

**Самостоятельная работа № 1 (СР1).**  
**«Анализ и прогнозирование временных рядов: ARIMA-модели»**

**Срок сдачи: 23:59 11 апреля 2024 г. (четверг)**

**Оценка: 1-10 баллов**

**Вес СР1 в итоговой оценке = 0,2**

**Примерные критерии оценивания:**

Общее содержание, четкость и последовательность изложения материала	Владение инструментарием: использование всех моделей по заданию и необходимых тестов	Структура и оформление, наличие файла с данными
20%	60%	20%

**ВНИМАНИЕ!** В случае точного совпадения содержания отчета других студентов (с работами предыдущих лет), за СР1 – 0 баллов и итоговая оценка за курс – «неуд»!

**! Отчет загружается в систему SmartLMS.**

**Работа выполняется в группе из 2 человек (по желанию).**

**Построение временного ряда.**

	Ограничения/комментарии
<b>Исходные данные:</b>	Временной ряд ежегодных данных (без сезонной составляющей) для двух стран/регионов и т.д.
<b>Объем выборки:</b>	<b>N&gt;25</b> (Предпочтительно брать более длинные ряды). Возможно взять данные с сезонностью, сезонность <b>сгладить</b> любым способом(описать каким способом сглаживали).
<b>Временной интервал:</b>	без ограничений (можно взять исторические данные за период, например, 1890-1930).
<b>Рекомендованные данные:</b>	Данные Всемирного банка <a href="https://data.worldbank.org/indicator?tab=all">https://data.worldbank.org/indicator?tab=all</a> <b>Обзор баз данных.</b> Цыплаков А. (2009) Где найти данные в сети? // Квантиль, №6, стр. 59–71. <a href="http://quantile.ru/06/06-AT.pdf">http://quantile.ru/06/06-AT.pdf</a>  Можно взять свои данные (без сезонности(!), для данных с сезонностью необходимо подбирать и оценивать SARIMA).

**Примеры возможных показателей (можно и нужно взять что-то свое):**

Показатель	Страны/Регионы	Период
CO2 emissions (metric tons per capita) «Выбросы CO2 (метрические тонны (тонны) на душу населения)	Франция и Италия	1960-2014
Доля экспорта в ВВП (в %)	Малайзия и Таиланд	1960-2019
Children out of school, % of primary school age	EU (European Union) & MENA (Middle East and North Africa)	1970-2019
Население в возрасте 65 лет и старше в % от общей численности населения (в %)	США и Россия	1960-2019
Прирост ИПЦ, %	Финляндия и Норвегия	1960-2019

Суммарный коэффициент рождаемости, кол-во рождений на женщину	США и Россия	1960-2018
Выбросы углекислого газа в атмосферу, килотонн	Финляндия, Израиль	1960-2016
Занятость в сфере услуг (% от общего числа занятых)	США и Россия	1991 - 2020
ВВП на душу населения (ВВП в текущих ценах, US \$)	Великобритания, Франция	1970-2019
Объёмы годовых выбросов углекислого газа (в килотоннах)	Норвегия и Швеция	1960-2016
Площадь сельскохозяйственных земель, %	Сальвадор, США	1961-2016
Число населения в крупнейшем мегаполисе страны, количество человек	Франция и США	1960-2019
Занятость в сфере услуг (% от общего числа занятых)	США и Россия	1991 - 2020

(по материалам работ студентов прошлых лет)

В работе необходимо провести сравнительный анализ по выбранным странам/регионам, выводы структурировать.

**1. Введение.** Во введении опишите проблему Вашего исследования, анализируемый показатель, почему были выбраны именно эти страны/регионы, временной интервал (не обязательно брать новейшие данные, это может быть исторический временной интервал) и пр.

**2. Исходные данные.** Опишите исходные данные в виде таблицы:

анализируемый показатель и его единица измерения:	
анализируемые страны/регионы:	
источник данных (ссылка):	
период наблюдения:	
количество наблюдений:	

- постройте, включите в текст работы и опишите графики временных рядов для выбранных стран (не забудьте подписать график, указать шкалу по оси x и указать единицы измерения (см.примеры оформления)),

- проанализируйте динамику (как изменился анализируемый показатель за исследуемый период, можно использовать показатели динамики (прирост, темп прироста)).

**Пример описания:**

Важным показателем смертности и одновременно качества жизни является коэффициент младенческой смертности - число умерших в возрасте до 1 года в расчете на 1000 родившихся живыми. В отличие от общего коэффициента смертности, то поднимавшегося, то снижавшегося на протяжении двух последних десятилетий, коэффициент младенческой смертности довольно устойчиво снижался (рис). Наблюдавшиеся повышения значения коэффициента младенческой смертности были связаны в основном с улучшением качества учета и постепенным переходом на международный стандарт в определении живорождения - в 1993 году (на 11%)<sup>[12]</sup> и в 2012 году (на 17%)<sup>[13]</sup>. Дополнительное расширение критериев живорождения в 2013 году не привело к повышению показателей младенческой смертности. Число детей, умерших в возрасте до 1 года, быстро снижалось в 1960-е годы за счет сокращения и рождаемости, и смертности, но в 1972-1976 годах стало расти (отчасти за счет улучшения учета родившихся). С конца 1980-х годов число умерших в возрасте до 1 года неуклонно

сокращалось, снизившись с 48,5 тысячи в 1987 году до 13,2 тысячи человек в 2011 году. В 2012 году число зарегистрированных смертей в возрасте до 1 года в связи с расширением критериев живорождения увеличилось до 16,3 тысячи человек, что на 24% больше, чем за 2011 год. Значение коэффициента младенческой смертности возросло до 8,6‰ против 7,4‰ за 2011 год.

<http://www.demoscope.ru/weekly/2018/0761/barom05.php>

- Рассчитайте и опишите дескриптивные статистики. Проведите сравнительный анализ по выбранным странам/регионам.

3.1. **Основные компоненты временного ряда.** Проведите сравнительный анализ по выбранным странам/регионам. Исследуйте и опишите трендовую компоненту (вид тренда) временного ряда, его характеристики: коррелограмму, автокорреляционную и частную автокорреляционную функцию (ACF/PACF). Какой вывод можно сделать? Как можно по характеру ACF сделать предположение о стационарности ряда, о наличии трендовой компоненты? (включайте основные графики в текст, не выносите все в приложение)

3.2. Проверьте гипотезу об отсутствии тренда, укажите какие тесты можно использовать. В тексте работы приведите основные значения статистик. Сделайте вывод.

3.3. Исследуйте структурные сдвиги в данных, с чем (событие в экономике или политике и др.) может быть связано наличие структурного сдвига, укажите какие тесты можно использовать. В тексте работы приведите основные значения статистик, на основании которых делаются выводы. Сделайте вывод.

	Страна 1	Страна 2
Наличие тренда	Но, статистика критерия, р-значение, вывод	Но, статистика критерия, р-значение, вывод
Наличие структурного сдвига	Но, статистика критерия, р-значение, вывод	Но, статистика критерия, р-значение, вывод

4. **Методология Бокса-Дженкинса.** Проведите сравнительный анализ по выбранным странам/регионам, выводы структурируйте.

4.1. **Порядок интегрируемости процесса.** Какие тесты используют для выявления стационарности ряда? Проведите 3 теста единичного корня. В чем их достоинства и недостатки? Какие проблемы возникают? Результаты тестирования на стационарность сведите в сравнительную таблицу по странам/регионам.

В случае нестационарности ряда, приведите его к стационарному виду взятием последовательных разностей. Как определить порядок интегрируемости процесса?

Страна 1				
ВР	Тест	Нулевая гипотеза	Статистика критерия, р-значение	Вывод
у	Тест1(с трендом/без)			
	Тест2 (с трендом/без)			
	Тест3(с трендом/без)			

Δy	Тест1(с трендом/без)			
	Тест2 (с трендом/без)			
	Тест3(с трендом/без)			
Страна 2				
BP	Тест	Нулевая гипотеза	Статистика критерия, р-значение	Вывод
y	Тест1(с трендом/без)			
	Тест2 (с трендом/без)			
	Тест3(с трендом/без)			
Δy	Тест1(с трендом/без)			
	Тест2 (с трендом/без)			
	Тест3(с трендом/без)			

Сделайте общий вывод относительно типа процесса (TS/DS). В приложении приведите скрины тестов.

4.2. На основе анализа ACF и PACF обоснуйте выбор ARIMA-моделей для дальнейшего оценивания. Какие ARIMA-модели существуют, какие используют на практике, как их подбирают?

4.3. Оцените 4-5 предполагаемых моделей ARIMA (обязательно добавьте модели с детерминированным трендом, модель с фиктивной переменной, соответствующей предполагаемому структурному сдвигу). Опишите все модели и статистические свойства оценок, запишите математическую форму моделей (используйте лаговый оператор). Опишите критерии качества моделей, выполнение предпосылок ARIMA-моделей (стационарность, обратимость). Рассчитайте и сравните значения информационных критериев.

**Замечание.** При построении ARIMA-модели с трендом, исследуйте трендовую составляющую. Оцените параметры 2 трендовых моделей (линейную и параболическую, например).

4.4. **Анализ остатков.** Проведите анализ остатков и выберите наиболее адекватную модель. Ответ обоснуйте. Результаты сравнения приведите в виде сводной таблицы.

Страна 1				
модели ARIMA	Стационарность, обратимость модели	Ошибка модели, инф.критерии	Анализ остатков (автокорреляция, нормальность)	Наилучшая модель
1. ARIMA				
2. ARIMA				
3. ARIMA				
4. ARIMA+детермин.тренд				
5. ARIMA+фиктивные переменные				

<b>Страна 2</b>				
<b>модели ARIMA</b>	<b>Стационарность, обратимость модели</b>	<b>Ошибка модели, инф.критерии</b>	<b>Анализ остатков (автокорреляция, нормальность)</b>	<b>Наилучшая модель</b>
1. ARIMA				
2. ARIMA				
3. ARIMA				
4. ARIMA+детермин.тренд				
5. ARIMA+фиктивные переменные				

**4.5. Прогноз.** Рассчитайте точечную и интервальную оценку прогноза на 1-6 шагов вперед (в стат пакете). Постройте совмещенные графики в одной системе координат: исходные данные и предсказанные значения по модели, добавьте точечную оценку прогноза и 95%доверительный интервал. Опишите и сравните, как будут различаться прогнозы показателя в выбранных Вами странах. Насколько изменятся показатели по странам в будущем?

Страна1	Страна 2

**4.6. Качество прогноза.** Оцените качество прогноза, приведите соответствующие 3-4 характеристики качества прогноза, сделайте вывод относительно полученных характеристик.

Страна1	Страна 2

**5. Заключение.** В заключении подведите итог Вашего мини-исследования, результаты моделирования, сравните результаты по странам. Укажите, с какими трудностями Вы столкнулись при подборе и оценивании ARIMA-моделей.

#### *Требования к оформлению отчета*

Работа сдается в эл.виде преподавателю через SmartLMS вместе с файлом исходных данных и расчетами.

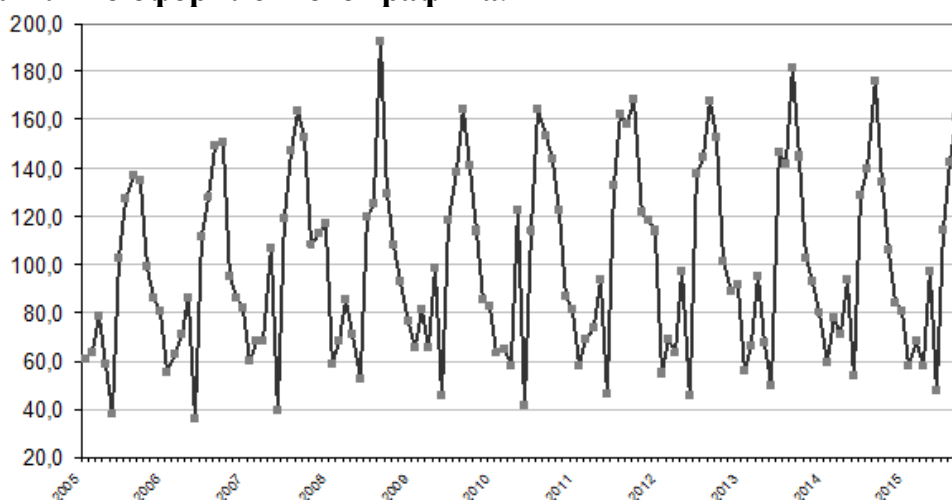
Отчет должен содержать следующие разделы:

- Титульный лист
- Содержание,
- Введение: постановка задачи (актуальность и пр.)
- Главы (если необходимо),
- Заключение,
- список литературы,
- приложения.
- В тексте отчета допустимы только грамотно оформленные модели в виде формул и **форматируемых** таблиц! (см оформление статей в реферируемых журналах, например,

«Прикладная эконометрика»). Все **скрины** *вспомогательных* расчетов (оценивание моделей, результаты тестов) разместите и оформите в приложениях.

- *Оформление тестов.* В каждом тесте необходимо описать содержательную идею применения, нулевую гипотезу, выводы.
- Текст отформатировать. Формат **PDF** (другие форматы при выгрузке искажаются или не читаются)!
- Страницы самостоятельной работы должны иметь сквозную нумерацию. Первой страницей является титульный лист, на котором номер страницы не проставляется.
- Формулы располагают отдельными строками в центре листа или внутри текстовых строк.
- Таблицы и рисунки должны иметь названия и порядковую нумерацию (например, табл. 1, рис. 3). Нумерация таблиц и рисунков должна быть сквозной для всего текста домашней работы. Порядковый номер таблицы проставляется в правом верхнем углу над ее названием. В каждой таблице следует указывать единицы измерения показателей и период времени, к которому относятся данные.
- Порядковый номер рисунка и его название проставляются под рисунком. При построении графиков по осям координат вводятся соответствующие показатели, буквенные обозначения которых выносятся на концы координатных осей, фиксируемые стрелками. При необходимости вдоль координатных осей делаются поясняющие надписи.
- Таблицы, полученные в статистических пакетах («скрины»), размещать только в приложении. Все основные результаты и выводы по таблицам описывать в работе.
- Отчет загружается в систему **SmartLMS**.
- К отчету необходимо приложить **файл с данными!**

### Пример правильно оформленного графика.



**Рис.1. Число зарегистрированных браков (тыс.)  
по месяцам регистрации (январь 2005 – ноябрь 2015)**

Источник: данные Росстата [[www.gks.ru](http://www.gks.ru)]

### *Примеры плохо оформленных графиков*

(по материалам работ студентов прошлых лет)

**Пример 1. Недостаток:** ось У- какая шкала?; Ось х – шкала (годы, месяцы)?

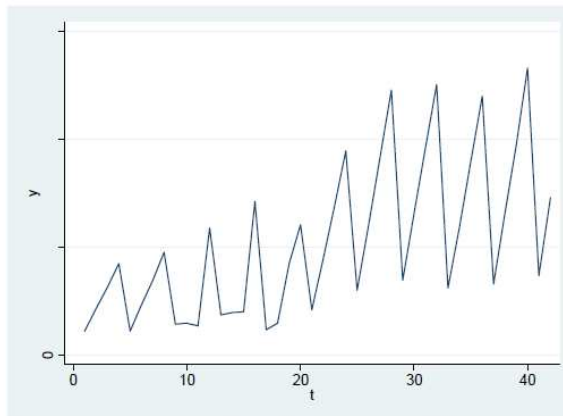
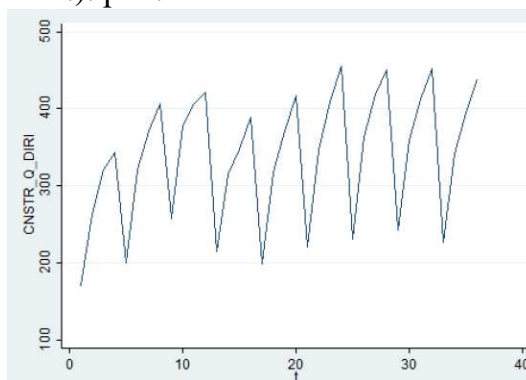


Рисунок 1. График показателя «выручка» ПАО «Нижнекамскнефтехим», тыс. руб., 2005 – 2015 гг.

**Пример 2.** Рассмотрим динамику показателя объёмов завершённого строительства (поквартально, за 2006-2014 гг.). рис. 1



**Недостаток:** Нет подписи рисунка, ось У- единицы измерения?; Ось х – шкала?

**Пример 3.** Исходные данные представлены динамикой ежемесячных продаж домов за период с 1 января 2008 года по 1 августа 2015 года. Данные взяты с сайта Федерального Резервного Банка Сент-Луиса: <https://research.stlouisfed.org/fred2/series/EXHOSLUSM495N>

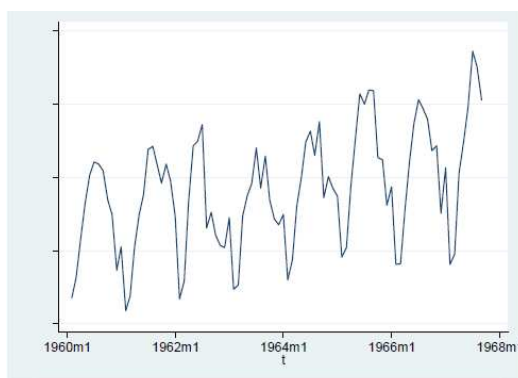


Рисунок 1. Динамика ежемесячной продажи домов, 2008-2015

**Недостаток:** ось У- какая шкала и единица измерения?