

Применение методов анализа и прогнозирования временных рядов эпидемиологической ситуации новой коронавирусной инфекции (COVID-19) в Тюменской области.

Т. Ф. Степанова¹, Г.В. Шарухо², А.Н. Летюшев^{1,2}, А.П. Ребещенко¹, А.Я. Фольмер³, А.А.Шепоткова¹.

¹ ФБУН «Тюменский научно-исследовательский институт краевой инфекционной патологии» Роспотребнадзора, г. Тюмень

² Управление Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека по Тюменской области

³ ФБУЗ «Центр гигиены и эпидемиологии в Тюменской области»

Первый случай новой коронавирусной инфекции (COVID-19) был зарегистрирован на территории Тюменской области 31 января 2020 года у гражданина КНР, проходящего обучение в Тюменском индустриальном институте и прибывшего после каникул из г. Цзинань провинции Шаньдун 28 января в город Тюмень. Спорадические случаи заболевания в регионе регистрировались до начала апреля 2020 года. Ежедневно новые случаи на территории Тюменской области регистрируются с 9 апреля. На 1 июля на территориях 25 из 26 муниципальных образований субъекта COVID-19 диагностирован у 3 231 человек. Необходимость точного прогноза новых случаев заболевания новой коронавирусной инфекции COVID-19 в Тюменской области в еженедельном режиме продиктована потребностью оперативного принятия управленческих решений и формирования комплекса противоэпидемических мер, планирование коечного фонда, объемов тестирования.

Цель: составление прогноза новых случаев коронавирусной инфекции в Тюменской области с помощью методов одномерных временных рядов на краткосрочный период.

Процедуры статистического анализа выполнялись с помощью статистического пакета SPSS- версия 22.0, предназначенным для научных исследований, отвечающих требованиям клинической эпидемиологии.

Материалами для настоящего исследования послужили ежедневные статистические данные о случаях новой коронавирусной инфекции (COVID-19), опубликованные на официальном сайте <https://стопкоронавирус.рф> и информации Управления Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека по Тюменской области за период с 09.04.2020 по 30.06.2020 г.

В работе исследуются и сравниваются подходы, относящиеся к адаптивным методам прогнозирования, способные учитывать новые условия развития эпидемического процесса COVID-19: простое экспоненциальное сглаживание, модель Хольта, модель Брауна, модель Бокса – Дженкинса

(ARIMA) с определением 95% доверительный интервал прогноза. На отсутствие автокорреляции в остатках был проведен тест Льюинга– Бокса [2, 4,5]. Для оценки качества и точности модели использован нормализованный байесовский информационный критерий (BIC) и средняя абсолютная ошибка (MAPE).

Наиболее простым и широко известным подходом, позволяющим устранить случайные и периодические колебания в краткосрочном периоде, является метод простого экспоненциального сглаживания. Временной ряд генерируется моделью (формула 1), где $a_{1,t}$ - варьирующий во времени средний уровень ряда; ε_t - случайные неавтокоррелированные отклонения с нулевым математическим ожиданием и дисперсией σ^2 .

$$x_t = a_{1,t} + \varepsilon_t \text{ (формула 1)}$$

Для проверки отсутствие автокорреляции в остатках и определения качества прогнозной модели рассчитаны Q-тест Льюнга и нормализованный байесовский информационный критерий (BIC). Результаты анализа представлены в таблице 1. Согласно полученным данным, можно принять нулевую гипотезу об отсутствии автокорреляции в остатках.

Таблица 1

Результаты Q-теста Льюнга – Бокса и BIC

Модель	Q-статистика Льюнга-Бокса			Нормализованный BIC
	Статистика	DF	Значимость	
Простая	21,189	17	0,218	5,360
Хольта	20,795	16	0,187	5,414
Брауна	24,452	17	0,108	5,465
ARIMA	12,594	17	0,763	4,676

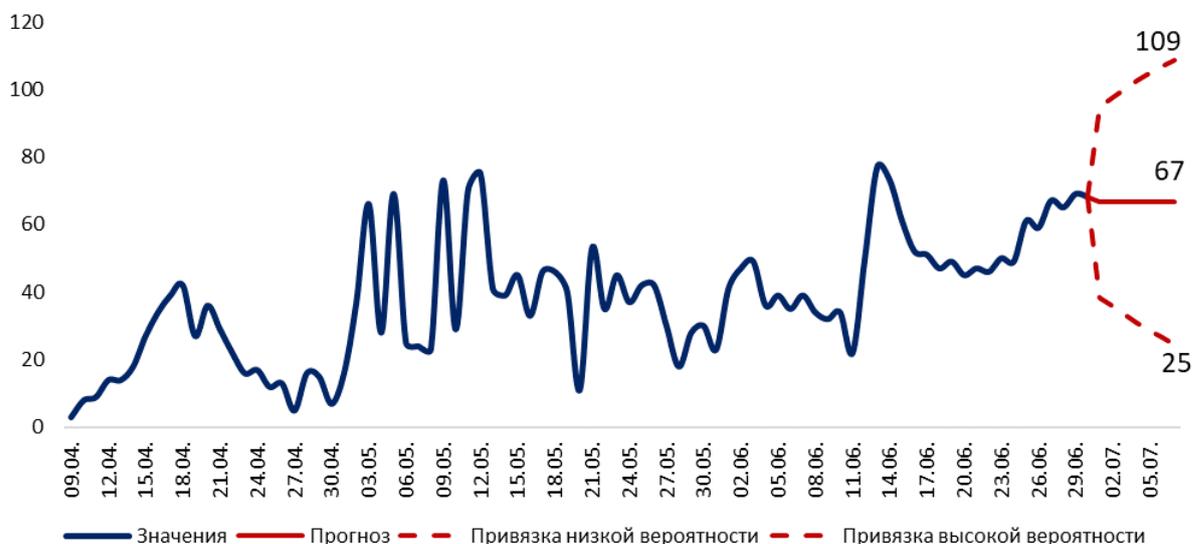


Рисунок 1 Динамика и прогноз случаев коронавирусной инфекции в Тюменской области методом простого экспоненциального сглаживания

Полученная по методу простого экспоненциального сглаживания модель имела следующие статистические характеристики: средняя абсолютная ошибка (МАРЕ) – 0,36; ВИС – 5,360, критерий Льюнга – Бокса, показывающие отсутствие корреляции ($p=0,218$).

В среднем субъекте в день будет регистрироваться порядка 67 случаев (95% ДИ от 25 до 109), точность прогноза =64%.

Вариантами адаптивных моделей, использующими процедуру экспоненциального сглаживания, являются модели линейного роста (модели Хольта и Брауна).

Оценка коэффициентов модели Хольта производится по следующим формулам 2 и 3 [4,6], где коэффициенты α_1 и α_2 – параметры экспоненциального сглаживания или адаптации.

$$\hat{a}_{(1,t)} = a_1 x_t + (1 - a_1) \hat{a}_{(1,t-1)} \quad (\text{формула 2})$$

$$\hat{a}_{(2,t)} = \hat{a}_2 (\hat{a}_{(1,t)} - \hat{a}_{(1,t-1)}) + (1 - \hat{a}_2) \hat{a}_{(2,t-1)} \quad (\text{формула 3})$$

Результаты теста Льюнга – Бокса и нормализованные значения информационного критерия Байеса (ВИС) приведены в таблице 1.

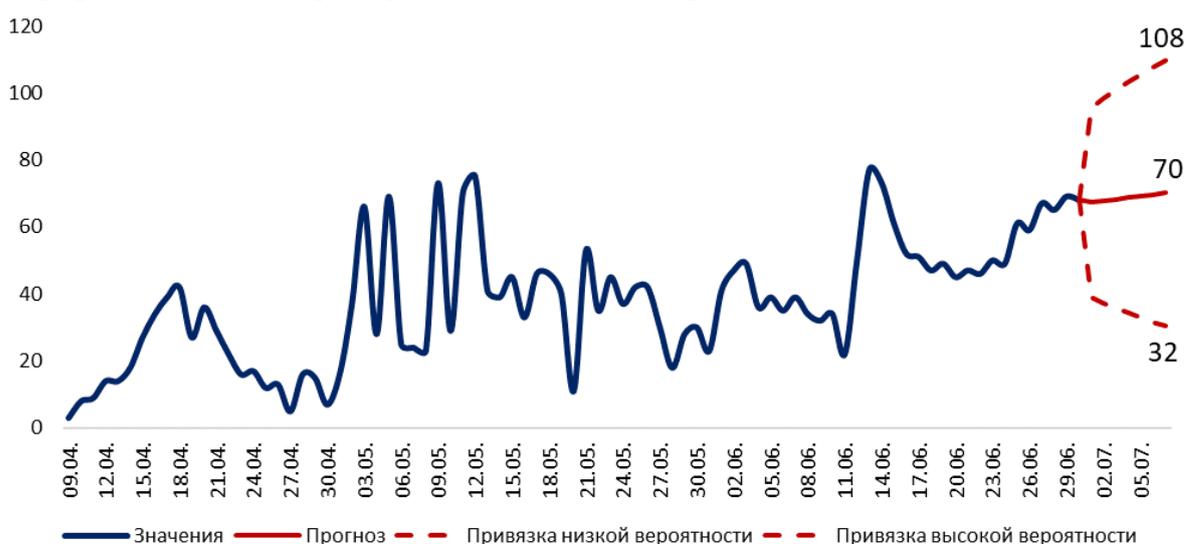


Рисунок 2 Динамика и прогноз случаев коронавирусной инфекции в Тюменской области методом Хольта

Полученная по методу Хольта модель имела следующие статистические характеристики: средняя абсолютная ошибка (МАРЕ) – 0,37; ВИС – 5,414, критерий Льюнга – Бокса, показывающие отсутствие корреляции ($p=0,187$).

В среднем субъекте в день будет регистрироваться порядка 67 случаев (95% ДИ от 25 до 109), точность прогноза =63%.

Частным случаем модели Хольта является однопараметрическая модель Брауна и рассчитываются по формулам 4 и 5 [4,6], где $e_t = x_t - \hat{x}_1(t-1)$ – ошибка прогноза, $0 < \beta < 1$ – коэффициент дисконтирования.

$$\hat{a}_{(1,t)} = \hat{a}_{(1,t-1)} \quad (\text{формула 4})$$

$$\hat{a}_{(2,t)} = \hat{a}_{(2,t-1)} + (1 - \beta^2) e_t \text{ (формула 5)}$$

Результат теста Льюнга – Бокса и ВИС представлены в таблице 1.

Полученная по методу Брауна модель имела следующие статистические характеристики: средняя абсолютная ошибка (MAPE) – 0,46; ВИС – 5,465, критерий Льюнга – Бокса - $p=0,107$.

В среднем субъекте в день будет регистрироваться порядка 77 случаев (95% ДИ от 33 до 121), точность прогноза =54%

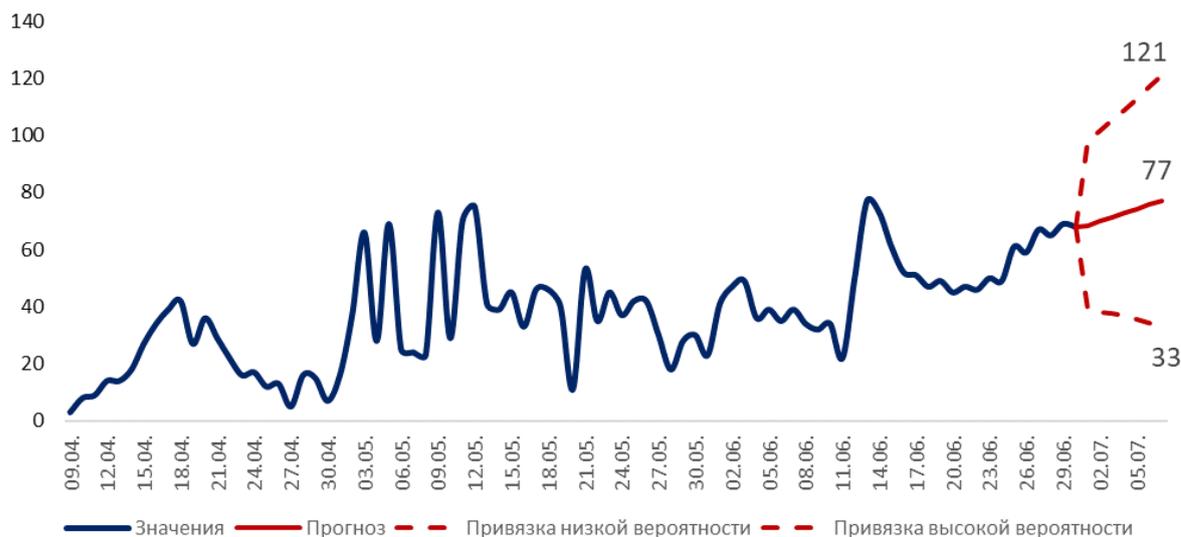


Рисунок 3 Динамика и прогноз случаев коронавирусной инфекции в Тюменской области методом Брауна

Мощным инструментом прогнозирования временного ряда является интегрированная модель авторегрессии– скользящего среднего ARIMA [1, 3]

Для использования данной модели необходимо соблюдение условия стационарности зависимой переменной (отсутствие тренда, постоянная дисперсия (гомоскедастичность), постоянное значение автокорреляции для всего ряда. После дифференцирования исходного временного ряда и анализа эмпирических автокорреляционных и частных автокорреляционных функций для прогнозирования была выбрана модель ARIMA (1, 1, 0).

Адекватность выбора модели временного ряда проверена с помощью АКФ и ЧАКФ остатков (рис3). Согласно визуализированным данным, ошибки не коррелируют.

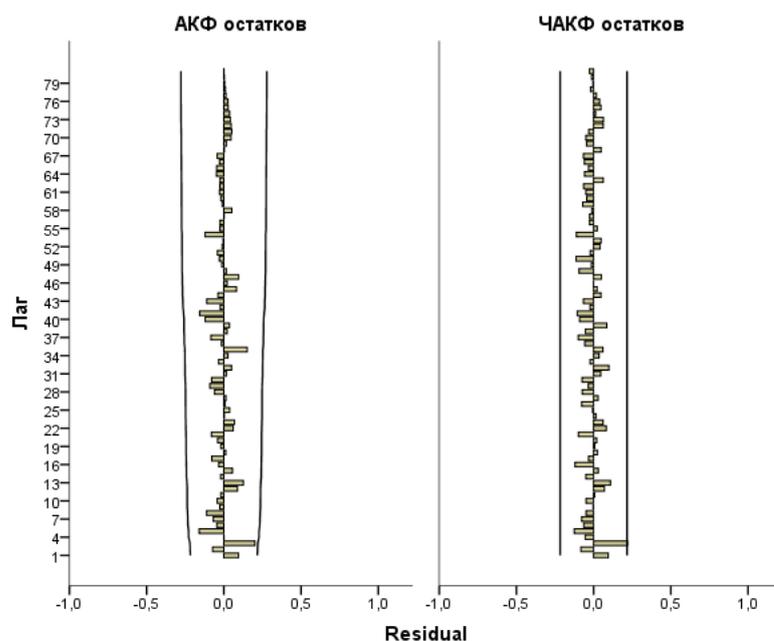


Рисунок 4 Автокорреляционная и частная автокорреляционная функции для остатков ARIMA (1, 1, 0)

Результаты теста Льюнга – Бокса, а также нормализованные значения информационного критерия Байеса приведены в таблице 1.

Полученная модель имела следующие статистические характеристики: средняя абсолютная ошибка (MAPE) – 0,22; BIC – 4,676, критерий Льюнга – Бокса, показывающие отсутствие корреляции ($p=0,763$).

Таким образом, полученные прогнозные значения заболеваемости COVID-19 с помощью модели ARIMA (1, 1, 0) с константой показывают, что в среднем субъекте в день будет регистрироваться порядка 68 случая (95% ДИ от 40 до 96), точность прогноза =78%.

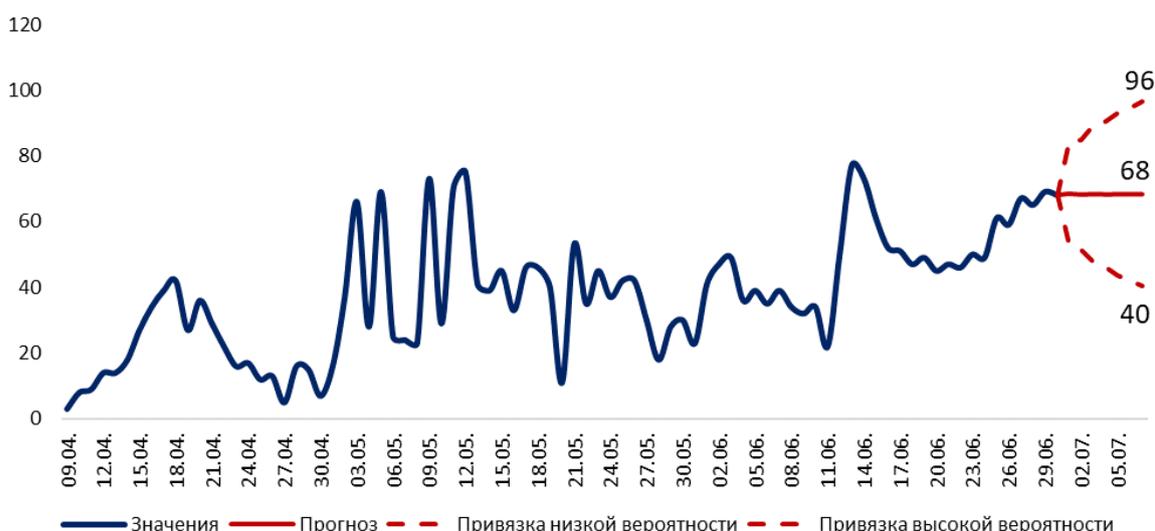


Рисунок 5 Динамика и прогноз случаев коронавирусной инфекции в Тюменской области методом ARIMA (1, 1, 0)

Как видно из таблицы 1, наименьшее значение показателя BIC соответствует модели ARIMA (4,676). Из этого следует, что указанная модель является лучшей среди рассмотренных в данной статье.

Методология выбора различных моделей прогнозирования используется при еженедельном анализе эпидемиологической ситуации COVID-19 в Ханты-Мансийском автономном округе- Юрге и Ямало-Ненецком автономном округе.

1. Бокс Дж., Дженкинс Г.М. Анализ временных рядов. Прогноз и управление. — Мир, 1974.
2. Ибрагимова З.Ф. Прогнозирование уровня бедности населения в Российской Федерации на основе методов временных рядов //Вестник экономики, права и социологии. – 2017. – № 4. – С. 44-48
3. Крюков Ю.А., Чернягин Д.В. ARIMA – модель прогнозирования значений трафика // Информационные технологии и вычислительные системы. – 2011. – № 2. – С. 41-49
4. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие. – М.: Финансы и статистика, 2003. – 416 с.
5. Ostertagova, E., and Ostertag, O. Forecasting using simple exponential smoothing method. – 2012. Acta Electrotechnica et Informatica, 12(3), pp. 62
6. Prajakta S. Kalekar (2004). Time series forecasting using Holt-Winters Exponential Smoothing. Kanwal Rekhi School of Information Technology.