

ИДЕНТИФИКАЦИЯ ТИПА РАСПРЕДЕЛЕНИЯ НА ПРИМЕРЕ НЕКОТОРЫХ ПОКАЗАТЕЛЕЙ ЦЕННОСТНОГО МИРА РОССИЙСКОЙ МОЛОДЕЖИ

Роман Александрович Жуков¹

¹ Доктор экономических наук, кандидат физико-математических наук, доцент, доцент кафедры математики и информатики, Финансовый университет при Правительстве Российской Федерации (Тульский филиал), Тула, Россия, e-mail: pluszh@mail.ru

Светлана Васильевна Прокопчина²

² Доктор технических наук, профессор, профессор кафедры "Системный анализ в экономике", Финансовый университет при Правительстве Российской Федерации, Москва, Россия, e-mail: svprokopchina@mail.ru

ИНФОРМАЦИЯ

Ключевые слова:

байесовские интеллектуальные технологии
лингвистические измерения
факторы
ценностный мир молодежи
модель
плотность распределения вероятностей
программный модуль

АННОТАЦИЯ

Целью исследования является определение возможности идентификации типа распределения по информации анкетных опросов. Методологическая сложность такой задачи состоит в том, что данные являются, по сути, уникальными экспериментами, для обработки которых невозможно использовать методы математической статистики. Идентификация типа распределения осуществляется на основании байесовских интеллектуальных технологий с использованием системы Пирсона, по методике, разработанной одним из авторов в 70-х годах прошлого столетия. Методика успешно применялась для идентификации законов распределений в условиях малых выборок и значительной неопределенности. Но для анкетной информации она применена впервые. Это дает возможность по полученному закону распределения генерировать выборки любого объема и использовать их для обработки мощными методами математической статистики. В статье приведен пример ее успешного применения для определения законов распределения показателей модели ценностного мира российской молодежи по информации, полученной на основе результатов анкетирования молодых лиц в возрасте от 17 до 30 лет в 10 субъектах Российской Федерации.

ДЛЯ ЦИТИРОВАНИЯ: Жуков Р.А., Прокопчина С.В. Идентификация типа распределения на примере некоторых показателей ценностного мира российской молодежи // Мягкие измерения и вычисления. 2023. № 7. Т. 68. С. 42–50; <https://doi.org/10.36871/2618-9976.2023.07.004>.

IDENTIFICATION OF THE TYPE OF DISTRIBUTION BY THE EXAMPLE OF SOME INDICATORS OF THE VALUE SYSTEM OF RUSSIAN YOUTH

Roman A. Zhukov¹

¹ Doctor of Economic Sciences, Candidate of Physical and Mathematical Sciences, Docent, Associate Professor, Dept. of Mathematics and Informatics, Financial University under the Government of the Russian Federation (Tula Branch), Tula, Russia, e-mail: pluszh@mail.ru

Svetlana V. Prokopchina²

² Doctor of Economic Sciences, Professor, Professor of the Department of "System Analysis in Economics", Financial University under the Government of the Russian Federation, Moscow, Russia, e-mail: svprokopchina@mail.ru

ARTICLE INFO

Keywords:

Bayesian intelligent technologies
Linguistic measurements
Factors
Value system of youth
Model
Probability distribution density
Software module

ABSTRACT

The purpose of the study is to determine the possibility of identifying the type of distribution for the information of questionnaire surveys. The methodological complexity of such a task lies in the fact that the data are essentially unique experiments, for the processing of which it is impossible to use the methods of mathematical statistics. Identification of the distribution type is carried out on the basis of Bayesian intelligent technologies using the Pearson system, according to the methodology developed by one of the authors in the 70s of the last century. The technique has been successfully applied to identification of law distribution in the situations with small samples and under the conditions of considerable uncertainty of information. But for the task of questionnaire data processing this approach is used for the first time. This makes it possible to generate samples of any size according to the obtained distribution law and use them for processing by powerful methods of mathematical statistics. The article provides an example of its successful application to determine the laws of the distribution of indicators of the model of the value world of Russian youth, according to information obtained on the basis of the results of a survey of young people aged 17 to 30 years in 10 subjects of the Russian Federation.

FOR CITATION: Zhukov R.A., Prokopchina S.V. (2023) Identification of the type of distribution by the example of some indicators of the value system of Russian youth. *Soft measurements and computing*, vol. 68, no. 7, pp. 42–50 (In Russ.); <https://doi.org/10.36871/2618-9976.2023.07.004>.

Введение

При изучении сложных систем и процессов, характеризующихся неполнотой и неточностью числовой или лингвистической информации, возникает проблема их адекватного описания, причем в большинстве своем измерения являются единичными и зависят от конкретных условий и периодов времени получения данных, что может привести к некорректности сопоставления результатов для одного и того же объекта или явления. Это характерно и для социально-экономических систем (далее – СЭС) и процессов, информация о которых поступает зачастую в искаженном виде не только из официальных источников, но и из результатов проведенных социологических опросов и анкетирования. Два последних вида статистического наблюдения являются яркими тому примерами. Полную информацию о процессе может дать функция или плотность распределения вероятностей, однако в условиях ограниченности данных ее восстановление представляется трудной задачей.

Применение байесовских интеллектуальных технологий (далее – БИТ) дает возможность работать с неполной и нечеткой информацией, получать статистические оценки процессов в условиях малых выборок, характерных для СЭС, в том числе при изучении такой системы как ценностный мир молодежи (далее – ЦММ). Молодежь – молодые люди в возрасте от 17 до 30 лет – является будущим базисом устойчивого развития государства и его административно-территориальных единиц, а формирование ориентиров (ценностей) этой группы населения является одной из приоритетных задач органов управления всех уровней. Достоверное знание о характере ЦММ может быть получено, если будут известны статистические характеристики объекта исследования и функция распределения, восстановленная посредством соответствующей методологии.

Методология и информационная база исследования

В качестве базовой методологии обработки результатов анкетирования используются БИТ [1], которые хорошо зарекомендовали себя в ряде прикладных исследований [2–6].

Для определения вида закона распределения показателей по анкетной информации используется подход и метод на основе байесовских интеллектуальных измерений (БИИ) коэффициентов асимметрии и эксцесса на двумерной функциональной шкале БИИ, построенной на плоскости моментов Пирсона. Метод обоснован и приведен в [1].

Согласно методологии БИИ, каждый объект или процесс может быть представлен в виде модели объекта (процесса) в виде иерархической структуры подсистем (под-процессов) и элементов (элементарных работ), которые могут характеризоваться интегральными (обобщенными) и частными факторами, представленными на числовой и лингвистической шкалах в виде нечеткого числа.

Для получения вероятности значений факторов используется байесовская свертка вида [7]:

$$P^{(ap)}(h_{k,t} | \{Mx_{i,t}\} | Y_{k,t}) = \frac{P_{k,t}^a(h_{k,t} | \{Mx_{i,t}\} | Y_{i-1,t})P(\tilde{h}_{k,t} | \{Mx_{i,t}\} | Y_{i,t})}{\sum_{j=1}^K P_{k,t}^a(h_{j,t} | \{Mx_{i,t}\} | Y_{i-1,t})P(\tilde{h}_{j,t} | \{Mx_{i,t}\} | Y_{i-1,t})}, \quad (1)$$

где $h_{k,t}$ – список результатов или значимых альтернативных решений из множества решений $H_{k,t}$ (в данном случае $q^{(j)}(t)$); $\tilde{h}_{k,t}$ – их оценка; $x_{i,t}$ – набор данных из множества $X_{i,t}$ или набор числовых или лингвистических переменных, отражающих свойства атрибутов; $Y_{i,t}$ – условия реализации измерения, включающие в себя множество метрологических требований $\{MX\}_{i,t}$, множество априорной информации A и ограничений и допущений O ; a – априорное значение; ap – апостериорное значение; t – время.

Для свертки двух показателей и формирования интегрального фактора (обозначение (in)) используется формула:

$$P_{k,i,t}^{(in)}(h_{k,t} | h_{i,t}) = \frac{P^{(ap)}(h_{k,t} | \{Mx_{k,t}\} | Y_{k-1,t})P^{(ap)}(h_{i,t} | \{Mx_{i,t}\} | Y_{i-1,t})}{\sum_{k=1}^K \sum_{i=1}^I P^{(ap)}(h_{k,t} | \{Mx_{k,t}\} | Y_{k-1,t})P^{(ap)}(h_{i,t} | \{Mx_{i,t}\} | Y_{i-1,t})}, \quad (2)$$

где K, I – соответствующие значимые альтернативные значения для первого и второго факторов.

В случае наличия нескольких индикаторов, подлежащих свертке, процедура (2) применяется для полученного и следующего факторов, повторяясь до тех пор, пока номер итерации не достигнет номера последнего фактора в соответствующем поддереве факторов, характеризующих модель объекта.

Такая процедура реализована в программном комплексе "Инфоаналитик" [8].

Каждый из полученных факторов отображается на числовой (значение репера (количество реперов задается пользователем), вероятность) и лингвистической (один из 9 классов (от предельно ниже нормы до предельно выше нормы), вероятность) шкалах с левой и правой границами. Частные факторы могут представляться на двух шкалах, интегральный фактор – только на лингвистической шкале.

Следующим этапом является генерация промежуточной выборки на заданных интервалах, например 9 интервалах числовой или лингвистической шкал (для частных

и интегральных факторов соответственно), случайных величин требуемого объема. Для этого предложено использовать распределение Симпсона. Полученная выборка служит основой для идентификации типа распределения по рассчитываемым коэффициентам асимметрии и эксцесса [9]. На основе методологии определения разделяющих границ на плоскости моментов Пирсона, определяется наиболее вероятные типы распределения и вычисляются их характеристики.

После идентификации типа распределения можно генерировать выборку произвольного объема, к которой обоснованно можно применять эконометрические методы и строить соответствующие модели.

Представленная методология была применена для оценки и анализа ценностного мира молодежи на базе построенной модели [10].

Информационной базой исследования стали данные, полученные на основе результатов анкетирования молодых лиц в возрасте от 17 до 30 лет из 10 субъектов Российской Федерации.

Результаты и обсуждение

В качестве факторов представлены индикаторы, характеризующие отношение к получению знаний (1 интегральный 5 частных факторов) молодых лиц – мужчин – в возрасте от 17 до 20 лет, учащихся вузов Тульской области (объем выборки 7). Объем сгенерированной промежуточной выборки составил 7000.

Фрагмент модели ЦММ представлен на [рисунке 1](#).

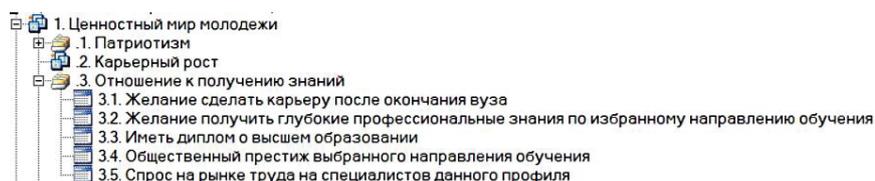


Рис. 1. Фрагмент модели ЦММ

Источник: составлено авторами.

Результаты анкетирования и соответствующие им типы распределений, характеристики выборок представлены на [рисунках 2–9](#).

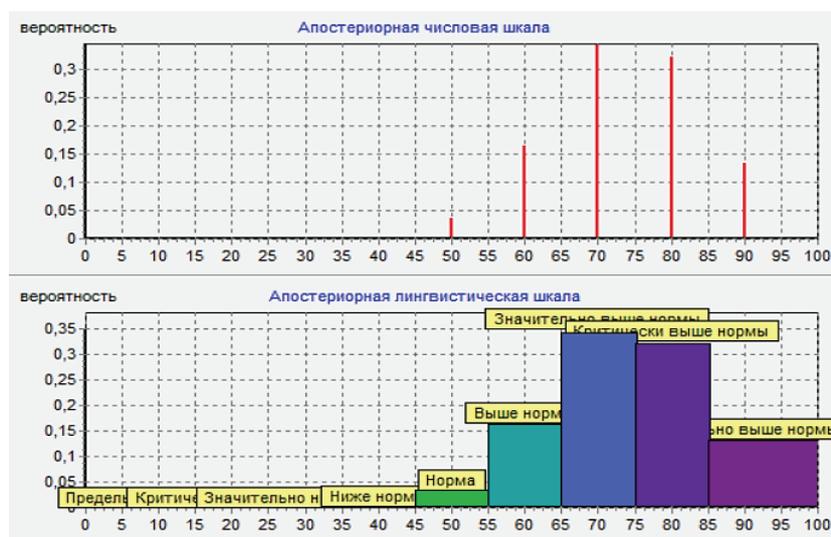


Рис. 2. Фактор 3.1. Желание сделать карьеру после окончания вуза

Источник: составлено авторами.

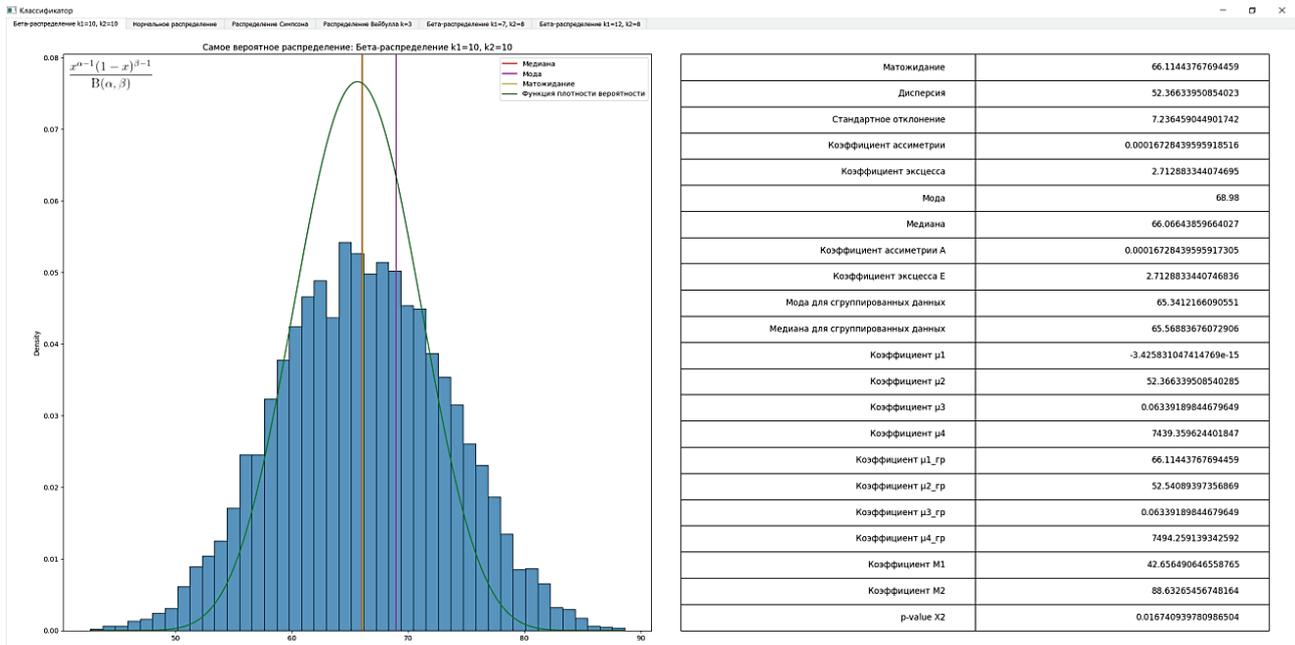


Рис. 3. Наиболее вероятное распределение и числовые характеристики.
Фактор 3.1. Желание сделать карьеру после окончания вуза
Источник: составлено авторами.

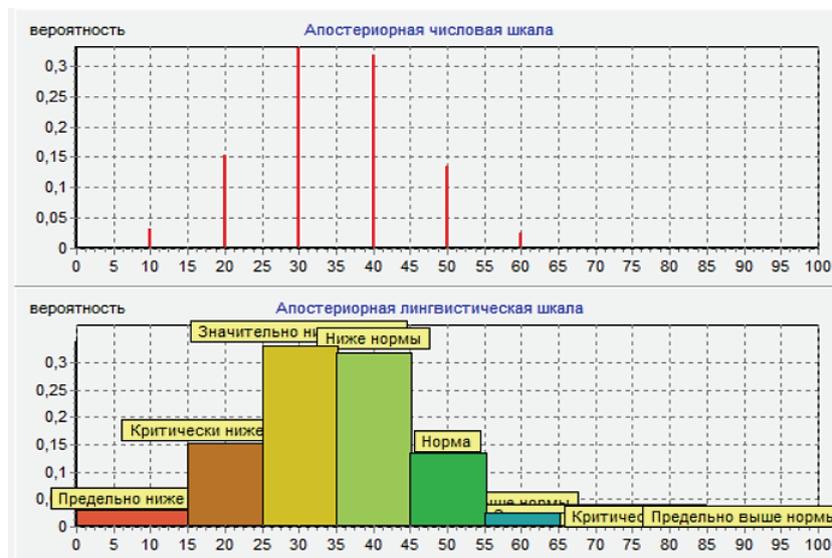


Рис. 4. Фактор 3.2. Желание получить глубокие профессиональные знания по избранному направлению обучения
Источник: составлено авторами.

Самое вероятное распределение: Нормальное распределение
Также вероятное распределение: Бета-распределение k1=10, k2=10
Маловероятное распределение: Распределение Симпсона
Маловероятное распределение: Распределение Вейбулла k=3
Маловероятное распределение: Бета-распределение k1=7, k2=8
Маловероятное распределение: Бета-распределение k1=12, k2=8

Рис. 5. Результат идентификации типа распределения. Фактор 3.2. Желание получить глубокие профессиональные знания по избранному направлению обучения
Источник: составлено авторами.

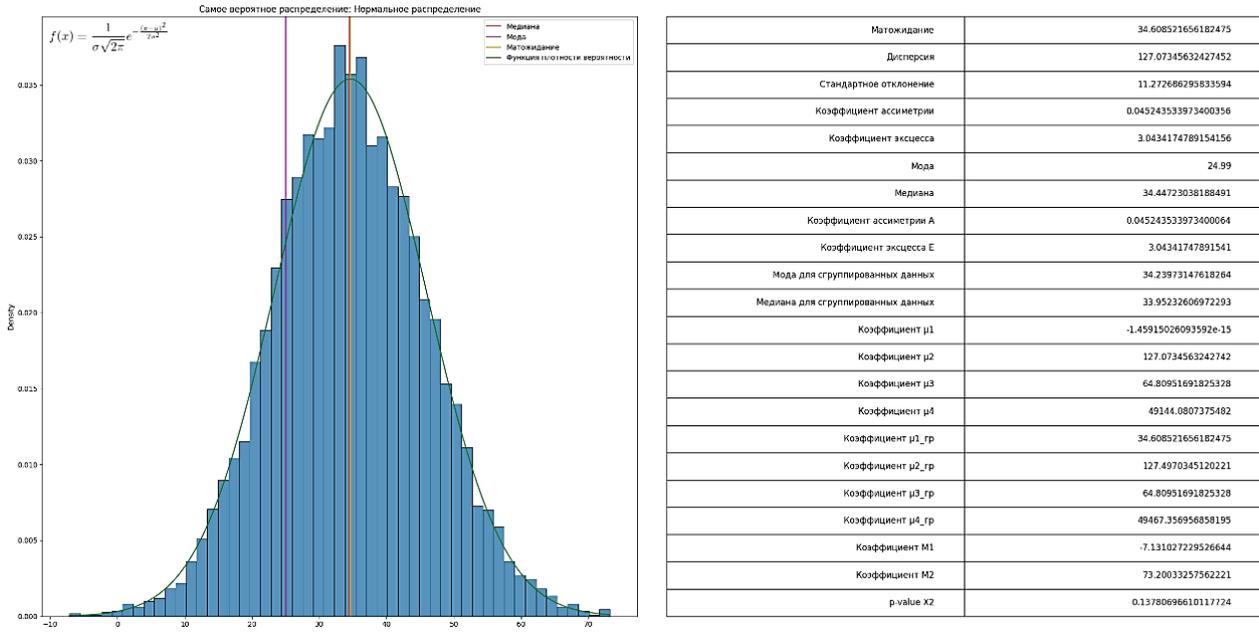


Рис. 6. Наиболее вероятное распределение и числовые характеристики. Фактор 3.2. Желание получить глубокие профессиональные знания по избранному направлению обучения
 Источник: составлено авторами.

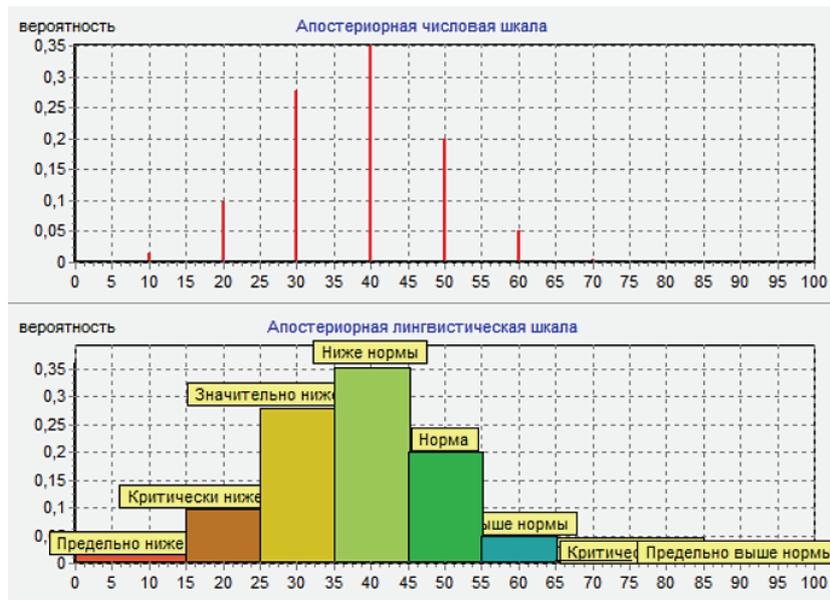


Рис. 7. Фактор 3.3. Иметь диплом о высшем образовании
 Источник: составлено авторами.

Самое вероятное распределение: Нормальное распределение
 Также вероятное распределение: Бета-распределение k1=10, k2=10
 Маловероятное распределение: Распределение Симпсона
 Маловероятное распределение: Распределение Вейбулла k=3
 Маловероятное распределение: Бета-распределение k1=7, k2=8
 Маловероятное распределение: Бета-распределение k1=12, k2=8

Рис. 8. Результат идентификации типа распределения.
 3.3. Иметь диплом о высшем образовании
 Источник: составлено авторами.

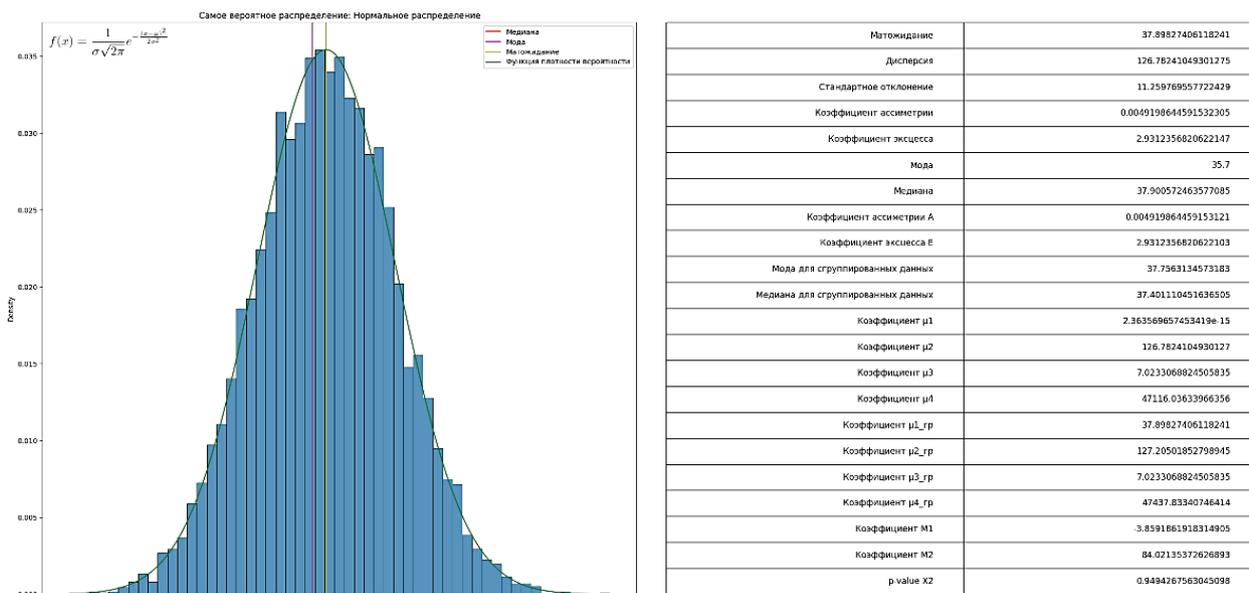


Рис. 9. Наиболее вероятное распределение и числовые характеристики.

3.3. Иметь диплом о высшем образовании

Источник: составлено авторами.

Аналогичные характеристики были получены для факторов 3.4, 3.5 и 3. Результаты идентификации типа распределения представлены на рисунках 10–12.

Самое вероятное распределение: Бета-распределение k1=6, k2=10
 Также вероятное распределение: Распределение Вейбулла k=3
 Маловероятное распределение: Бета-распределение k1=2, k2=4
 Маловероятное распределение: Бета-распределение k1=10, k2=10

Рис. 10. Результат идентификации типа распределения.

3.4. Общественный престиж выбранного направления обучения

Источник: составлено авторами.

Самое вероятное распределение: Нормальное распределение
 Также вероятное распределение: Бета-распределение k1=10, k2=10
 Маловероятное распределение: Распределение Симпсона
 Маловероятное распределение: Распределение Вейбулла k=3
 Маловероятное распределение: Бета-распределение k1=7, k2=8
 Маловероятное распределение: Бета-распределение k1=12, k2=8

Рис. 11. Результат идентификации типа распределения.

3.5. Спрос на рынке труда на специалистов данного профиля

Источник: составлено авторами.

Для интегрального фактора 3. Отношение к получению знаний были идентифицированы типы распределения для промежуточной выборки N=2000.

Самое вероятное распределение: Нормальное распределение
 Также вероятное распределение: Бета-распределение k1=10, k2=10
 Маловероятное распределение: Распределение Симпсона
 Маловероятное распределение: Распределение Вейбулла k=3
 Маловероятное распределение: Бета-распределение k1=7, k2=8
 Маловероятное распределение: Бета-распределение k1=12, k2=8

Рис. 12. Результат идентификации типа распределения.

3. Отношение к получению знаний (интегральный)

Источник: составлено авторами.

Из полученных результатов видно, что факторы, в большинстве своем соответствуют нормальному закону распределения. Это говорит о том, что основная часть молодежного сообщества придает этим ценностям примерно одинаковое значение. Однако, значения других ценностей могут быть и полярно противоположными как, например, в примерах, приведенных в [10].

Заключение

В результате исследования применение байесовских интеллектуальных технологий и методологии идентификации типов распределений случайных величин позволили определить плотность распределения и числовые характеристики для некоторых факторов, характеризующих ценностный мир российской молодежи, которые могут быть использованы для формирования обоснованных и корректных моделей функционирования и формирования ЦММ.

Продемонстрированная методология расширяет область применения БИТ и может стать связующим звеном между байесовским подходом к изучению сложных процессов и систем и классическими эконометрическими методами в условиях значительной неопределенности, уникальных экспериментов и малых выборок, содержащих числовые и нечисловые данные.

Список источников

- [1] Прокопчина С.В. Байесовские интеллектуальные технологии в задачах моделирования закона распределения в условиях неопределенности: монография. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2020. 292 с.
- [2] Прокопчина С.В., Щербаков Г.А., Ефимов Ю.В. Моделирование социально-экономических систем в условиях неопределенности: учебное пособие. М.: Научная библиотека, 2019. 508 с.
- [3] Березин А.С., Жуков Р.А., Прокопчина С.В. Байесовские интеллектуальные измерения индексов и показателей региональной обеспеченности объектами культуры // Мягкие измерения и вычисления. 2022. Т. 53. № 4. С. 5–15.
- [4] Жуков Р.А., Григорьев Е.В., Плинская М.А., Желуницина М.А. О динамике смертности в Тульской области // Мягкие измерения и вычисления. 2022. Т. 60. № 11–2. С. 30–38.
- [5] Жуков Р.А., Григорьев Е.В., Плинская М.А., Желуницина М.А. Прогнозирование показателя смертности населения Тульской области с учетом влияющих факторов // Мягкие измерения и вычисления. 2023. Т. 65. № 4. С. 15–24.
- [6] Pedroza C. A Bayesian forecasting model: predicting U.S. male mortality. *Biostatistics*, 2006, vol. 7, no. 4, pp. 530–550.
- [7] Прокопчина С.В. Основы теории шкалирования в экономике: учебное пособие. М.: Научная библиотека, 2021. 272 с.
- [8] Прокопчина С.В. Инфоаналитик (свидетельство об официальной регистрации программы для ЭВМ № 2004611741 от 12.08.2004).
- [9] Жуков Р.А., Прокопчина С.В., Гиниатов И.А., Николина Е.М. Применение библиотеки "Байесовская математическая статистика" в программном комплексе "Инфоинтегратор" // Мягкие измерения и вычисления. 2022. Т. 54. № 5. С. 99–108.
- [10] Бормотов И.В., Жуков Р.А., Прокопчина С.В. Методология байесовских интеллектуальных измерений для информационных систем и ее применение к изучению ценностного мира российской молодежи // Международная конференция по мягким вычислениям и измерениям. 2023. Т. 1. С. 6–9.

References

- [1] Prokopchina S.V. (2020) Bayesian intelligent technologies in the problems of modeling the distribution law under uncertainty: monograph. M.: Publishing House "SCIENTIFIC LIBRARY", 292 p.

- [2] *Prokopchina S.V., Shcherbakov G.A., Efimov Yu.V.* (2019) Modeling of socio-economic systems in conditions of uncertainty: textbook. Moscow: Scientific Library, 508 p.
- [3] *Berezin A.S., Zhukov R.A., Prokopchina S.V.* (2022) Bayesian intellectual measurements of indices and indicators of regional availability of cultural objects. *Soft measurements and calculations*, vol. 53, no. 4, pp. 5–15.
- [4] *Zhukov R.A., Grigoryev E.V., Plinskaya M.A., Zhelunitsina M.A.* (2022) On mortality dynamics in Tula region. *Soft measurements and calculations*, vol. 60, no. 11-2, pp. 30–38.
- [5] *Zhukov R.A., Grigoryev E.V., Plinskaya M.A., Zhelunitsina M.A.* (2023) Forecasting the mortality rate of the population in the Tula region, according to influenced factors. *Soft measurements and calculations*, vol. 65, no 4, pp. 15–24.
- [6] *Pedroza C.* (2006) A Bayesian forecasting model: predicting U.S. male mortality. *Biostatistics*, vol. 7, no. 4, pp. 530–550.
- [7] *Prokopchina S.V.* (2021) Fundamentals of scaling theory in economics: textbook. Moscow: Scientific Library, 272 p.
- [8] *Prokopchina S.V.* Infoanalitik (certificate of official registration of computer programs No. 2004611741 from 12.08.2004).
- [9] *Zhukov R.A., Prokopchina S.V., Giniyatov I.A., Nikolina E.M.* (2022) Application of the Bayesian Mathematical Statistics library in the Infointegrator software package. *Soft measurements and Calculations*, vol. 54, no. 5, pp. 99–108.
- [10] *Bormotov I.V., Zhukov R.A., Prokopchina S.V.* (2023) Methodology of Bayesian intellectual measurements for information systems and its application to the study of the value system of Russian youth. *International Conference on Soft Computing and Measurements*, vol. 1, pp. 6–9.

БЛАГОДАРНОСТЬ: *Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета № 15841п-П 8.*

*Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.
Авторы заявляют об отсутствии конфликта интересов.
Contribution of the authors: the authors contributed equally to this article.
The authors declare no conflicts of interests.*

Статья поступила в редакцию 13.07.2023; одобрена после рецензирования 20.07.2023; принята к публикации 24.07.2023.

The article was submitted 13.07.2023; approved after reviewing 20.07.2023; accepted for publication 24.07.2023.
