МЕТОДЫ И МЕТОДОЛОГИЯ

DOI: 10.14515/monitoring.2020.4.1253

А. Д. Зотьева

О СОЦИОЛОГИЧЕСКИХ ИНДЕКСАХ С НОРМАЛЬНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ

Правильная ссылка на статью:
Зотьева А. Д. О социологических индексах с нормальным законом распределения // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 4. С. 4—16. https://doi.org/10.14515/monitoring.2020.4.1253.
For citation:
Zoteva A. D. (2020) On Sociological Indices with Normal Distribution Law. *Monitoring of Public Opinion: Economic and Social Changes.* No. 4. P. 4—16. https://doi.org/10.14515/monitoring.2020.4.1253. (In Russ.)

О СОЦИОЛОГИЧЕСКИХ ИНДЕКСАХ С НОРМАЛЬНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ

ЗОТЬЕВА Анна Дмитриевна — студентка Высшей школы современных социальная наук, Московский государственный университет имени М.В. Ломоносова, Москва, Россия
E-MAIL: anna_zoteva1011@mail.ru
https://orcid.org/0000-0003-4888-7731

Аннотация. В социологии широко применяются индексы, выраженные линейно через доли положительных и отрицательных ответов респондентов. Например, индексы социальных и потребительских настроений, индекс ожиданий безработицы, публикуемые «Левада-центром», индексы социальных оценок от ВЦИОМ. Подобные показатели применяются в различных областях знания помимо социологии (экономика, медицина и т.д.). Вопрос о нормальности их распределений имеет важное значение для анализа и прогнозирования, так как методы математической статистики наиболее эффективны и доступны применительно к нормальным величинам.

В статье выдвинута и теоретически обоснована гипотеза о том, что такие индексы имеют нормальные законы распределения на тех промежутках времени, где они являются квазистационарными. Это условие проявляет себя в том, что сезонная компонента индекса отсутствует, а его тренд является (приблизительно) стационарным. Полученный результат можно использовать для предварительных оценок размеров выборок при опросах, что могло бы снизить расходы на социальный мониторинг.

ON SOCIOLOGICAL INDICES WITH NORMAL DISTRIBUTION LAW

Anna D. ZOTEVA[1] *— Student, Higher School of Modern Social Sciences*
E-MAIL: anna_zoteva1011@mail.ru
https://orcid.org/0000-0003-4888-7731

[1] Lomonosov Moscow State University, Moscow, Russia

Abstract. Indices expressed linearly through the shares of positive and negative answers of respondents are widely used in sociology. They are indices of social and consumer moods, the unemployment expectations index published by Levada Center, VCIOM's social assessments indices. These indicators are also used in other fields (economy, health care, etc.). The normality of their distributions is an important issue in the analysis and forecasts as the methods of mathematical statistics are more effective and available when applied to normal variables.

The article puts forward and substantiates the hypothesis that indices have normal laws of distribution for those time intervals where they are quasi stationary. This manifests itself in the fact that the index's seasonal component is absent and the index trend is (approximately) stationary. The results obtained can be used in preliminary assessments of sample sizes in surveys which, in turn, could reduce social monitoring costs.

Ключевые слова: социологический индекс, индекс социальных настроений, нормальное распределение, теорема Ляпунова, временной ряд

Keywords: sociological index, index of social moods, normal distribution, Lyapunov theorem, time series, Gaussian distribution

Введение

Социологические показатели (*индексы*) вычисляются на основании опросных данных или иных результатов исследований социальных групп. В настоящей статье изучаются индексы, которые линейно выражаются через частоты ответов респондентов [1]. Они являются достаточно распространенными в социологии.

При дополнительном условии квазистационарности разумно предположить, что временные ряды таких индексов имеют нормальные распределения. Вопрос о справедливости этой гипотезы важен в связи с тем, что для нормально распределенных индексов можно найти доверительные интервалы средних значений и среднеквадратических отклонений, а также оценить достаточные размеры выборок. Это позволило бы снизить расходы на социологические опросы. Значительное (многократное) уменьшение размеров выборки, которое благодаря этому становится возможным, помогло бы отобрать вполне случайных респондентов или, во всяком случае, приблизиться к случайности. Последняя необходима для того, чтобы обеспечить репрезентативность и теоретически обосновать полученные статистические данные [Neyman, 1934].

При каких условиях индекс может называться квазистационарным? Допустим, значения $x(t_1), x(t_2), \ldots, x(t_n)$ некоторого социологического индекса X, подсчитанные в последовательные моменты времени $t_1, t_2, \ldots, t_n$, образуют так называемый временной ряд. В нем выделяют тренд $y(t_i)$ (линию регрессии), сезонную и циклическую компоненты $s(t_i)$ и $c(t_i)$, а также ошибку вычислений $e(t_i)$ [Крыштановский, 2000]. Пренебрегая этой ошибкой, предположим, что на промежутке времени $[t_1; t_n]$ тренд $y(t)$ изменяется так медленно, что его можно считать постоянным. В таком случае сезонная компонента $s(t_i)$ отсутствует или незначительна (впрочем, слишком грубый тренд может «срезать» сезонные колебания, пройдя через их средние значения). Соответственно, изменение значений X в основном обусловлено циклической компонентой $c(t_i)$, где $i = 1, \ldots, n$. При этих условиях можно предположить, что индекс X определяет случайную величину с почти неизменным во времени законом распределения. Такие индексы мы будем называть *квазистационарными.* Это свойство проявляется лишь на некотором промежутке времени, в ходе которого не происходит резких социальных изменений. При удлинении этого промежутка постепенное изменение ситуации в обществе (или в релевантной социальной группе) приведет к качественным сдвигам.

Многие социологические показатели линейно выражаются через частоты ответов, полученных от респондентов. Например, индексы социальных оценок от ВЦИОМ [2]. В этой статье теоретически доказано, что такой индекс X должен иметь

---

[1] Для унификации далее мы будем называть изученных респондентами, а метод будем называть опросом, хотя фактически могут быть использованы другие методы.

[2] Индексы социальных оценок // ВЦИОМ. URL: https://wciom.ru/news/ratings/indeksy_soc_nastroenij/ (дата обращения: 25.08.2020).

приблизительно нормальный закон распределения, параметры которого могут изменяться со временем. При этом опросы должны происходить на случайных выборках, что означает равновероятное попадание в опрос любого представителя релевантной социальной группы. На практике такую равновероятность обеспечить сложно, поэтому в социологии активно применяются методы неслучайных выборок [Отчет рабочей группы AAPOR…, 2016]. Особенность таких методов состоит в том, что выбор респондентов обусловлен прежде всего их доступностью. По-видимому, преобладание неслучайных выборок объясняет то обстоятельство, что для социологии нормальные распределения не характерны [Давыдов, 1995: 113]. Ф. Н. Ильясов, основываясь на мнениях А. О. Крыштановского и А. И. Орлова, утверждает, что идея нормального распределения в социологии не обоснована и такие распределения почти не встречаются на практике [Ильясов, 2014: 37].

При этом отмечается, что многие измерения и подходы к анализу данных в социологии строятся на основании гипотезы о нормальном распределении, несмотря на то, что нельзя говорить о достаточности подтверждений данного предположения, что, в свою очередь, ставит под вопрос основанные на ней методы [Ильясов, 2014: 37]. В данной статье теоретически обосновывается свойство нормальности распределений линейных, квазистационарных индексов социологии при условии, что они вычисляются по данным из случайных выборок. Это позволяет применять к ним методы математической статистики без риска методологических ошибок.

Мнение о том, что временной ряд некорректно соотносить с гауссовым распределением [Ильясов, 2014: 37], несомненно верно в случае, когда соответствующий индекс *X* не является квазистационарным. Тогда его последовательные значения нельзя рассматривать как независимые измерения случайной величины (имеющей неизменный закон распределения). Однако очевидно, что на коротком промежутке времени $[t_1; t_n]$, на котором индекс *X* является квазистационарным, числа $x(t_1)$, $x(t_2)$, …, $x(t_n)$ можно считать выборкой значений случайной величины. Последняя может быть (или не быть) нормальной. Настоящая статья посвящена проблеме соотношения нормального распределения с временными рядами и направлена на восполнение пробела в дискуссии на эту тему, а именно математико-статистическому анализу данной проблемы в социологическом контексте.

Линейные социологические индексы

Пусть проводится опрос в социальной группе, при этом членам выборки размера *n* задают по *Q* вопросов, на каждый из которых можно дать один из предлагаемых ответов. Пронумеруем респондентов $i = 1, \ldots, n$, вопросы $q = 1, \ldots, Q$, ответы $r = 1, \ldots, R_q$ (число ответов на вопрос № *q* в общем зависит от *q*). Обозначим S_{qr} долю ответов № *r* на вопрос № *q*, выраженную в % от общего числа полученных ответов, где $r \in \{1, \ldots, R_q\}$. Пусть некоторый индекс *X*, связанный с этим опросом, выражается формулой:

$$X = A + \sum_{q=1}^{Q} \sum_{r=1}^{R_q} K_{qr} S_{qr} . \qquad (1)$$

Постоянные K_{qr} и *A* определяются при разработке показателя. В этой статье рассматриваются индексы *X* вида (1), которые мы будем называть линейными.

Примером служит индекс социальных настроений (ИСН [3]), который ежемесячно публикуется «Левада-Центром». Респондентам из выборки размером $n = 1600$ предлагается по $Q = 12$ вопросов, на каждый из которых нужно дать один из нескольких заданных ответов. ИСН вычисляется как разность между долями положительных и отрицательных ответов, выраженными в % от общего числа ответов, к которой прибавляют 100 для исключения отрицательных значений. Таким образом, если X есть ИСН, то $A = 100$, для положительных ответов $K_{qr} = 1$, для отрицательных $K_{qr} = -1$, для нейтральных $K_{qr} = 0$. Аналогично выражаются индекс потребительских настроений (ИПН), индекс ожидания безработицы (ИБ) от «Левада-Центра», индексы социальных оценок от ВЦИОМ (где $A = 0$).

Чтобы принять во внимание разницу между ответами по важности или степени уверенности респондентов, можно присвоить ответам на вопрос № q уровни значимости $l_{q1}, \ldots, l_{qP_q}$, где P_q — число вариантов положительных ответов, которое равно числу вариантов отрицательных ответов, $q = 1, \ldots, Q$. Уровни значимости определяются интуитивно — методом экспертного оценивания, обычно по шкале баллов от 1 до 5 или от 1 до 10, хотя возможны и другие шкалы. Для такого *дифференцированного*, социального индекса формула (1) приобретает следующий вид:

$$X = A + \sum_{q=1}^{Q} \sum_{s=1}^{P_q} w_{qs} \cdot \left(S_{qs}^{+} - S_{qs}^{-}\right) w_{qs} = \frac{l_{qs} P_q}{l_{q1} + \cdots + l_{qP_q}}, \qquad (2)$$

где S_{qs}^{+} и S_{qs}^{-} есть доли положительных и отрицательных ответов на вопрос № q, имеющих номер s и общий уровень значимости l_{qs}. Как положительные, так и отрицательные ответы пронумерованы от 1 до P_q, нейтральные ответы не принимаются в расчет ($R_q = 2P_q + 1$). При каждом q весовые коэффициенты w_{qs} пропорциональны l_{qs}, где $\sum_{s=1}^{P_q} w_{qs} = P_q$.

Если $l_{q1} = \ldots = l_{qP_q}$ для каждого q, то есть все ответы на один вопрос равнозначны (кроме нейтральных), то $w_{qs} = 1$и формула (2) сводится к следующей:

$$X = A + \sum_{q=1}^{Q} \left(S_q^{+} - S_q^{-}\right) = A + \sum_{q=1}^{Q} S_q^{+} - \sum_{q=1}^{Q} S_q^{-}, \qquad (3)$$

где S_q^{+} и S_q^{-} — доли положительных и отрицательных ответов на вопрос № q, выраженные в % от числа всех ответов: $S_q^{+} = S_{q1}^{+} + \ldots + S_{qP_q}^{+}$ и $S^{-}_{q} = S_{q1}^{-} + \ldots + S_{qP_q}^{-}$. Формула (3) отвечает ИСН, ИПН и ИБ от «Левада-Центра», индексам социальных оценок от ВЦИОМ и другим показателям.

Формула (1) может выражать не только социологические индексы. Например, в работе [Дуганов, Калашников, 2001] описан показатель уровня преждевременной смертности от болезней (ПГПЖ). Здесь $Q = 1$ и параметр R_1 на 1 больше так называемого базового возраста, все смерти раньше достижения которого считаются преждевременными, S_{1r} равно доле случаев смерти в возрасте $r - 1$ лет в % от числа N умерших преждевременно, $A = 0$ и $K_{1r} = N(R_1 - \mathrm{r})/100$.

Другой пример связан с композитным индексом I материального благосостояния [Балацкий, Саакянц, 2006]. Он вычисляется по формуле (1) при $Q = 1$ и $R_1 = 5$,

[3] Обновленная методика измерения Индекса социальных настроений (ИСН) // Левада-Центр. URL: https://www.levada.ru/obnovlennaya-metodika-izmereniya-indeksa-sotsialnykh-nastroenii-isn/ (дата обращения: 03.09.2020).

где S_{1r} равно доле в % респондентов, относящих себя к r-й группе благосостояния (определенной в таблице 1 [там же]), $A=0$, весовые коэффициенты $K_{11}=0$, $K_{12}=0{,}25$, $K_{13}=0{,}5$, $K_{14}=0{,}75$, $K_{15}=1$. Примером социологического показателя, не выражаемого формулой (1), является индекс Джини, характеризующий материальное неравенство [там же].

Рассмотрим произвольный индекс X, определяемый формулой (1). Пусть он вычисляется в последовательные моменты времени $t_1, t_2, \ldots, t_m$. Для определенности будем считать, что респонденты нумеруются в порядке занесения ответов в базу данных. Введем случайную величину X^i_{qr}, которая зависит от ответа респондента № i на вопрос № q. Пусть $X^i_{qr}=1$, если дан ответ № r, и $X^i_{qr}=0$ при любом другом ответе. Тогда из (1) следует, что

$$X = A + \sum_{q=1}^{Q}\sum_{r=1}^{R_q} K_{qr}\cdot\frac{\sum_{i=1}^{n} X^i_{qr}}{nQ}\cdot 100 = A + 100\cdot\frac{\sum_{i=1}^{n} X_i}{n}, \qquad (4)$$

где nQ — общее число полученных ответов и при каждом $i=1, \ldots, n$.

$$X_i = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{r=1}^{R_q} K_{qr}X^i_{qr}.$$

Предполагается, что респонденты выбираются случайным образом, поэтому случайные величины X_i/n можно считать независимыми между собой. Поскольку $|X_i/n| \ll |Y|$, величина $Y=\Sigma_{i=1}^{n}X_i/n$ имеет близкое к нормальному распределение [Гмурман, 2003: 135]. Отсюда можно сделать вывод о (приблизительной) нормальности временного распределения значений индекса X, так как линейная функция $X=100\cdot Y+A$ от нормальной величины Y также является нормальной [там же: 141]. Тогда закон распределения X однозначно определяется параметрами α и $\sigma>0$, имеющими смысл математического ожидания и среднего отклонения, так что функция плотности распределения имеет вид:

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(t-a)^2}{2\sigma^2}\right). \qquad (5)$$

Параметры α и σ могут зависеть от t, так что $\alpha=\alpha(t)$ и $\sigma=\sigma(t)$. Но на любом промежутке времени, где индекс X является квазистационарным, α и σ можно считать постоянными.

Свойство нормальности статистического показателя X, который линейно выражается через взятые в большом числе независимые, случайные величины, было установлено Р. Фишером еще в начале XX в. [Neyman, 1934: 564]. Чтобы обосновать его теоретически, воспользуемся теоремой Ляпунова.

Каждая величина X_i представляет собой результат измерения, то есть определения в результате опроса респондента № i значения случайной величины.

$$\chi = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{r=1}^{R_q} K_{qr}x_{qr}, \qquad (6)$$

где $x_{qr}=1$, если на вопрос № q был дан ответ № r, а при других ответах $x_{qr}=0$.

Ограничимся промежутком времени, на котором индекс X является квазистационарным. Тогда X — случайная величина, не зависит от времени (приблизительно). Предположим, что респонденты нумеруются индексом i, который изменяется от 1 до бесконечности. Ни одна социальная группа не состоит из бесконечного числа членов. Но если она достаточно велика, то такая идеализация не приведет к ошибке. Рассмотрим последовательность величин $Y_n = \Sigma_{i=1}^{n} X_i / n$, где $n = 1, 2, \dots$ и X_i есть результат измерения показателя χ для i-го респондента. Каждая X_i есть случайная величина, поскольку респонденты выбираются случайным образом. Согласно следствию из теоремы Ляпунова, при $n \to \infty$ закон распределения величины Y_n неограниченно приближается к нормальному [Ляпунов, 1948: 245].

Условия теоремы Ляпунова также требуют, чтобы для некоторого $\delta > 0$ математическое ожидание величины $|\chi - M(\chi)|^{2+\delta}$ было конечным [там же: 245]. В данном случае это верно для всех $\delta > 0$, так как из (6) следует, что полученное при любом измерении значение x величины χ удовлетворяет неравенству $|x| < \Sigma_{q=1}^{Q} \Sigma_{r=1}^{R_q} |K_{qr}| / Q$. Следовательно, величина $|\chi - M(\chi)|^{2+\delta}$ является ограниченной при любом $\delta > 0$. Поэтому ее математическое ожидание конечно.

Таким образом, условия теоремы Ляпунова выполняются. Следовательно, любой социологический индекс X, определяемый формулой (1), должен иметь приблизительно нормальное распределение на каждом промежутке времени, где его можно считать квазистационарным.

Последствия нормальности

Условие квазистационарности означает, что значения индекса в различные моменты времени определяют случайную величину X, закон распределения которой (почти) не зависит от времени. О выполнении этого условия можно судить по тому, что индекс X не имеет существенной, сезонной компоненты, а его тренд является (приблизительно) константой.

Еще одно важное условие, связанное с квазистационарностью, состоит в том, что каждый опрос проводится среди респондентов, выбираемых случайно в соответствующей социальной группе. Если выборки осуществляются случайно, но условие квазистационарности при $\tau \leqslant t \leqslant T$ не выполняется, то нужно разделить промежуток $[\tau; T]$ на N частей $[\tau_j; \tau_{j+1}]$, где $\tau_0 = \tau$ и $\tau_N = T$, так что для каждого $j = 0, \dots, N-1$ индекс X можно считать квазистационарным при $\tau_j \leqslant t \leqslant \tau_{j+1}$. Тогда на каждом промежутке времени $[\tau_j; \tau_{j+1}]$ случайная величина X будет иметь (почти) неизменные параметры α и σ нормального распределения (5), где $\alpha = \alpha(\tau_j) \approx \alpha(\tau_{j+1})$ и $\sigma = \sigma(\tau_j) \approx \sigma(\tau_{j+1})$.

Пусть некоторый социологический индекс X является квазистационарным на промежутке $[t_1; t_n]$ и в последовательные моменты времени t_j были вычислены его значения x_j, так что $x_j = x(t_j)$ при $j = 1, 2, \dots, n$. Если известно, что случайная величина X имеет (приблизительно) нормальное распределение с функцией плотности (5) при $t_1 \leqslant t \leqslant t_n$, то параметры α и σ определяются по приближенным формулам:

$$a \approx \bar{x}_{\text{в}} = \frac{\sum_{j=1}^{n} x_j}{n} \quad \sigma \approx s = \sqrt{\frac{\sum_{j=1}^{n}(x_j - \bar{x}_{\text{в}})^2}{n-1}}. \qquad (7)$$

При заданном $0 < p < 1$ можно оценить погрешности приближенных формул (7) с вероятностью ошибки не больше $1 - p$ (уровень значимости). Соответственно, надежность таких оценок будет равна $100p$ %. Они основаны на следующих фактах [Гмурман, 2003].

1. С вероятностью p имеет место:

$$\bar{x}_{\text{в}} - t_{n,p} \cdot \frac{s}{\sqrt{n}} < a < \bar{x}_{\text{в}} + t_{n,p} \cdot \frac{s}{\sqrt{n}}, \tag{8}$$

где $t_{n,p}$ обозначает $(p+1)/2$ — квантиль распределения Стьюдента с $n-1$ степенями свободы. В пакете EXCEL квантиль $t_{n,p}$ вычисляется как значение функции СТЬЮДЕНТ.ОБР.2Х с аргументами $1-p$ и $n-1$. Проверка правильности вычислений: при $n = 16$ и $p = 0{,}95$ должно получаться $t_{n,p} \approx 2{,}13$ [там же: 217].

2. С вероятностью p имеет место:

$$s \cdot \left(1 - q_{n,p}\right) < \sigma < s \cdot \left(1 + q_{n,p}\right), \tag{9}$$

где $q_{n,p}$ есть относительная погрешность приближения $\sigma \approx s$, выраженная в долях единицы. Величина $q_{n,p}$ определяется из уравнения:

$$F_{\chi^2, n-1}\left(\frac{n-1}{(1-q_{n,p})^2}\right) - F_{\chi^2, n-1}\left(\frac{n-1}{\left(1+q_{n,p}\right)^2}\right) = p, \tag{10}$$

где $F_{\chi^2, n-1}(x)$ есть функция распределения χ^2 с $n-1$ степенями свободы [там же: 221]. Уравнение (10) можно решить с помощью процедуры «Поиск решения» в меню «Данные» пакета EXCEL. При этом значения $F_{\chi^2, n-1}(x)$ вычисляются функцией ХИ2.РАСП со значением ИСТИНА аргумента «интегральная». Проверка правильности вычислений: при $n = 25$ и $p = 0{,}95$ должно получаться $q_{n,p} \approx 0{,}32$ [там же: 222].

Таким образом, параметры α и σ можно оценить неравенствами (8) и (9) с надежностью $100p$ %. Следует иметь в виду, что при повышении надежности *доверительные интервалы* (8) и (9) растягиваются, то есть растут погрешности оценок. Если подобрать число n таким образом, чтобы неравенства (8) и (9) выполнялись с достаточной надежностью, то размеры доверительных интервалов (8) и (9) определят погрешности в определенииα и σ.

В пакете SPSS оценки (7) выдаются автоматически, например, в окне «Анализ → Описательные статистики → Описательные». При этом доверительные интервалы (8) и (9) не определяются, но для оценки параметра α (среднего значения X) в SPSS вычисляется его стандартная ошибка.

Полученный в этой работе результат можно использовать для оценивания размеров выборок, достаточных для измерений линейных, квазистационарных, социологических индексов. Такой теоретический размер может оказаться многократно меньше применяемых на практике. Помимо экономии финансовых ресурсов, это даст возможность проводить более тщательные и подготовленные опросы, что обеспечит стохастичность выборок. В классической работе [Neyman, 1934: 561] отмечалось, что выборки респондентов, не являющиеся случайными, не позволяют обосновать социологические измерения также надежно, как и стохастические выборки.

Практически следует получить временной ряд значений индекса по случайным выборкам некоторого начального размера, например, 100. Число значений временного ряда может быть равно, например, 30. Таким образом, все исследование при ежедневном опросе займет чуть больше месяца. Затем следует проверить закон распределения индекса во времени, и если он окажется недостаточно нормальным, то размер выборки нужно будет увеличить, например до 500. Далее процесс повторяется, и если тесты на нормальность пройдены успешно, то выборку можно уменьшить и т. д. Разумеется, эти соображения нуждаются в практической проверке, главная сложность которой будет состоять в том, чтобы обеспечить случайный отбор респондентов. Примеры проверки временных рядов на нормальность рассмотрены в следующем параграфе.

Индекс социальных настроений

Используя данные, опубликованные на сайте «Левада-Центра»[4], проверим гипотезу о нормальности ИСН на промежутке времени с марта 2009 по август 2019 г. (см. рис. 1). Абсолютные значения на сайте не указаны, но даны индексы семьи, России, ожиданий и власти за каждые два месяца. Значения ИСН вычисляются как среднеарифметические этих четырех индексов. Получается ряд всех чисел от 101 до 128, за исключением 103, 121 и 123 По этим данным построена диаграмма (см. рис. 1). Сезонная компонента на рисунке 1 не наблюдается (локальные максимумы и минимумы приходятся как на летние, так и на зимние месяцы).

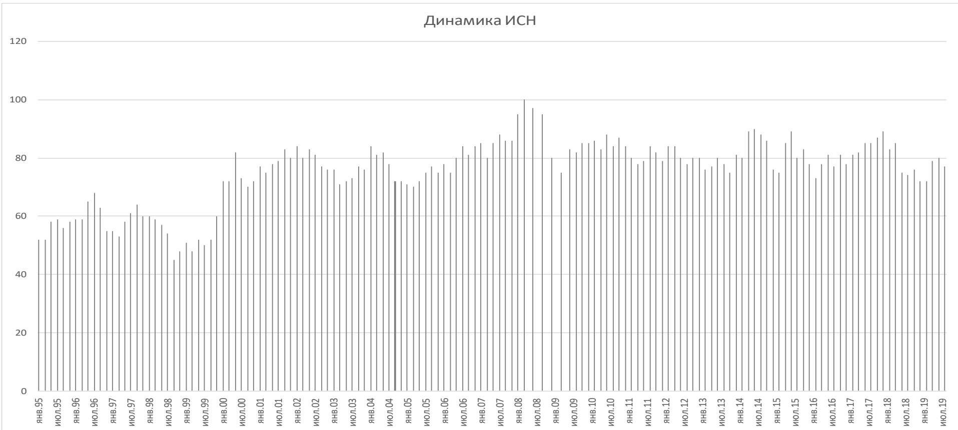

Рис. 1. Значения ИСН с января 1995 по август 2019 г., в % от показателя марта 2008 5

Можно предположить, что с марта 2009 по август 2019 г. ИСН был квазистационарным. Это соответствует официальной точке зрения о том, что за первое

[4] Социально-экономические индикаторы // Левада-Центр. URL: https://www.levada.ru/indikatory/sotsialno-ekonomicheskie-indikatory/ (дата обращения: 25.08.2020).

[5] Это максимальное значение ИСН за период измерений.

десятилетие XXI века Россия достигла стабильного социально-экономического состояния. Является ли случайная величина X, связанная с ИСН на данном промежутке времени, хотя бы приблизительно нормальной?

Пусть $x_j = 100 + j$ и n_j — частота значения x_j, где $j = 1, \ldots, 28$. Число измерений $n = \Sigma_{j=1}^{28} n_j = 62$. Найдем эмпирическую плотность распределения $f_E(x)$:

$$F_{\chi^2,n-1}\left(\frac{n-1}{(1-q_{n,p})^2}\right) - F_{\chi^2,n-1}\left(\frac{n-1}{\left(1+q_{n,p}\right)^2}\right) = p\,, \qquad (11)$$

где $f_E(x_j)\cdot(x_{j+1} - x_j)$ приближенно равно вероятности, с которой индекс X принимает значение в полуинтервале $[x_j; x_{j+1})$. Полагаем $x_{29} = 129$ и получаем эмпирические оценки $\bar{x}_в = 113{,}65$ и $\sigma_в = 6{,}14$. Принимаем эти числа как параметры α и σ теоретической плотности $f(x)$ распределения величины X (5). Вычислим $f(x)$ в Excel с помощью функции НОРМ.РАСП со значением ЛОЖЬ аргумента «Интегральная».

Для полученных таким образом теоретической $f(x)$ и эмпирической $f_E(x)$ плотностей распределения ИСН построены гистограммы значений в точках x_j (см. рис. 2). Ряду 2 отвечает теоретическая гистограмма (светлые столбики), ряду 3 — эмпирическая (темные столбики). Для проверки согласия между теоретическим и эмпирическим распределениями следует использовать критерий Пирсона [Гмурман, 2003: 329], однако и без вычислений видно, что они не согласованы. Таким образом, выборка значений ИСН с марта 2009 по август 2019 г. не имеет никаких видимых признаков нормального распределения.

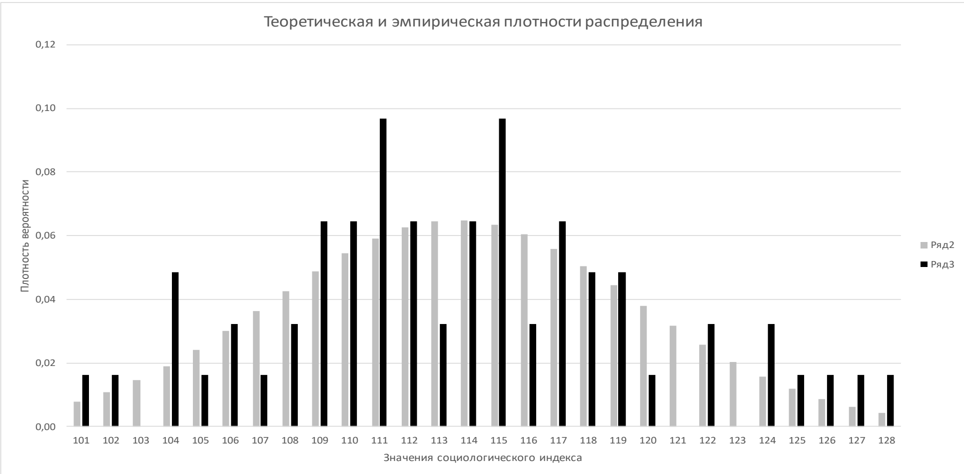

Рис. 2. Теоретическая и эмпирическая плотности распределения ИСН

Отсутствие признаков нормальности на рисунке 2, по-видимому, связано прежде всего с тем, что при вычислении ИСН используются выборки респондентов, которые не являются случайными (квотные выборки и т. п.) [Отчет рабочей группы AAPOR…, 2016]. Квазистационарность в период с марта 2009 по август 2019 г.

является спорным допущением, так как квадратичный тренд данного временного ряда, построенный в пакете STATISTICA, снижается на 22 % от размаха колебаний ИСН (см. рис. 3).

Рассмотрим промежуток времени между отметками 0 и 40, на котором тренд снижается лишь на 6 % от размаха колебаний индекса. Можно предположить, что на этом интервале ИСН является квазистационарным.

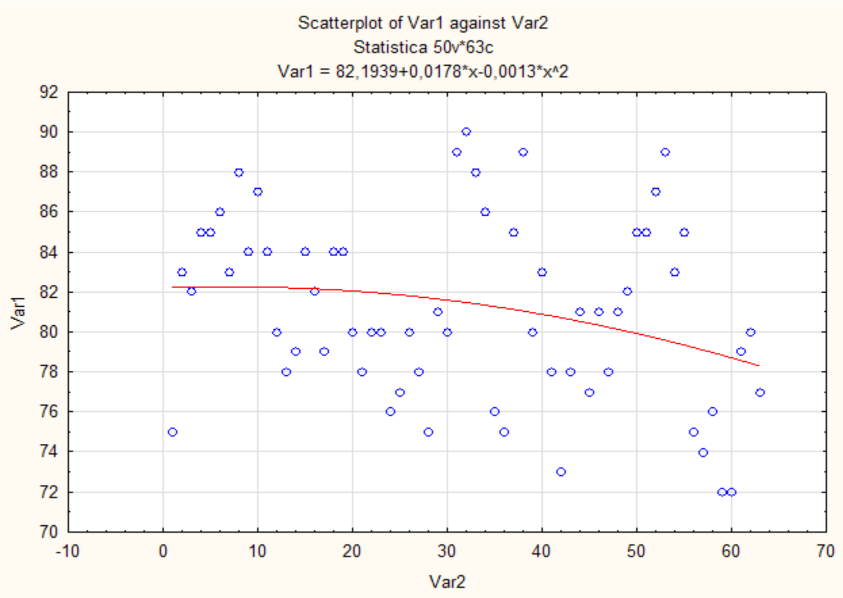

Рис. 3. Квадратичный тренд временного ряда ИСН с марта 2009 по август 2019 г.

Проверим нормальность распределения ИСН на этом интервале времени, используя тест Шапиро-Уилка пакета STATISTICA. Получим так называемую критическую вероятность $p = 0{,}268$, которая удовлетворяет неравенству $p > 0{,}05$ и, следовательно, у нас есть *формальное* основание принять нулевую гипотезу о нормальности. Однако этот тест ничего определенного не выражает. Его результат означает следующее: если бы ИСН имел нормальное распределение, то найденное значение статистики Шапиро-Уилка могло быть получено с вероятностью $p = 0{,}268$. Таким образом, в результате теста наступило событие, которое могло произойти с небольшой, но и не слишком малой вероятностью. Для иллюстрации предположим, что в результате теста было получено значение $p = 0{,}04 < 0{,}05$. В случае нормальности ИСН оно имело бы (малую) вероятность 4 %, но, тем не менее, было получено. Отсюда можно сделать вывод, что ИСН не является нормальным. В нашем случае значение критерия $p = 0{,}268$. Превышение уровня значимости 0,05, который принят в социологии, фактически означает лишь отсутствие весомых оснований, чтобы отвергнуть гипотезу о нормальности. Считается, что в этом случае ее можно принять. Но одного результата этого теста недостаточно, чтобы сделать вывод о нормальном распределении ИСН. Действительно, проверка распределения на равномерность (Rectangular) с помощью теста χ^2 дает критическую вероятность $p = 0{,}294$. Поэтому временное распределение ИСН можно было бы с таким же основанием считать равномерным.

Таким образом, тесты пакета STATISTICA не позволяют сделать уверенное заключение о нормальности ИСН на данном промежутке времени, хотя и не опровергают ее. По-видимому, свойство нормальности может отсутствовать из-за не вполне случайного характера выборок при социологических опросах. Другим источником «анормальности» закона временного распределения может быть недостаточно обоснованное предположение о квазистационарном характере индекса.

Заключение

В статье теоретически обоснована гипотеза о том, что социологические индексы, вычисляемые по формуле (1), должны иметь нормальные законы распределения на тех промежутках времени, где эти индексы являются квазистационарными. При этом каждый отбор респондентов должен быть стохастическим, т. е., все члены данной социальной группы должны иметь равные вероятности попасть в выборку. Важность последнего условия характеризуется тем фактом, что после 1948 г. случайная выборка стала для США предпочтительным методом по сравнению с квотными выборками, opt-in панелями и т. д. [Отчет рабочей группы AAPOR…, 2016: 22]. Отмечается, что при использовании неслучайных источников для выборки следует избегать претензии на репрезентативность [Отчет рабочей группы AAPOR…, 2016: 24].

Полученный результат может оказаться полезным для оценок размеров выборок при опросах, что помогло бы снизить расходы на социальный мониторинг. Для проверки этого предположения необходимы практические исследования, которые в случае успеха помогут выработать метод определения достаточных размеров выборки для измерения линейного, квазистационарного, социологического индекса, а также алгоритм случайного опроса.

Список литературы

Отчет рабочей группы AAPOR о неслучайных выборках: июнь 2013 / Американская ассоциация исследователей общественного мнения (AAPOR); пер. с англ. Д. М. Рогозина, А. А. Ипатовой. М.: Фонд «Общественное мнение», 2016.
AAPOR (2016) Report of the American Association for Public Opinion Research (AAPOR) Task Force on Non-Probability Sampling. June 2013. Moscow: Public Opinion Foundation (FOM). (In Russ.)

Балацкий Е. В., Саакянц К. М. Индексы социального неравенства // Мониторинг общественного мнения: экономические и социальные перемены. 2006. № 2. С. 122—128.
Balatskiy E. V., Saakyants K. M. (2006) Indices of Social Inequality. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 122—128. (In Russ.)

Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов. М.: НИУ ВШЭ, 2003.

Gmurman V. E. (2003) Probability Theory and Mathematical Statistics. Moscow: National Research University Higher School of Economics. (In Russ.)

Давыдов А. А. Анализ одномерных частотных распределений в социологии: эволюция подходов // Социологические исследования. 1995. № 5. С. 113—116.
Davydov A. A. (1995) Analysis of One-Dimensional Frequency Distributions in Sociology: the Evolution of Approaches. *Sociological Studies*. No. 5. P. 113—116. (In Russ.)

Дуганов М. Д., Калашников К. Н. Методологические подходы к оценке эффективности регионального здравоохранения // Экономические и социальные перемены: факты, тенденции, прогноз. 2011. № 6. С. 93—105.
Duganov M. D., Kalashnikov K. N. (2011) Methodological Approaches to Assessing the Effectiveness of Regional Health Care. *Economic and Social Changes: Facts, Trends, Forecast*. No. 6. P. 93—105. (In Russ.)

Ильясов Ф. Н. Типы шкал и распределений в социологии // Мониторинг общественного мнения: экономические и социальные перемены. 2014. № 4. С. 24—40. https://doi.org/10.14515/monitoring.2014.4.03.
Iliassov F. N. (2014) Scales and Specific Sociological Measurement. *Monitoring of Public Opinion: Economic and Social Changes*. No. 4. P. 24—40. https://doi.org/10.14515/monitoring.2014.4.03. (In Russ.)

Крыштановский А. О. Методы анализа временных рядов // Мониторинг общественного мнения: экономические и социальные перемены. 2000. № 2. С. 44—51.
Kryshtanovsky A. O. (2000) Methods of Analysis of Time Series. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 44—51. (In Russ.)

Ляпунов А. М. Избранные труды. М.: Издательство АН СССР, 1948.
Lyapunov A. M. (1948) Selected Works. Moscow: Publishing House of the Academy of Sciences of the USSR. (In Russ.)

Neyman J. (1934) On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*. Vol. 97. No. 4. P. 558—625. https://doi.org/10.2307/2342192.