашем случае – расстоянием городских кварталов (5). Это



основной результат представленной работы.

Для исследуемой аудитории такие расчеты, выполненные уже с использованием метрики, соответствующей шкале показали следующие результаты. Примерно 50% респондентов характеризуются нейтральным восприятием (уровень от -0,2 до 0,2) обоих уровней власти, что говорит об их политической отчужденности. Около 33% положительно (уровень больше 0,2) воспринимают политические порядки. Оставшаяся часть опрашиваемых (17%) - отрицательно, с уровнем восприятия меньше -0,2. Они как раз и являются подгруппой риска по отношению к возможности формирования конфликтогенных процессов. Отметим, что эти цифры практически не отличаются от уровня Парето, что свидетельствует о нормальности процессов конфликтогенности в исследуемой аудитории.

Работа выполнена по гранту РФФИ № 14-06-00230а.

Литература

1. Толстова Ю.Н. Измерение в социологии: Курс лекций.- М.: Инфра-М, 1998.- 224 с.
2. Наследов А.Д. SPSS: Компьютерный анализ данных в психологии и социальных науках. СПб.: Питер, 2006. 416 с.
3. Green S. B., Salkind N. J. Using SPSS for Windows and Macintosh: Analyzing and understanding data. Prentice Hall Press, 2010. 480 p.
4. Nie N. H., Bent D. H., Hadlai Hull C. SPSS statistical package for the social sciences. New York. McGraw-Hill, 1970. 363 p.
5. Теслер Г.С. Метрики и нормы в иерархии категориальных семантик и функций. Математические машины и системы. № 2, 2005. С. 63-75.
6. Иванова М.И., Снежков В.И. Динамика политических установок и социальной напряженности среди студенчества РГСУ за 2013-2014 гг.// Инженерный вестник Дона, 2015, №1. URL: ivdon.ru/ru/magazine/archive/n1y2015/2965.

7. Иванова М.И., Мощенко И.Н., Бугаян И.Ф. Анализ отношения студентов РГСУ к политической жизни в середине 2014 года // Инженерный вестник Дона, 2015, №1. URL: ivdon.ru/ru/magazine/archive/n1y2015/2949.
8. Osgood С.Е. The nature and measurement of meaning // "Psychological Bulletin", Vol. 49, No. 3, May, 1952. Pp.197 – 327.
9. Розин М.Д., Мощенко И.Н., Джикаев Д.А. Моделирование политической напряженности методами семантического дифференциала и теории катастроф // «Математический форум. (Итоги науки. Юг России)», 2010, Т. 4 , С. 341-353
10. Мощенко И.Н., Иванова М.И., Бугаян И.Ф. Типичные модели группового эмоционального восприятия политического порядка //Научное обозрение, 2013 г., №2. URL: [sced.ru/ru/index.php?option=com_content &view=article&id=107%3Aq&catid=21&Itemid=18](http://sced.ru/ru/index.php?option=com_content&view=article&id=107%3Aq&catid=21&Itemid=18).
11. Мощенко И.Н., Иванова М.И. Стохастическая интерпретация психосемантической феноменологической модели оценок социальных установок //Инженерный вестник Дона, 2015, №2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2948.
12. Мощенко И.Н., Алботов А.М. Агентно-ориентированная модель конфликтности студенчества КЧР. Инженерный вестник Дона, 2015, №2 ч.2. URL: ivdon.ru/ru/magazine/archive/n2p2y2015/3177.

References

1. Tolstova Ju.N. Izmerenie v sociologii: Kurs lekciy [Measurement in Sociology: lectures]. M.: Infra-M, 1998. 224 p.
 2. Nasledov, A.D. SPSS: Komp'yuternyy analiz dannykh v psikhologii i sotsial'nykh naukakh [SPSS: the Computer analysis of data in psychology and the social Sciences]. SPb.: Piter, 2006. 416 p.
 3. Green S. B., Salkind N. J. Using SPSS for Windows and Macintosh: Analyzing and understanding data. Prentice Hall Press, 2010. 480 p.
-



4. Nie N. H., Bent D. H., Hadlai Hull C. SPSS statistical package for the social sciences. New York. McGraw-Hill, 1970. 363 p.
5. Tesler G.S. Matematicheskie mashiny i sistemy. № 2, 2005. pp. 63-75.
6. Ivanova M.I., Snezhkov V.I. Inzhenernyj vestnik Dona (Rus), 2015, №1. URL: ivdon.ru/ru/magazine/archive/n1y2015/2965.
7. Ivanova M.I., Moshhenko I.N., Bugajan I.F. Inzhenernyj vestnik Dona (Rus), 2015, №1. URL: ivdon.ru/ru/magazine/archive/n1y2015/2949.
8. Osgood C.E. The nature and measurement of meaning. Psychological Bulletin, Vol. 49, No. 3, May, 1952. Pp.197-327.
9. Rozin M.D., Moshhenko I.N., Dzhikaev D.A. Matematicheskij forum. (Itogi nauki. Jug Rossii) , 2010, T. 4, pp. 341-353.
10. Moshhenko I.N., Ivanova M.I., Bugajan I.F. Nauchnoe obozrenie, 2013 g., №2 URL: sced.ru/ru/index.php?option=com_content&view=article&id=107%3Aq&catid=21&Itemid=18.
11. Moshhenko I.N., Ivanova M.I. Inzhenernyj vestnik Dona (Rus), 2015, №2. URL: ivdon.ru/ru/magazine/archive/n2y2015/2948.
12. Moshhenko I.N., Albotov A.M. Inzhenernyj vestnik Dona (Rus), 2015, №2 ч.2. URL: ivdon.ru/ru/magazine/archive/n2p2y2015/3177.