

## ***Лабораторная работа.***

### **Анализ временных рядов в среде R. Метод последовательной идентификации составляющих ВР**

#### ***1. Цель лабораторной работы***

- изучить методы и алгоритмы прогнозирования временных рядов на примере решения конкретной задачи ИАД;
- исследовать эффективность использования различных методов прогнозирования временных рядов для решения прикладной задачи;
- ознакомиться и получить практические навыки работы с языком R для решения задач исследования и прогнозирования временных рядов.

#### ***2. Задание к лабораторной работе***

Прочитайте содержательную постановку задачи ИАД для вашего варианта. В файле “л.р.2.csv” представлены исходные данные для решения поставленной задачи.

2.1. Загрузите исходные данные для проведения интеллектуального анализа в R.

2.2. Постройте график временного ряда. Выполните декомпозицию временного ряда на трендовую, сезонную и шумовую составляющие. Постройте автокорреляционную и частную автокорреляционную функции ВР. На основе визуального анализа графиков сделайте предварительные выводы о структуре временного ряда:

- наличие тренда; характер основной тенденции (монотонность; существование вертикальных и/или горизонтальных асимптот; рост (спад) уровней ряда с течением времени); тип функции тренда (линейная, нелинейная);
- наличие сезонной составляющей и характер сезонной составляющей (периодичность; амплитуда колебаний; постоянство (изменчивость) амплитуды колебаний с течением времени).

2.3. Выполните исследование и модельное описание временного ряда на основе метода последовательной идентификации:

2.3.1. Идентификация тренда:

- определите структурную модель тренда временного ряда (в табл. 1 приведено описание основных моделей тренда, заданных в вариантах);
- выполните параметрическую идентификацию выбранной модели тренда – Модель 1;

- оцените статистические характеристики точности модели (характеристики остатков), результаты занесите в табл. 2 (Модель 1);

- постройте график исходного временного ряда с наложением Модели 1;

- постройте периодограмму остатков для Модели 1, автокорреляционную и частную автокорреляционную функции остатков для Модели 1. Сделайте выводы о структуре остатков.

2.3.2. Идентификация сезонной составляющей временного ряда:

- определите структурную модель сезонной составляющей временного ряда;

- выполните параметрическую идентификацию модели (тренд+сезонная составляющая) – Модель 2;

- оцените статистические характеристики точности модели (характеристики остатков), результаты занесите в табл. 2 (Модель 2);

- постройте график исходного временного ряда с наложением Модели 2;

- постройте периодограмму остатков для Модели 2, автокорреляционную и частную автокорреляционную функции остатков для Модели 2. Сделайте выводы о структуре остатков и адекватности модели.

2.3.3. Идентификация авторегрессионной составляющей временного ряда

- определите структурную модель авторегрессионной составляющей временного ряда (авторегрессионная составляющая может не выделяться во временном ряду после выделения тренда и сезонной составляющей, в этом случае пункт 2.3.3 не выполняется);

- выполните параметрическую идентификацию модели (тренд+сезонная составляющая+авторегрессионная составляющая) – Модель 3;

- оцените статистические характеристики точности модели (характеристики остатков), результаты занесите в табл. 2 (Модель 3);

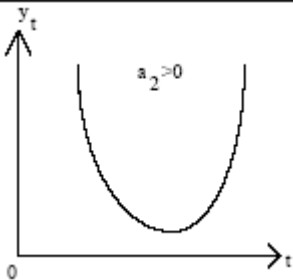
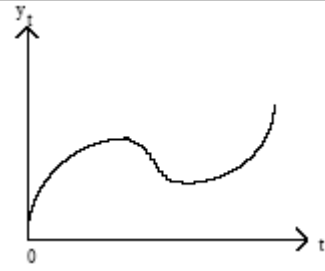
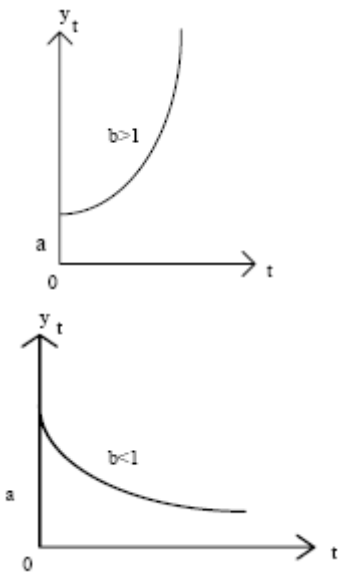
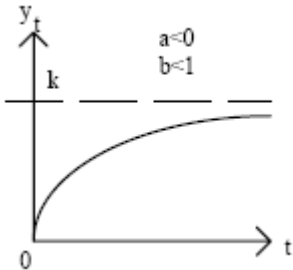
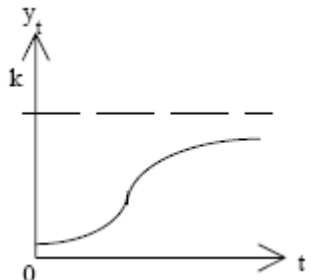
- постройте график исходного временного ряда с наложением Модели 3;

- постройте периодограмму остатков для Модели 3, автокорреляционную и частную автокорреляционную функции остатков для Модели 3. Сделайте выводы о структуре остатков и адекватности Модели 3.

2.4. Сравните модели по точности. Выполните прогноз значений ВР на 3 шага вперед на основе полученной итоговой модели.

Таблица 1 – Типы моделей тренда

	Модель тренда	Название модели	График функции
1	$y(t) = a_0 + a_1 t$	Полином первой степени (линейная)	

2	$y(t) = a_0 + a_1 t + a_2 t^2$	Полином второй степени	
3	$y(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$	Полином третьей степени	
4	$y(t) = ab^t$	Экспонента	
5	$y(t) = k + ab^t$	Модифицированная экспонента	
6	$y(t) = \frac{k}{1 + ae^{bt}}$	Логистическая кривая	

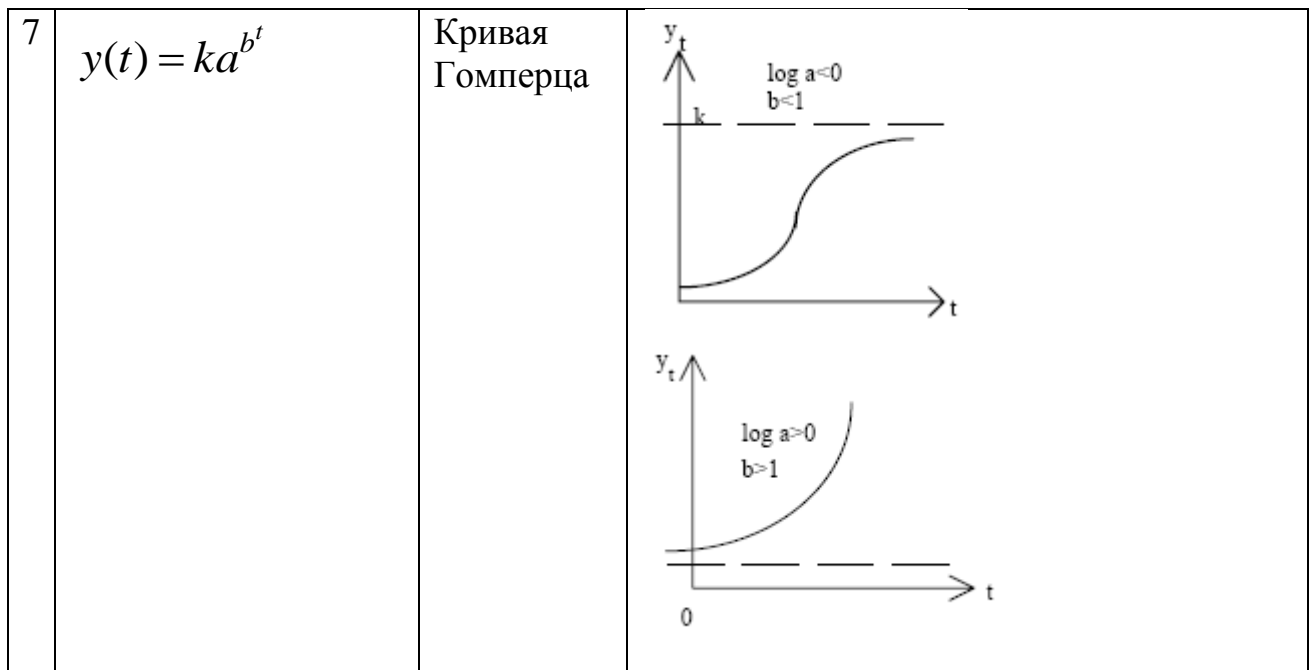


Таблица 2 – Характеристики точности прогнозных моделей

	Модель 1	Модель 2	Модель 3
1. Прогнозная модель			
2. Минимальный остаток (Min error)			
3. Максимальный остаток (Max error)			
4. Средняя ошибка (Mean error)			
5. СКО ошибки (Std. dev.)			
6. Средняя абсолютная ошибка (Mean absolute error)			

7. Средняя ошибка в процентах (Mean percentage error)			
8. Средняя абсолютная ошибка в процентах (Mean absolute percentage error)			
9. Средний квадрат ошибки (Root mean squared error)			
10. Коэффициент детерминации			

2.5. По результатам проведенного исследования сделайте выводы в свободной форме.

### ***3. Методические указания к лабораторной работе***

#### **Пояснение к пп. 2.1-2.2**

Для исследования временного ряда установите пакеты:

```
>install.packages("tseries")
>install.packages("forecast")
>library(tseries)
>library(forecast)
```

Для загрузки данных из CSV-файла с заголовками используется команда:

```
>data <- read.table("C:\\Users\\Desktop\\л.п.2.csv", header=TRUE, sep=";")
```

# read.table - это функция (процедура).  
# В круглых скобках - аргументы функции.  
# ("C:\\Users\\Desktop\\л.р.2-2019.csv" - путь к файлу с данными  
# header= указываем, есть в первой строке названия переменных или нет  
# sep= указываем, какой символ отделяет поля  
# row.names= столбец, в котором хранятся названия строк (параметр не обязателен).

Чтобы взять только необходимый столбец (временной ряд, соответствующий варианту) из данных воспользуйтесь механизмом Slice.

Пример использования:

```
> tsData <- data[,y] , где y - номер столбца.
```

Далее нужно преобразовать массив во временной ряд. Это можно сделать с помощью функции ts(x), где x - массив данных.

Пример:

```
> tsData <- ts(tsData)
```

Если измерения временного ряда сделаны на регулярных интервалах меньших, чем один год, например, по месяцам или кварталам, можно указать количество измерений, сделанных за год, используя параметр frequency функции ts(). Для временных рядов по месяцам: frequency=12, а для данных по кварталам frequency=4. Также можно указать первый год, в который собирались данные, и первый интервал в году с помощью параметра start функции ts(). Например, если первая точка данных соответствует второму кварталу 2000 года, то start=c(2000,2).

Пример:

```
> tsData <- ts(tsData, frequency=4, start=c(2000,2))  
> tsData
```

Для построения графика временного ряда используется функция:

```
> plot.ts(tsData)
```

Чтобы разложить временной ряд на основные компоненты (тренд, сезонность и шум) необходимо воспользоваться функцией decompose.

Описание функции:

**decompose**(x, type = c("additive", "multiplicative"), filter = NULL)

Аргументы:

- x - временной ряд;

- `type` - тип сезонного компонента;
- `filter` - вектор коэффициентов фильтра в обратном порядке времени (как для коэффициентов AR или MA). Используется для фильтрации сезонной составляющей. Если NULL, применяется скользящее среднее с симметричным окном.

Пример использования:

```
>tsDataComponents <- decompose(tsData)
>tsDataComponents
>plot(tsDataComponents)
```

Функция `decompose()` возвращает список объектов в качестве результата, где содержатся оценки периодической составляющей, тренда и нерегулярной компоненты, хранящиеся в именованных элементах этого списка объектов, называемых «seasonal», «trend» и «random» соответственно.

### Пояснение к п. 2.3

Идентификация тренда может быть выполнена на основе регрессионной модели, в которой зависимая переменная – временной ряд, а независимая переменная – отсчеты времени ( $t=1,2,3,\dots$ ).

В R основная функция для подгонки линейной регрессионной модели – это `lm()`. Формат ее применения таков:

```
>myfit <- lm(formula, data),
```

где `formula` описывает вид модели, которую нужно подогнать, а `data` – это таблица с данными, которые используются для создания модели. Полученный объект (`myfit`) – это список, содержащий обширную информацию о подогнанной модели. Формула обычно записывается в таком виде:

$$Y \sim X_1 + X_2 + \dots + X_k,$$

где  $\sim$  отделяет зависимую переменную слева от независимых переменных (разделенных знаками  $+$ ) справа.

В дополнение к `lm()` в табл. 3 перечислены несколько полезных функций для выполнения простого или множественного регрессионного анализа. Каждая из них применяется к объекту, созданному при помощи функции `lm()`, чтобы получить дополнительную информацию о подогнанной модели.

Таблица 3.

Функция	Действие
<code>summary()</code>	Показывает детальную информацию о подогнанной модели
<code>coefficients()</code>	Перечисляет параметры модели (свободный член и регрессионные коэффициенты)
<code>confint()</code>	Вычисляет доверительные интервалы для параметров модели (по умолчанию 95%)
<code>fitted()</code>	Выводит на экран предсказанные значения, согласно подогнанной модели
<code>residuals()</code>	Показывает остатки для подогнанной модели
<code>anova()</code>	Создает таблицу ANOVA (дисперсионного анализа) для подогнанной модели или таблицу ANOVA, сравнивающую две или более моделей
<code>vcov()</code>	Выводит ковариационную матрицу для параметров модели
<code>AIC()</code>	Вычисляет информационный критерий Акаике (Akaike's Information Criterion)
<code>plot()</code>	Создает диагностические диаграммы для оценки адекватности модели
<code>predict()</code>	Использует подогнанную модель для предсказания зависимой переменной для нового набора данных

Пример:

```
>myfit1 <- lm(y~t, data=data1)
```

Здесь y и t – это названия столбцов в таблице data1

```
>plot(myfit)
```

```
>r1<-residuals(myfit1)
```

```
>plot(r1)
```

Для подгонки нелинейных моделей используется функция `nls`:

```
>myfit2<-nls(y ~ 1+a*exp(b*t)),data=data1,start=c(a=0.1,b=0.1))
```

```
>r2<-residuals(myfit2)
```

```
>plot(r2)
```

Для корректной работы функции необходимо задать начальные значения `start= c(...)` для оцениваемых параметров, по умолчанию для оценки параметров модели используется метод Гаусса-Ньютона.

Для расчета точности модели (характеристик остатков) необходимо использовать функцию `accuracy`. Она возвращает суммарные показатели точности прогноза.

```
accuracy(f, x, test = NULL, d = NULL, D = NULL, ...)
```

Аргументы:



- `f` - объект класса “forecast” или числовой вектор, содержащий модельные значения (предсказанные моделью). Функция также будет работать с объектами `Arima`, `ets` и `lm`, если параметр `x` не задан, то рассчитываются характеристики точности набора `f`.
- `x` - необязательный числовой вектор, содержащий фактические значения временного ряда. Если указан параметр `x`, функция рассчитывает характеристики точности модельных остатков на основе значений `x-f`.
- `test` – задает, какие значения временных рядов `x` и `f` будут использоваться при расчете характеристик точности. Если `test` равен `NULL`, используются все элементы `BP`. В противном случае `test` – это числовой вектор, содержащий индексы элементов `BP`, используемых при расчете характеристик точности.
- `d` – целое число, указывающее значение лага, которое будет использоваться в знаменателе при расчете `MASE`. Значение по умолчанию: 1 для несезонных `BP` и 0 для сезонных `BP`.
- `D` - целое число, указывающее количество сезонных разниц, которые будут использоваться в знаменателе при расчете `MASE`. Значение по умолчанию: 0 для несезонных серий и 1 для сезонных.
- ... - дополнительные аргументы, в зависимости от конкретного метода.

Пример использования:

```
>accuracy(myfit1)
> accuracy(data1$y-r2, data1$y)
```

Функция `accuracy` не вычисляет все статистические характеристики, указанные в табл. 2. Для расчета статистических характеристик остатков удобно написать собственную функцию. Основные формулы для расчета характеристик остатков приведены по ссылке: <https://otexts.com/fpp2/accuracy.html>.

Функция `spec.pgram` вычисляет значения периодограммы временного ряда, используя быстрое преобразование Фурье.

```
spec.pgram(x, spans = NULL, kernel, taper = 0.1, fast = TRUE, demean = FALSE, detrend = TRUE, plot = TRUE, na.action = na.fail, ...)
```

Аргументы:

- `x` – временной ряд;
- `spans` – вектор, задающий ширину окна Даниэля для сглаживания периодограммы;

- kernel- тип сглаживания;
- taper – параметр масштаба;
- demean – вычитает среднее значение из элементов ВР, если TRUE;
- detrend – вычитает тренд из элементов ВР, если TRUE;
- plot - Если TRUE - рисует график;
- na.action - функция для обработки пропущенных значений.

Пример использования:

```
>sp <- spec.pgram(tsData,detrend = FALSE, log = "no", fast = FALSE, pad = FALSE, taper = 0, plot = TRUE)
```

Функция Aacf вычисляет и строит график оценки автокорреляционной функции (одномерного, многомерного) временного ряда.

Функция Pacf вычисляет и строит график оценки функции частной автокорреляционной функции (одномерного, многомерного) временного ряда.

Описание функций:

**Aacf**(x, lag.max = NULL, type = c("correlation", "covariance", "partial"), plot = TRUE, na.action = na.contiguous, demean = TRUE, ...)

Аргументы:

- x – временной ряд;
- lag.max – максимальная задержка, при которой рассчитывается автокорреляция, по умолчанию:  $10 * \log_{10}(N / m)$ , где N - количество наблюдений, m – количество рядов;
- type – тип вычисляемой автокорреляции. Допустимые значения: «correlation» (по умолчанию), «covariance» или «partial»;
- plot – если TRUE – отображается график;
- na.action – функция для обработки пропущенных значений. По умолчанию это na.contiguous. Полезными альтернативами являются na.pass и na.interp;
- demean – если TRUE - вычитает среднее значение из элементов ВР.
- ... – дополнительные аргументы для функции plot.

Пример использования:

```
>tsAcf <- Aacf(tsData, lag.max = 15, plot=TRUE, na.action = na.contiguous, demean = TRUE)
```

**Pacf**(x, lag.max = NULL, plot = TRUE, na.action = na.contiguous, demean = TRUE, ...)

Аргументы:

- x – временной ряд;

- `lag.max` – максимальная задержка, при которой рассчитывается частная автокорреляция, по умолчанию:  $10 * \log_{10}(N / m)$ , где  $N$  – количество наблюдений,  $m$  – количество ВР;
- `plot` – если TRUE – отображает график;
- `na.action` – функция для обработки пропущенных значений. По умолчанию это `na.contiguous`. Полезными альтернативами являются `na.pass` и `na.interp`;
- `demean` – если TRUE - вычитает среднее значение из элементов ВР;
- ... - дополнительные аргументы для функции `plot`.

Пример использования:

```
tsPacf <- pacf(tsData, lag.max = 15, plot = TRUE)
```

#### ***4. Отчет о работе***

##### **Содержание**

1. Постановка задачи ИАД
2. Графический анализ временного ряда
3. Построение модели на основе метода последовательной идентификации
4. Код на R, соответствующий выполненному анализу. Заполненная таблица
2. Результаты сравнения моделей по точности, выводы по результатам сравнения. Прогнозирование исходного временного ряда на основе построенной итоговой модели.
5. Выводы по работе в произвольной форме.

#### ***5. Вопросы к работе***

1. Постановка задачи прогнозирования временного ряда, как одной из задач ИАД.
2. Определение временного ряда. Принципиальные отличия временного ряда от случайной выборки.
3. Типы факторов, под воздействием которых формируются значения временного ряда. Структурные составляющие временного ряда.
4. Метод декомпозиции ВР.
5. Как определить структурную модель временного ряда?
6. Методы определения вида трендовой составляющей временного ряда и их использование на примере решаемой задачи: графический метод;

метод характеристик приростов; метод последовательных разностей. Достоинства и недостатки методов.

7. Как определить структуру периодической гармонической функции, описывающей сезонную составляющую временного ряда?
8. Известно, что в формировании значений временного ряда участвуют колебания двух периодов: 12 и 6. Напишите структурную модель периодической гармонической функции.
9. Как определить порядок авторегрессионной составляющей временного ряда? В каких случаях выделяют авторегрессионную составляющую временного ряда?
10. Понятие стационарного/нестационарного временного ряда, автокорреляционных и частных автокорреляционных функций временного ряда, периодограммы.
11. Метод последовательной идентификации составляющих ВР.
14. Характеристики точности прогнозной модели.
15. Как проверить адекватность построенной прогнозной модели данным наблюдения?
16. Сравнительный анализ точности прогнозных моделей.
17. Средства языка R для анализа и прогнозирования временных рядов.

### *Литература*

1. <https://otexts.com/fpp2/accuracy.html> - Прогнозирование временных рядов в R.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. – М.: «Финансы и статистика», 1983. – 471 с.
3. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. – М.: ЮНИТИ, 1998. – 1022 с.
4. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, вып. 1, 1974. – 406 с.; вып. 2 – 197 с.
5. Дж. Д. Лонг и Пол Титор Книга рецептов: Проверенные рецепты для статистики, анализа и визуализации данных / пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2020. – 510 с.
6. Кабаков Р.И. R в действии. Анализ и визуализация данных в программе R. – М.: ДМК Пресс, 2014. – 588 с.
7. Кендэл М. Временные ряды. – М.: Финансы и статистика, 1981. – 199 с.
8. Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. – М.: Статистика, 1973. – 103 с.
9. Мاستицкий С. Э. (2020) Анализ временных рядов с помощью R. — Электронная книга, адрес доступа: <https://ranalytics.github.io/tsa-with-r>

10. Мاستицкий С.Э., Шитиков В.К. (2015) Статистический анализ и визуализация данных с помощью R. – М.: ДМК Пресс. – 496 с.
11. Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1977. – 199с.
12. Шитиков В. К., Мастицкий С. Э. (2017) Классификация, регрессия, алгоритмы Data Mining с использованием R. - Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>

## **Постановка задачи ИАД (данные в таблице Excel)**

### ***Вариант 1-3, 13-15, 25-27, 37-39.***

Представлен временной ряд объема производства промышленного предприятия (в млн. руб.) по месяцам с 2007 по 2022 гг. Необходимо построить прогнозную модель на основе имеющихся данных и осуществить прогнозирование объема производства на январь-март 2023.

### ***Вариант 4-6, 16-18, 28-30, 40-42.***

Представлен временной ряд средней заработной платы программистов крупной организации (в тыс. руб.) по месяцам с 2007 по 2022 гг. Необходимо построить прогнозную модель на основе имеющихся данных и осуществить прогнозирование средней заработной платы на январь-март 2023.

### ***Вариант 7-9, 19-21, 31-33, 43-45.***

Представлен временной ряд прибыли, полученной промышленным предприятием (в млн. руб.) по месяцам с 2007 по 2022 гг. Необходимо построить прогнозную модель на основе имеющихся данных и осуществить прогнозирование прибыли предприятия на январь-март 2023.

### ***Вариант 10-12, 22-24, 34-36, 46-48.***

Представлен временной ряд выработки электроэнергии крупной электростанции (в млн. кВт. Ч) по месяцам с 2007 по 2022 гг. Необходимо построить прогнозную модель на основе имеющихся данных и осуществить прогнозирование выработки электроэнергии на январь-март 2023.