



УДК 519.2

© 2024 г. М.З. Ермолицкая<sup>1,2</sup>, канд. биол. наук

<sup>1</sup>(Институт автоматизации и процессов управления ДВО РАН, Владивосток)

<sup>2</sup>(Владивостокский государственный университет)

## КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ ВРЕМЕННОГО РЯДА ЗАБОЛЕВАЕМОСТИ БОЛЕЗНЯМИ ОРГАНОВ ПИЩЕВАРЕНИЯ В ПРИМОРСКОМ КРАЕ

Проведено исследование временного ряда заболеваемости болезнями органов пищеварения в Приморском крае по данным Росстата. С помощью метода Хольта–Винтерса построен краткосрочный прогноз, показывающий рост заболеваемости в крае болезнями органов пищеварения.

**Ключевые слова:** анализ временных рядов, прогнозирование, болезни органов пищеварения.

DOI: 10.22250/18142400\_2024\_81\_3\_75

### Введение

Болезни органов пищеварения (БОП) относятся к наиболее распространенным заболеваниям среди населения Приморского края. С 2016 г. по 2022 г. на территории края наблюдается рост заболеваемости (зарегистрированных больных с диагнозом, установленным впервые в жизни) с 19.8 до 33.3 случая на 1000 человек населения, что составляет соответственно 2.71% и 3.76% от общего числа заболевших. Заболеваемость БОП встречается у всех возрастных групп, но наиболее распространена среди трудоспособного населения, удельный вес которого в общей численности населения края в 2016 г. составил 58.9%, в 2022 г. - 59.1% [1].

Проблема болезней органов пищеварения имеет ярко выраженную социально-экономическую направленность, так как в период обострения нередко сопровождается развитием осложнений, что приводит к длительной нетрудоспособности, преждевременной инвалидности и смерти больных [2,

---

Работа выполнена в рамках государственного задания FFWF-2022-0002.

3]. В Приморском крае смертность населения от БОП увеличилась с 84.1 случая на 100 тыс. населения в 2016 г. до 94.9 случая – в 2022 г.

Мониторинг и прогнозирование динамики заболеваемости населения позволяют планировать мероприятия по управлению ресурсами медицинских организаций, направленные на профилактику заболеваний на ранних стадиях и своевременное лечение пациентов [4, 5]. Для этих целей разработано большое число моделей с применением статистических методов анализа данных, методов математического моделирования и машинного обучения, основанных на исследовании структуры и динамики заболеваемости, выявлении модифицированных факторов риска [6 – 15].

Целью данного исследования является построение краткосрочного прогноза заболеваемости болезнями органов пищеварения в Приморском крае с помощью метода Хольта–Винтерса.

### Материалы и методы

Выборка данных по заболеваемости болезнями органов пищеварения (зарегистрированных больных с диагнозом, установленным впервые в жизни, на 1000 человек населения) в Приморском крае за 25-летний период (с 1998 г. по 2022 г.) получена из материалов официального сайта Федеральной службы государственной статистики [1]. Статистический анализ данных осуществлялся в программе RStudio Pro (Version 2022.12.0 Build 353.pro20). Проверка нулевых гипотез проводилась на заданном уровне значимости  $p = 0.05$  [16].

Преобразование ряда данных во временной ряд осуществлялось с помощью функция `ts{stats}` с частотой наблюдений в один год:

$$y_t = Q \cdot y_{t-1} + \varepsilon_t, \quad (1)$$

где  $y_t$  – значение временного ряда в момент  $t$ ;  $Q$  – коэффициент;  $\varepsilon_t$  – случайная ошибка.

Для проверки нулевой гипотезы о нестационарности временного ряда использовали тест Дики–Фуллера (`adf.test{tseries}`) [17]. Статистика теста имеет следующий вид:

$$\hat{t} = \frac{1-Q}{\sqrt{\frac{1-Q^2}{n}}}, \quad (2)$$

где  $n$  – объем выборки.

Для проверки нулевой гипотезы об отсутствии тренда применили тест тренда Манна–Кендалла (`MannKendall{Kendall}`). Статистика теста Манна–

Кендалла:

$$Z = \begin{cases} \frac{(S-1)}{\sigma}, S > 0, \\ 0, S = 0, \\ \frac{(S+1)}{\sigma}, S < 0; \end{cases} \quad S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(y_j - y_i), i < j, \quad (3)$$

где  $\sigma$  – стандартное отклонение  $S$ ;  $\text{sign}(y_j - y_i)$  – функция, которая возвращает  $-1$ , если аргумент отрицательный,  $0$  – если он равен нулю, и  $1$  – если он положительный.

Наличие тенденции и сезонности в ряде наблюдений проверяли с помощью расчета коэффициентов автокорреляции и построения коррелограммы (функции `acf{stats}`, `pacf{stats}`). Дополнительно использовали критерий Фостера–Стюарта (`foster.test{RecordTest}`) для проверки наличия тренда в средних значениях и в дисперсиях.

Статистика критерия Фостера–Стюарта для обнаружения тренда в средних:

$$t = \frac{D}{\sigma_D}, \quad D = \sum_{i=2}^n (m_i - k_i), \quad \sigma_D = \sqrt{2 \cdot \ln(n) - 0.8456}, \quad (4)$$

где  $m_i = 1$ , если  $y_i$  больше всех предшествующих уровней, иначе  $m_i = 0$ ;  $k_i = 1$ , если  $y_i$  меньше всех предшествующих уровней, иначе  $k_i = 0$ .

Статистика критерия Фостера–Стюарта для проверки тренда в дисперсиях:

$$t = \frac{S - \sigma_D^2}{\sigma_S}, \quad S = \sum_{i=2}^n (m_i + k_i), \quad \sigma_S = \sqrt{2 \cdot \ln(n) - 3.4253}. \quad (5)$$

Прогнозирование временного ряда осуществляли по данным с 1998 г. по 2021 г. с помощью модели экспоненциального сглаживания Хольта-Уинтерса (`HoltWinters{stats}`). Использовали модель двойного экспоненциального сглаживания с двумя составляющими временного ряда:  $l$  – уровень и  $b$  – тренд.

$$\begin{aligned} l_t &= \alpha \cdot y_t + (1 - \alpha) \cdot (l_{t-1} + b_{t-1}), 0 \leq \alpha \leq 1, \\ b_t &= \beta \cdot (l_t - l_{t-1}) + (1 - \beta) \cdot b_{t-1}, 0 \leq \beta \leq 1, \\ \hat{y}_{t+1} &= l_t + b_t, \end{aligned} \quad (6)$$

где  $\alpha$  – параметр сглаживания ряда вокруг тренда;  $\beta$  – параметр сглаживания самого тренда;  $\hat{y}_{t+1}$  – прогноз на момент  $(t + 1)$ .

В качестве оценок точности прогноза рассчитывали значения средней абсолютной ошибки (MAE) и среднеквадратической ошибки (MSE). Данные за 2022 г. использовали для сравнения с прогнозным значением

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}, \quad MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}. \quad (7)$$

### Исследование временного ряда

Графическое изображение ряда данных по заболеваниям БОП показывает, что временной ряд не является стационарным, присутствует тренд (рис. 1).

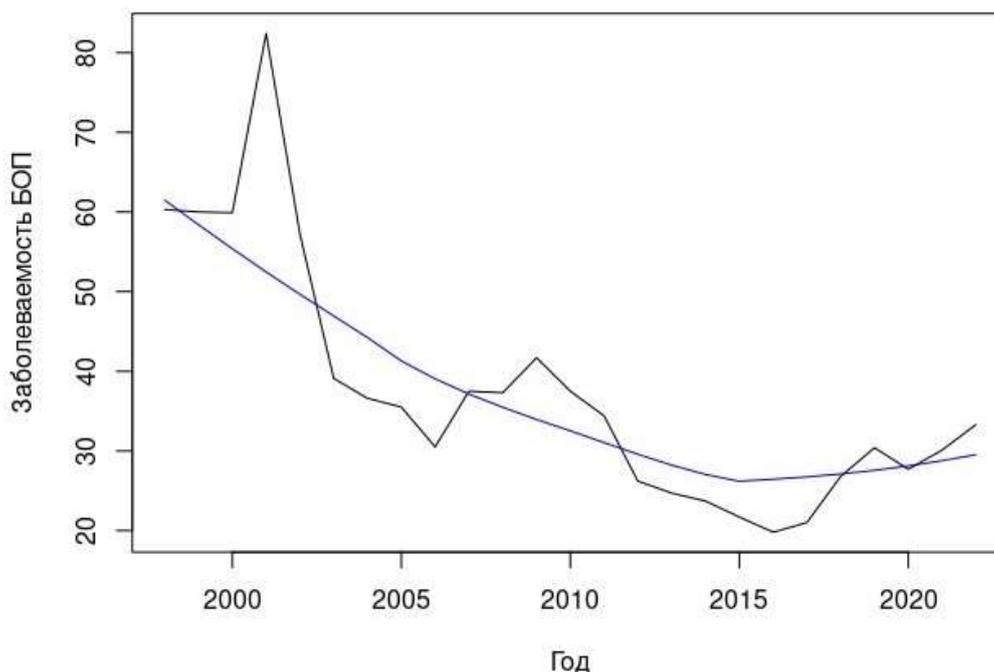


Рис. 1. Динамика временного ряда заболеваемости БОП с регрессионной кривой.

Примечание: регрессионная кривая получена с помощью локально-взвешенной полиномиальной регрессии с коэффициентом сглаживания 2/3 (функция `lowess {stats}`).

Это подтверждает и результат применения расширенного теста Дикки–Фуллера:  $Dickey-Fuller = -1.6029$ ,  $p\text{-value} = 0.7236$  (нулевую гипотезу приняли, временной ряд нестационарный). Согласно тесту Манна–Кендалла в ряде данных присутствует тренд ( $\tau = -0.598$ ,  $p = 3.206e-05$ ), т.е., временной ряд имеет некоторую структуру, зависящую от времени.

Судя по графику (рис.1), можно предположить, что разности первого порядка ряда данных стационарны. Применение теста Дикки–Фуллера подтверждает это предположение ( $Dickey-Fuller = -5.7973$ ,  $p = 0.01$ ). Графическое изображение разности первого порядка для данного временного ряда с линией тренда – на рис. 2.

Изучим характеристики исходного временного ряда – сезонность и цикличность. Для этого построим коррелограмму – график автокорреляции уровней временного ряда (рис. 3). Значимыми считаются показатели, лежащие выше/ниже доверительного интервала.

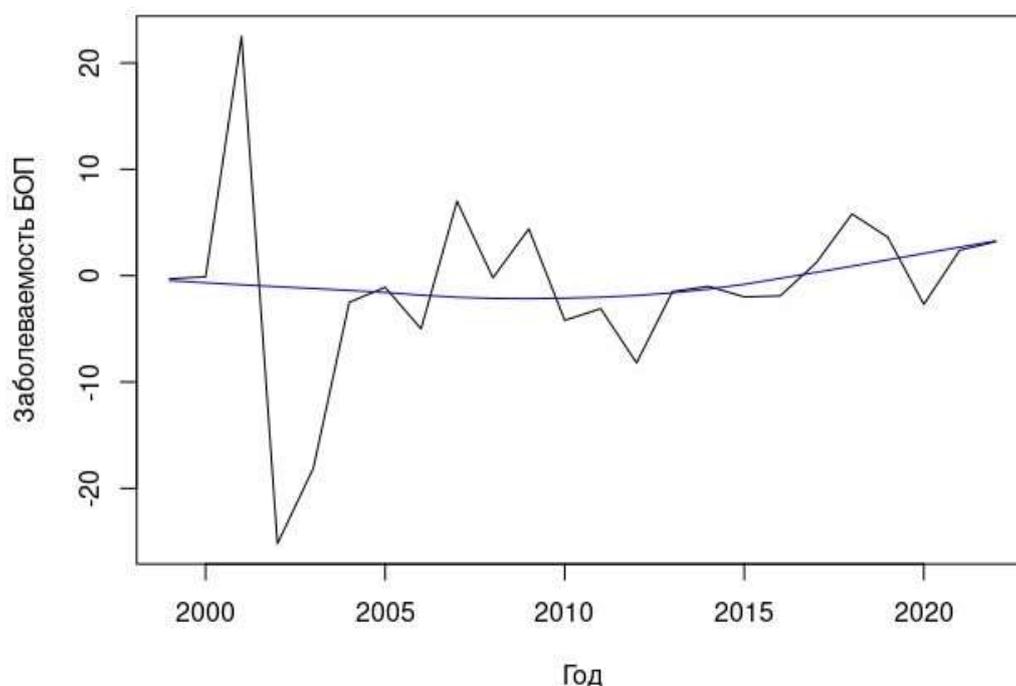


Рис. 2. График разности первого порядка для временного ряда заболеваемости БОП с регрессионной кривой. Примечание: регрессионная кривая получена с помощью локально-взвешенной полиномиальной регрессии с коэффициентом сглаживания 2/3 (функция `lowess {stats}`).

На рис. 3, а) видно, что имеются положительные, затем отрицательные коэффициенты корреляции. Максимальный коэффициент соответствует первому лагу (0.798). Это указывает на наличие во временном ряду убывающего тренда. Присутствуют коэффициенты корреляции, значимо отличающиеся от нуля (выше пунктирной линии – лаг 1:3). График частичной автокорреляции исключает влияние промежуточных уровней ряда (рис. 3, б)). Пики и значимые показатели указывают на потенциальные порядки в моделях временных рядов. В нашем случае все пики находятся в пределах значимости. Четкой сезонности не наблюдается, но можно видеть слабо выраженную цикличность.

Дополнительно использовали критерий Фостера–Стюарта. Нулевые гипотезы об отсутствии тенденции в средних и в дисперсиях не отклоняются ( $Z = -4.6352$ ,  $p = 1$  и  $Z = 1.0185$ ,  $p = 0.1542$  соответственно).

Таким образом, исследование временного ряда показало, что исходный ряд данных нестационарный, с наличием тренда. Взятие разности первого порядка приводит ряд данных к стационарному виду.

### Результаты прогнозирования

При прогнозировании временного ряда с помощью метода Хольта–Винтерса (HW) использовали модель двойного экспоненциального сглаживания (несезонная модель).

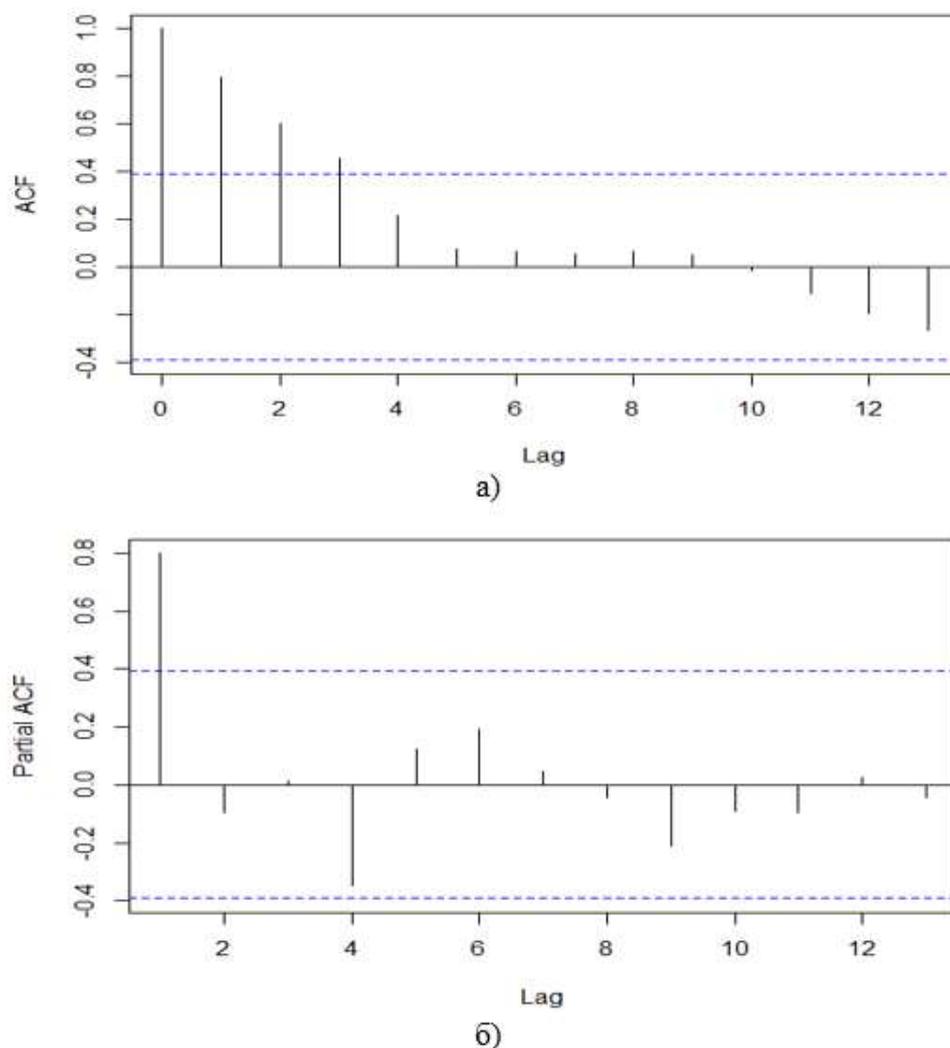


Рис. 3. Графики функции автокорреляции (ACF) – а) и частичной автокорреляции (Partial ACF) – б); пунктирные линии – доверительный интервал; Lag – лаг (число периодов, по которым рассчитывается коэффициент автокорреляции между парами элементов ряда).

Наилучшая аддитивная модель получена со следующими параметрами:  $\alpha = 0.9597$ ,  $\beta = 0.9658$ . Предсказанное значение временного ряда на 2022 г. равно 33.188 случая (рис. 4). По данным Росстата, в 2022 г. заболеваемость болезнями органов пищеварения (зарегистрированных больных с диагнозом, установленным впервые в жизни, на 1000 человек населения) в Приморском крае составила 33.3 случая. Коэффициент корреляции между наблюдаемыми и предсказанными значениями равен 0.998478. Оценки точности прогноза: MAE = 0.1779, MSE = 0.08428.

### Заключение

Исследование временного ряда по заболеваемости болезнями органов пищеварения позволило построить прогнозную модель. Полученные результаты показывают рост заболеваемости болезнями органов пищеварения в

Приморском крае, что соответствует данным Росстата. Для построения точного прогноза на долгий период необходимы данные за более длительный промежуток времени.

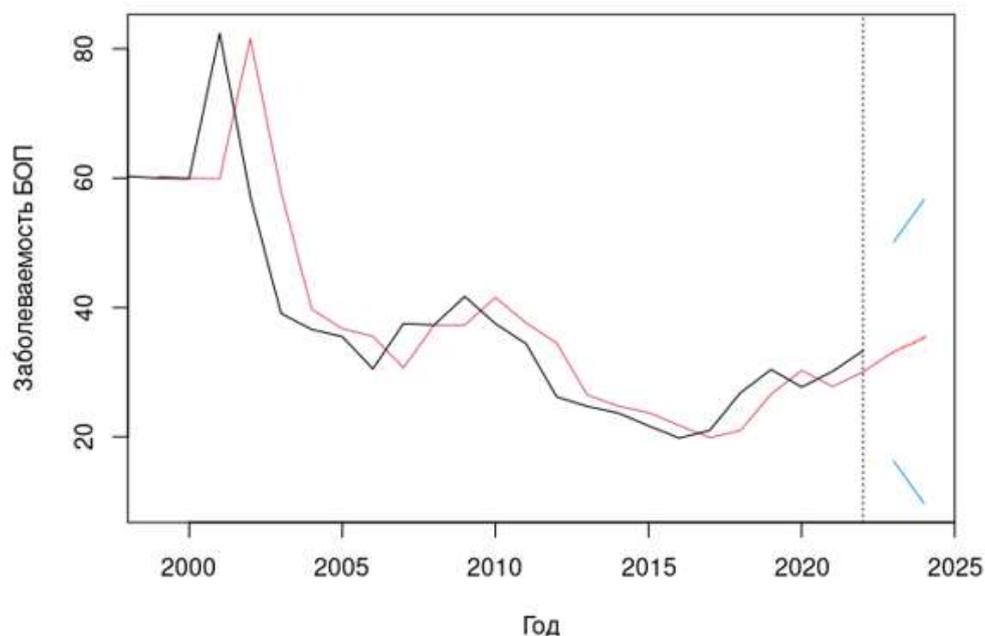


Рис. 4. График обследуемых и предсказанных значений заболеваемости болезнями органов пищеварения в Приморском крае по методу Хольта–Винтерса (красная линия – прогноз).

Полученные оценки точности прогноза дают возможность использовать результаты данного исследования в качестве дополнительного инструмента при совершенствовании санитарно-профилактических мероприятий на региональном уровне.

#### ЛИТЕРАТУРА

1. Федеральная служба государственной статистики: официальный сайт: URL: <https://rosstat.gov.ru> (дата обращения: 15.05.2024).
2. Будаев Б.С., Кицул И.С., Банзарова Л.П., Тармаева И.Ю., Богданова О.Г. Болезни органов пищеварения: структура и динамика на региональном уровне // Проблемы социальной гигиены, здравоохранения и истории медицины. - 2022. – Т. 30, №2. – С. 232–238.
3. Егянян Р.А., Кушунина Д.В., Калинина А.М. Актуальность и результативность раннего выявления заболеваний органов пищеварения при диспансеризации взрослого населения России // Профилактическая медицина. – 2017. – Т. 20, № 3. – С. 22–27.
4. Ермолицкая М.З. Прогнозирование временного ряда заболеваемости болезнями органов пищеварения // Здоровье населения и среда обитания - ЗНиСО. – 2023. – Т. 31, №6. – С. 20–26.
5. Шмакова Н.Н., Дмитриева Т.В., Дмитриев В.Н., Урусова М.А., Андреева Н.А., Пензев С.А. и др. Региональные особенности медико-демографических показателей здоровья населения Белгородской области (на примере болезней органов пищеварения за пе-

- риод 2000-2017 гг.) // Здоровье населения и среда обитания - ЗНиСО. – 2021. – Т. 29, №7. – С. 57–66.
6. Головин С.В., Козлов Ю.С., Козлова Л.В. Прогнозирование развития уровня общей заболеваемости населения на территории Воронежской области // Вестник ВГТУ. – 2011. – Т. 7, № 3. – С. 195–197.
  7. Sekeroglu B., Tuncal K. Prediction of cancer incidence rates for the European continent using machine learning models // Health Informatics Journal. – 2021. – Vol. 27(1): 1460458220983878.
  8. Тарасова С.А. Прогнозирование временного ряда инфекционной заболеваемости // Программные продукты и системы. – 2019. – № 2. – С. 337–342.
  9. Приземина И.Н., Луговсков А.Д., Знагован С.Ю., Козикова О.А., Перцова Ю.Г. Прогнозирование заболеваемости острыми кишечными инфекциями // Medicus. – 2019. – № 4(28). – С. 48–52.
  10. Загоруйченко А.А., Карпова О.Б. Актуальные подходы к прогнозированию и моделированию заболеваемости населения в России (обзор) // Санитарный врач. – 2022. – № 8. – С. 596–606.
  11. Филатова Е.Н., Солнцев Л.А. Смешанная техника прогнозирования инфекционной заболеваемости с применением методов декомпозиции временного ряда и SARIMA // Журнал микробиологии, эпидемиологии и иммунобиологии. – 2019. – № 3. – С. 76–83.
  12. Taylor S.J., Benjamin Letham B. Forecasting at Scale // The American Statistician. – 2018. – Vol. 72, № 1. – P. 37–45. URL: <http://lethalletham.com/ForecastingAtScale.pdf> (дата обращения: 15.05.2024).
  13. Asadi R., Regan A.C. A spatio-temporal decomposition based deep neural network for time series forecasting // Applied Soft Computing. – 2020. – Vol. 87.
  14. Косова А.А., Чалана В.И. Прогнозирование динамики заболеваемости норовирусной инфекцией с применением моделей временного ряда // Уральский медицинский журнал. – 2023. – № 22(3). – С.5 –63.
  15. Кондратьев М.А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование. – 2013. – Т.5, № 5. – С. 863–882.
  16. Мاستицкий С.Э. Анализ временных рядов с помощью R. [Электронная книга]. URL: <https://ranalytics.github.io/tsa-with-r/> (дата обращения: 15.05.2024).
  17. Лемешко Б.Ю., Комиссарова А.С., Щеглов А.Е. Вопросы применения некоторых критериев проверки случайности и отсутствия тренда // Метрология. – 2010. – № 12. – С. 3–25.

*Статья представлена к публикации членом редколлегии А.И. Абакумовым.*

*E-mail:*

*Ермолицкая Марина Захаровна – ermmz@mail.ru.*