

М. Г. Коротков, А. А. Петров, М. В. Куркина

**О ПРИМЕНИМОСТИ МОДЕЛЕЙ АРПСС ПРИ МОДЕЛИРОВАНИИ
И ПРОГНОЗИРОВАНИИ ДИНАМИКИ НОВЫХ СЛУЧАЕВ ЗАРАЖЕНИЯ COVID-19**

Целью работы является исследование применимости методологии построения моделей АРПСС в проблеме моделирования и прогнозирования динамики новых случаев заражения коронавирусом в осенне-зимний период 2020 года в ХМАО-Югре.

Ключевые слова: временной ряд, статистические методы, модель АРПСС, прогнозирование.

M. G. Korotkov, A. A. Petrov, M. V. Kurkina

**ON THE APPLICABILITY OF ARPS MODELS IN MODELING
AND FORECASTING THE DYNAMICS OF NEW CASES OF COVID-19 INFECTION**

The aim of the work is to study the applicability of the methodology for constructing ARIMA models in the problem of modeling and predicting the dynamics of new cases of coronavirus infection in the autumn-winter period of 2020 in the KhMAO-Yugra.

Keywords: time series, statistical methods, ARIMA model, forecasting.

Введение

При попытках построения математических моделей и прогноза процесса распространения COVID-19 для региона в качестве исходных данных из открытых источников можно рассматривать временной ряд количества зараженных за день.

Факторы, влияющие на процесс, можно условно разделить на два типа.

1. Разовые административные решения, приводящие к достаточно быстрому и резкому изменению тренда рассматриваемого процесса:

- объявление внеочередных нерабочих дней;
- ввод/отмена режима самоизоляции;
- ввод/отмена дистанционного формата обучения в образовательных учреждениях;
- ввод/отмена контроля и запрета въезда из-за пределов региона;
- ввод/отмена ограничений работы объектов общепита и досуга;
- и т. п.

2. Постоянно действующие факторы, информацию о которых достаточно сложно найти в открытых источниках:

- количество ежедневно проводимых тестирований по выявлению заражения коронавирусной инфекцией;
- количество ежедневно прививаемых;
- количество и ковидный статус ежедневно прибывающих и убывающих в регионе;
- и др.

Факторы первого типа характеризуются различным временем запаздывания влияния на рассматриваемый процесс и наличием мультипликативного эффекта.

Заметим, что логарифмирование исходных данных временного ряда позволяет учесть влияние мультипликативного эффекта факторов первого типа.

Существует несколько подходов при построении математических моделей процесса. Один из них – это учет влияния факторов через выделение нелинейного тренда (для всего

ряда). Такой подход целесообразен при сравнении процессов, протекающих в различных регионах, с целью выявления различий и причин появления этих различий.

Целью же данной работы построение модели АРПСС для одного региона.

Постановка задачи

Применимость методологии построения модели АРПСС [1, 2] и прогноза на ее основе рассмотрим на примере динамики новых случаев заражения COVID-19 в осенне-зимний период 2020 года в ХМАО-Югре (открытые данные сайта coronavirus-monitor.ru).



Рисунок 1

На рис. 1 представлен график за период с 01.09.2020 по 07.01.2021.

Возможность использования предлагаемого к рассмотрению подхода проиллюстрируем, предполагая, что прогноз будет строиться по данным с 01.09.2020 по 24.12.2020 (рис. 2) на двухнедельный период с 25.12.2020 по 07.01.2021



Рисунок 2

В соответствии с методикой, изложенной в [3], для построения модели АРПСС в качестве исходного ряда выбираем данные, начиная с 16.10.2020 (рис. 3) (убираем из рассмотрения данные с 01.09.20 по 15.10.20).

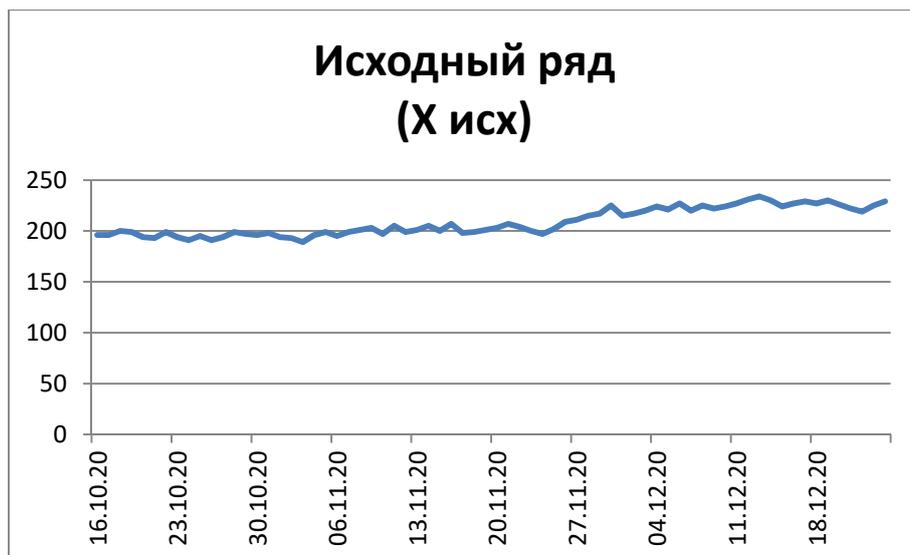


Рисунок 3

Этапы построения модели и прогноза

Этапы построения модели и прогноза проводятся методами, предлагаемыми в [4].

Логарифмирование исходного ряда

Для учета мультипликативных эффектов влияния факторов проводится логарифмирование исходного ряда (рис. 4).



Рисунок 4

Идентификация модели АРПСС

Приведение ряда к виду, близкому к стационарному, достигается взятием разностей второго порядка (рис. 5).

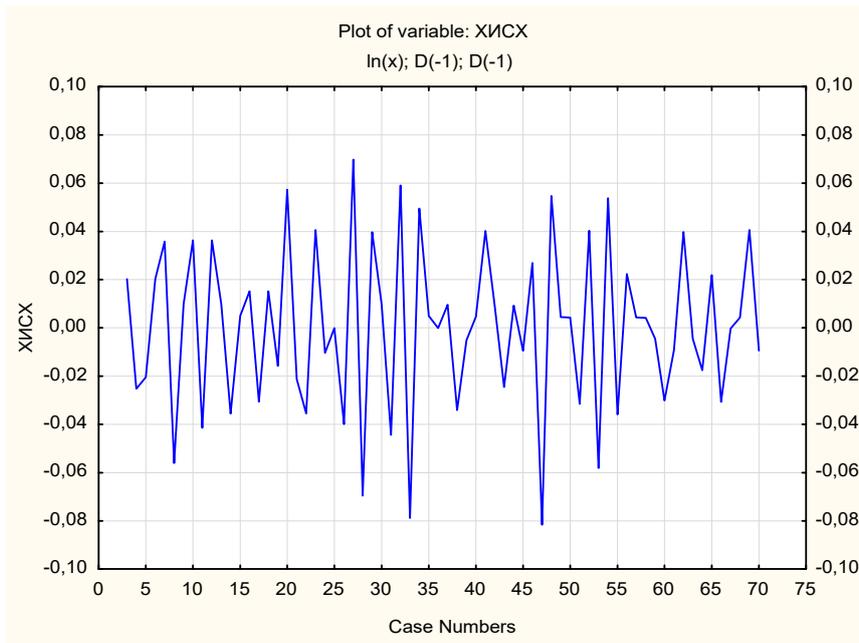


Рисунок 5

Анализ поведения автокорреляционной и частной автокорреляционной функций дает основания считать, что порядок авторегрессии модели АРСС $p=1$, порядок скользящего среднего $q=1$ (рис. 6а, 6б).

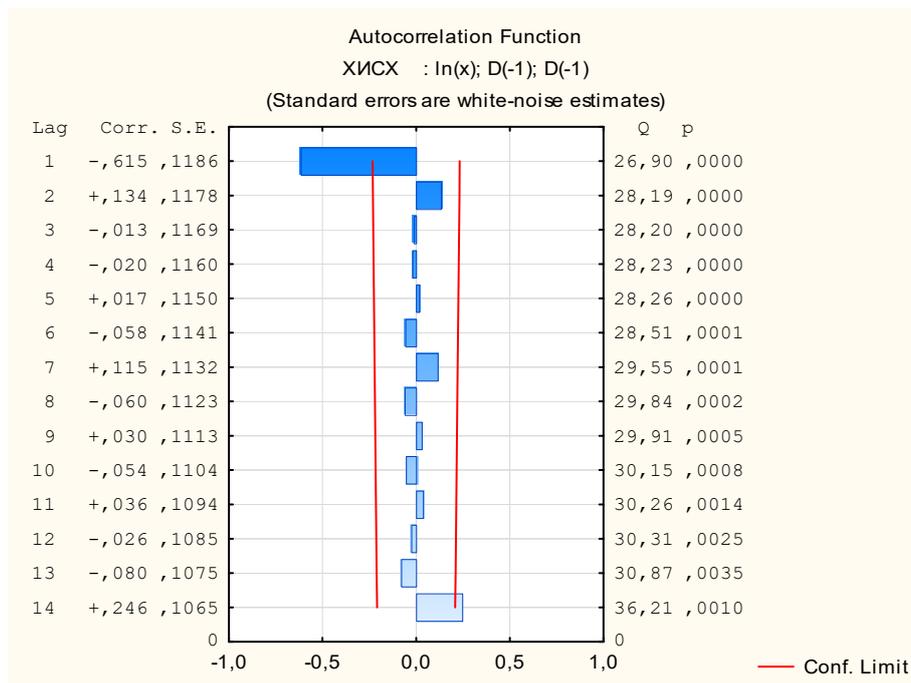


Рисунок 6а

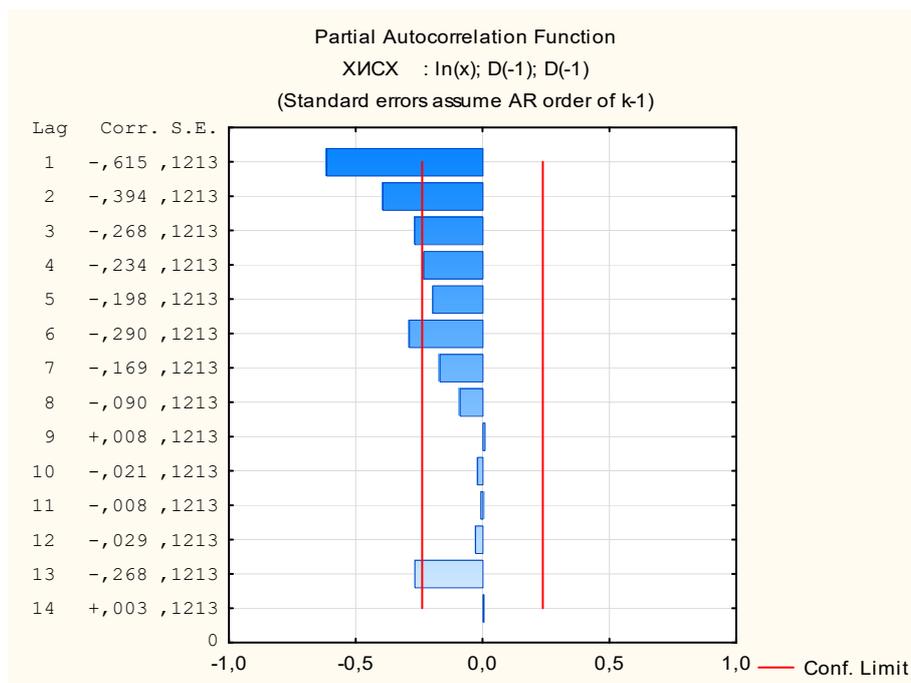


Рисунок 6б

Учитывая недельный цикл функционирования всех служб и систем, влияющих на отбор и обработку данных, можем предполагать наличие семидневной периодичности в рассматриваемом процессе.

Проведенный Фурье-анализ подтверждает предположение, что сезонный лаг $s=7$. Проведя процедуры, аналогичные вышеописанным, получаем, что сезонный параметр авторегрессии $P_s=1$, сезонная разность $d_s=1$, сезонный параметр скользящего среднего $Q_s=0$.

Таким образом, приходим к модели АРПСС (1,2,1) (1,1,0) с сезонным лагом, равным 7.

Оценка параметров модели АРПСС.

В результате процедуры оценивания получим значения параметров модели АРПСС (рис. 7).

Input: XICX (COR_2309_1_7_sta)						
Transformations: ln(x), 2*D(1), D(7)						
Model: (1,2,1)(1,1,0) Seasonal lag: 7 MS Residual= ,00060						
Paramet.	Param.	Asympt. Std.Err.	Asympt. t(58)	p	Lower 95% Conf	Upper 95% Conf
p(1)	-0,336549	0,134223	-2,50738	0,014985	-0,605226	-0,067872
q(1)	0,840715	0,072513	11,59393	0,000000	0,695564	0,985867
Ps(1)	-0,526717	0,118983	-4,42681	0,000043	-0,764888	-0,288546

Рисунок 7

Результаты показывают, что все значения параметров статистически значимы (с 5 % уровнем значимости).

Исследование адекватности модели АРПСС

Эта процедура проводится на основе анализа остатков модели.

На рис. 8 представлена гистограмма остатков. Отличие от нормальности распределения, возможно, обусловлено влиянием вышеописанных факторов, поскольку учет всех факторов в модели принципиально невозможен.

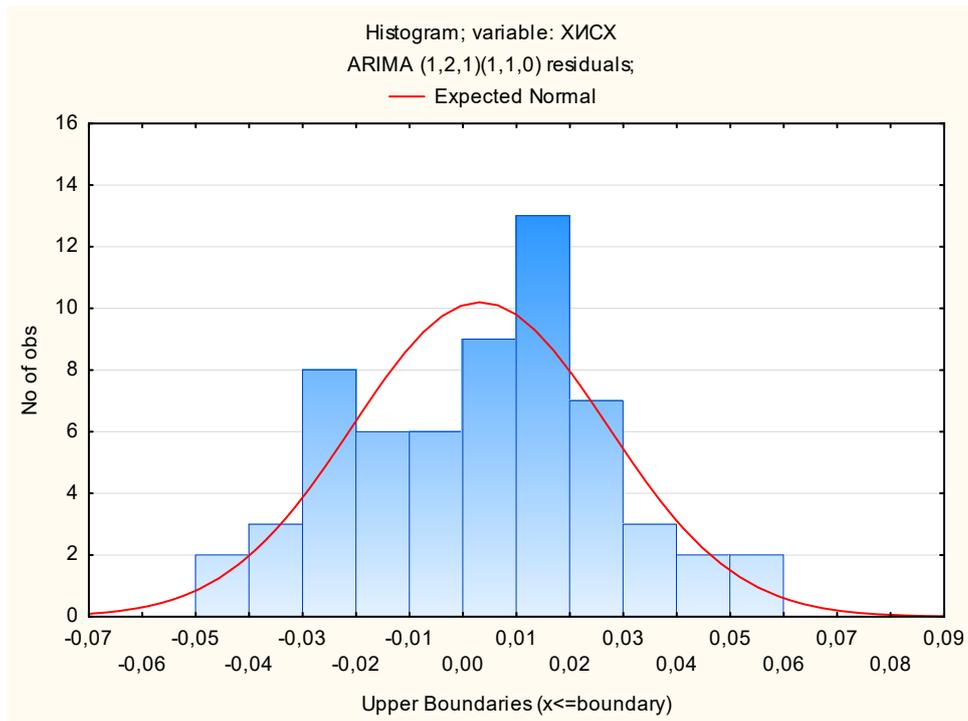


Рисунок 8

На рис. 9, 10 представлены автокорреляционная и частная автокорреляционная функции остатков соответственно. Их поведение позволяет прийти к выводу, что построенная модель АРПСС адекватна изучаемому процессу и может быть применена к построению прогноза.

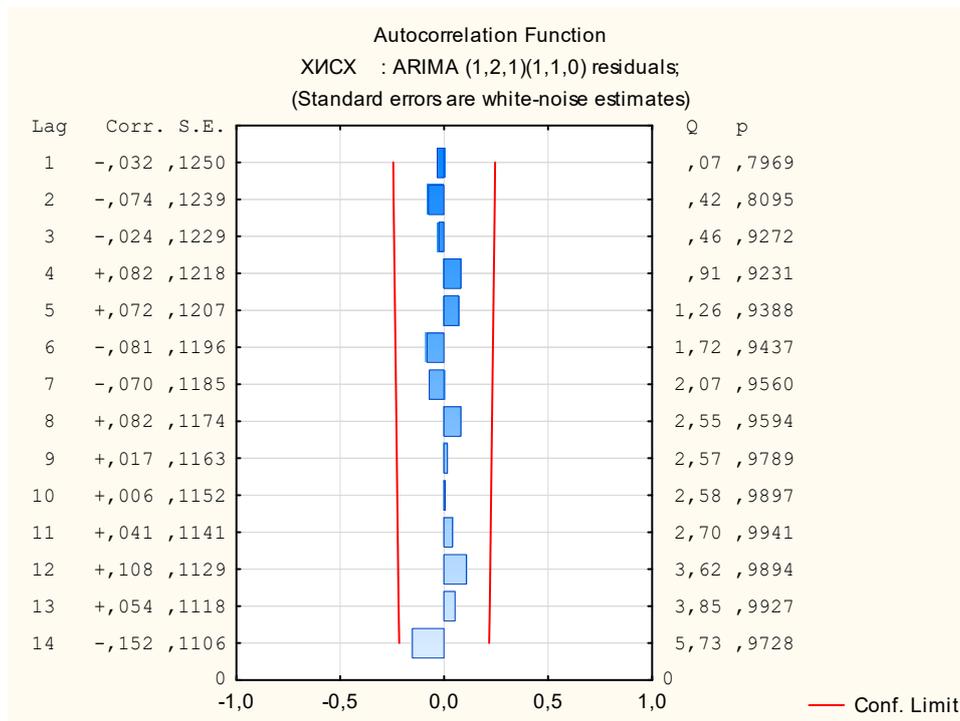


Рисунок 9

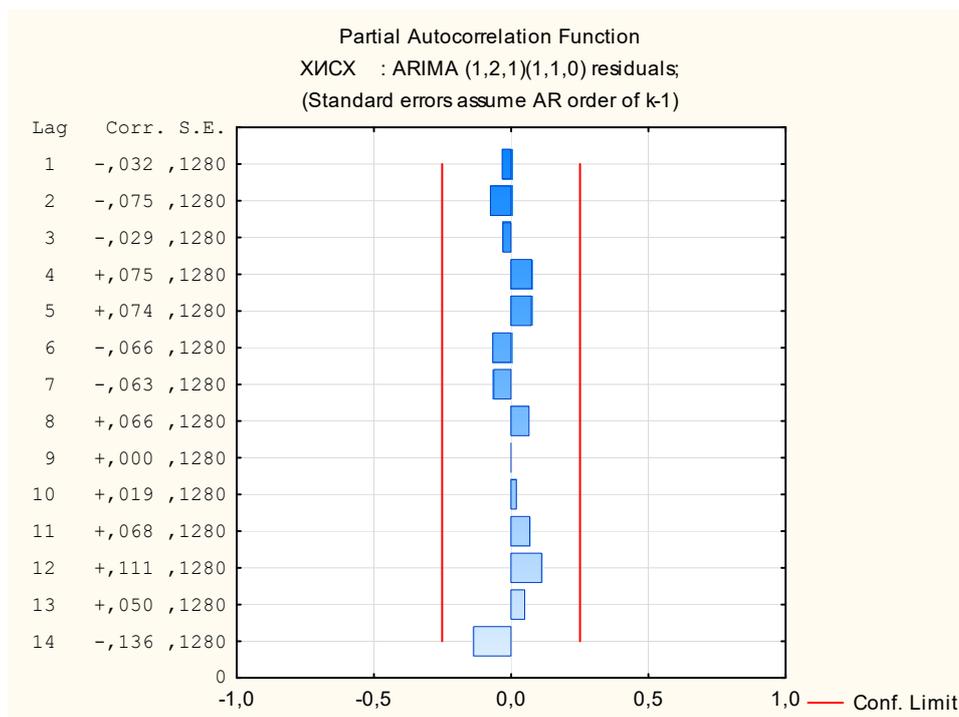


Рисунок 10

Построение прогноза

На рис. 11 представлен исходный ряд и прогноз на 2 недели (с 25.12.2020 по 07.01.2021).

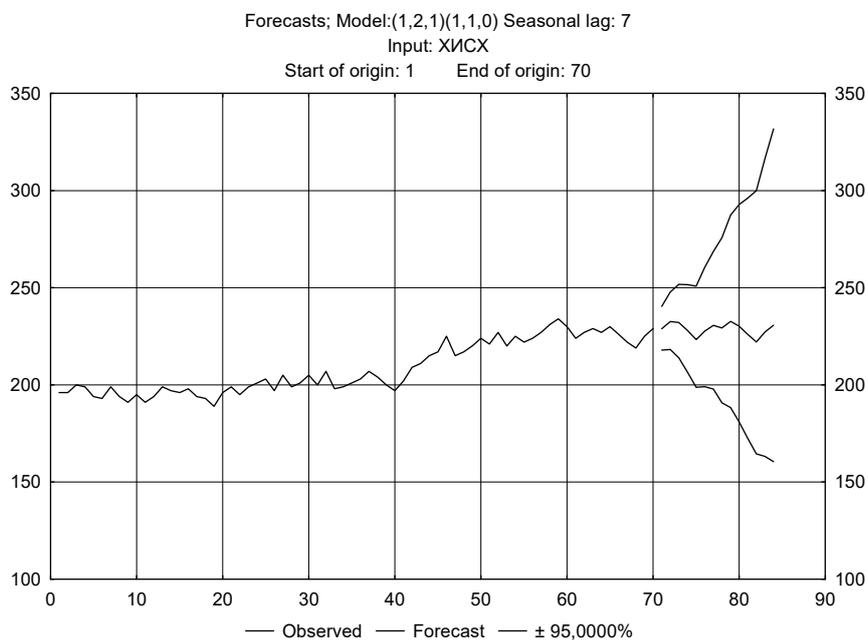


Рисунок 11

На рис. 12 представлены графики реальных значений и прогнозируемых на 2 недели.

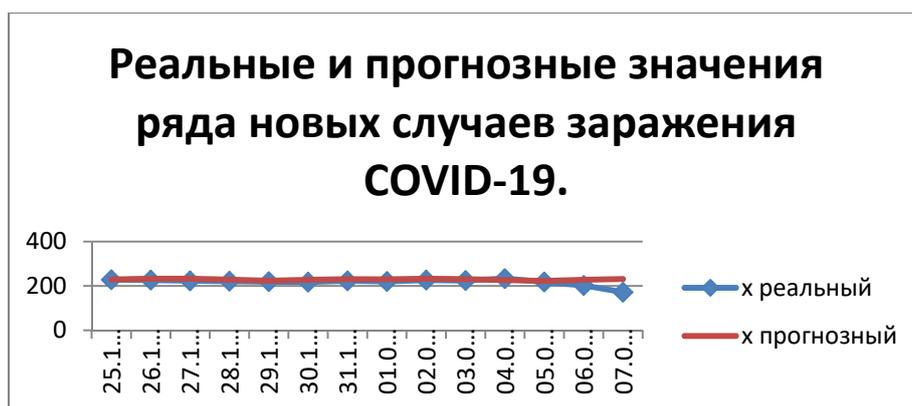


Рисунок 12

Прогнозные данные хорошо согласуются с реальными до 03.01.2021. Начиная с 04.01.2021 наблюдается существенное различие в поведении реальных и прогнозных данных. Скорей всего, это обусловлено началом новогодних праздничных дней, а хорошее согласие до 03.01.2021 обусловлено запаздыванием учета в реальных данных результатов предпраздничного тестирования. Поведение реальных данных позднее объясняется резким уменьшением количества тестируемых во время новогодних праздничных дней.

Выводы

Проведенные исследования показали, что использование модели АРПСС для описания динамики новых случаев заражения COVID-19 и прогнозирование на ее основе допустимы в случае, когда на характер процесса в прогнозируемый период не будет оказано существенное влияние в результате ожидаемых или внезапных административных решений и действий.

Отсутствие информации о последних приводит к невозможности получения правдоподобного прогноза. Весьма вероятно, что это справедливо и для других подходов к моделированию и прогнозированию динамики новых случаев заражения коронавирусом.

Литература

1. Мишулина, О. А. Статистический анализ и обработка временных рядов / О. А. Мишулина. – Москва : МИФИ, 2004. – 180 с. – Текст : непосредственный.
2. Дуброва, Т. А. Статистические методы прогнозирования : учебное пособие для вузов / Т. А. Дуброва. – Москва : Юнити, 2003. – 106 с. – Текст : непосредственный.
3. Коротков, М. Г. Проблемы идентификации интервалов временного ряда при прогнозировании динамики числа заражённых Covid-19 статистическими методами на примере Югры / М. Г. Коротков, А. А. Петров, М. В. Куркина. – Текст : непосредственный // Вестник Югорского государственного университета. – 2020. – № 16 (3). – С. 70–74.
4. Боровиков, В. П. Прогнозирование в системе STATISTICA в среде Windows / В. П. Боровиков, Г. И. Ивченко. – Москва : Финансы и статистика, 1999. – 384 с. – Текст : непосредственный.