ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

КАФЕДРА ВТ

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

«Анализ временных рядов в среде R. Метод последовательной идентификации составляющих ВР» по дисциплине «Компьютерные технологии анализа и обработки данных»

Выполнили: студенты

гр. АММ2-24

Атласюк Игорь Романович

Ириков Евгений Алексеевич

Проверил: к.т.н., доцент Кафедры ВТ Альсова Ольга Константиновна

Новосибирск 2024

## Содержание

[Постановка задачи 3](#_bookmark0)

[Ход работы 4](#_bookmark1)

[Заключение 23](#_bookmark2)

[Приложение 24](#_bookmark3)

# **Постановка задачи**

Изучить методы и алгоритмы прогнозирования временных рядов на примере решения конкретной задачи ИАД. Исследовать эффективность прогнозирования временных рядов для решения прикладной задачи. Ознакомиться и получить практические навыки работы с языком R для решения задач исследования и прогнозирования временных рядов.

# **Ход работы**

1. **Загрузка данных и построение графика временного ряда и его декомпозиций (рис.1)**

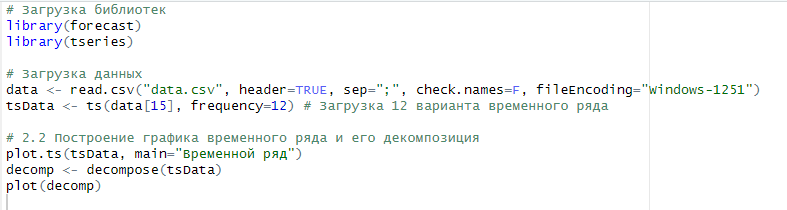


Рисунок 1. Начало работы

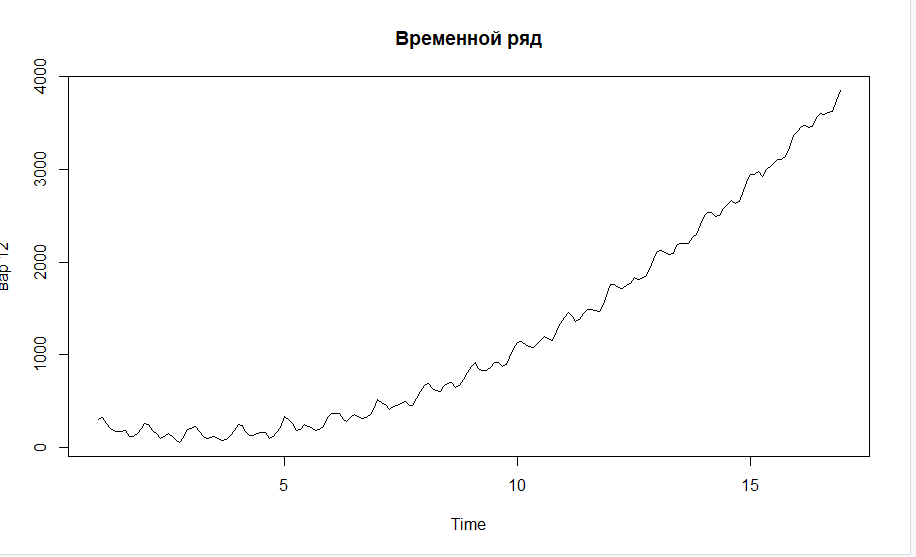


Рисунок 2. Временной ряд

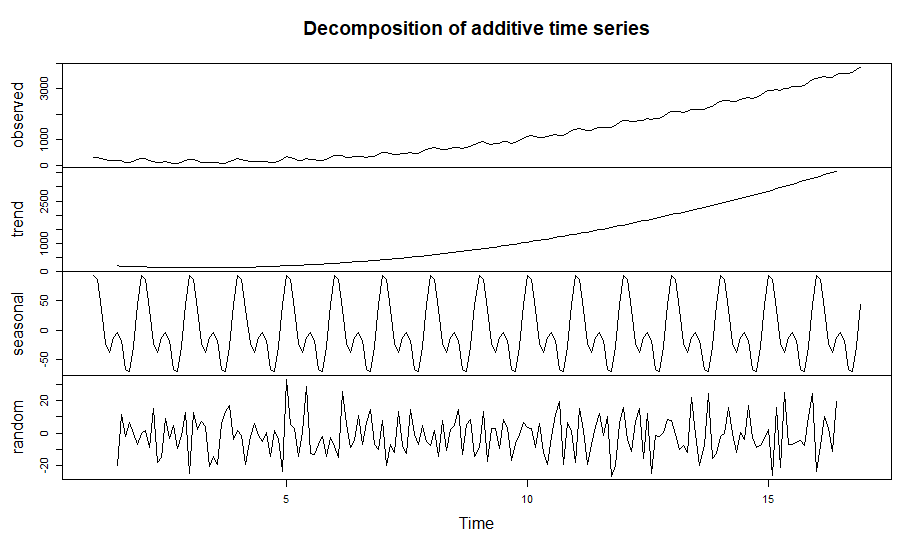


Рисунок 3. Декомпозиция временного ряда

На рисунке 3 изображен график декомпозиции временного ряда. На основании графика можно сделать следующие выводы:

* Трендовая компонента показывает возрастающий тренд. Это свидетельствует о том, что уровни временного ряда растут с течением времени. Тренд, является нелинейным, поскольку график имеет плавное, слегка ускоряющееся повышение.
* Сезонная компонента на графике представлена периодическими колебаниями с четкой периодичностью. Амплитуда колебаний примерно постоянна, что говорит о стабильности сезонной составляющей.
* Случайная компонента (шум) характеризуется небольшими, непостоянными колебаниями, которые не показывают четкой структуры. Это типично для шумовой компоненты, которая включает непредсказуемые отклонения, не объясняемые трендом или сезонностью.
* На графике тренда не наблюдаются горизонтальные или вертикальные асимптоты. Это говорит о том, что в текущем временном интервале тренд растет без замедления.

1. **Построение автокорреляционной и частной автокорреляционной функции временного ряда (рис.4).**



Рисунок 4. Автокорреляционные функции

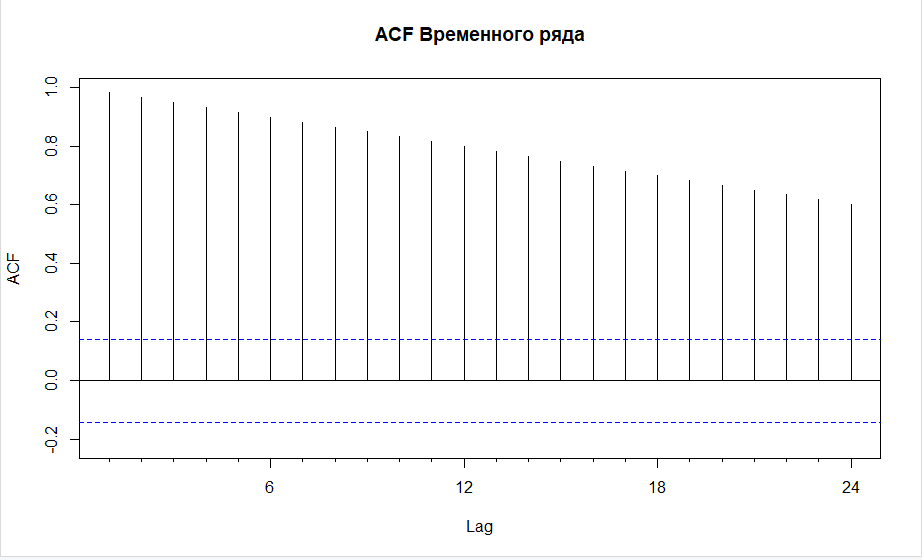


Рисунок 5. Автокорреляционная функция

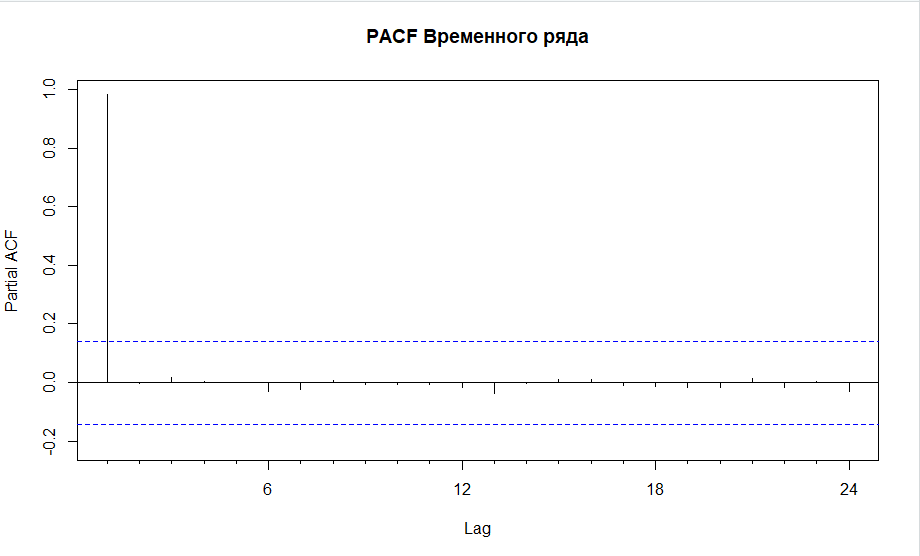


Рисунок 6. Частная автокорреляционная функция

1. **Исследование и модельное описание временного ряда на основе метода последовательной идентификации: Структурная модель тренда временного ряда: полином первой степени (линейная)**

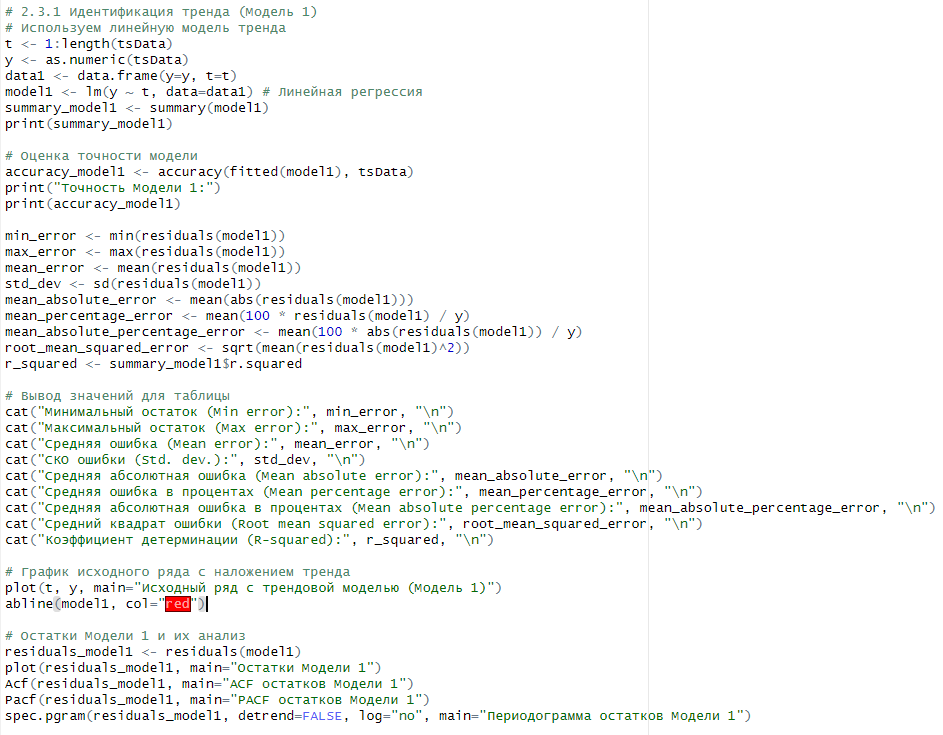


Рисунок 7. Идентификация тренда

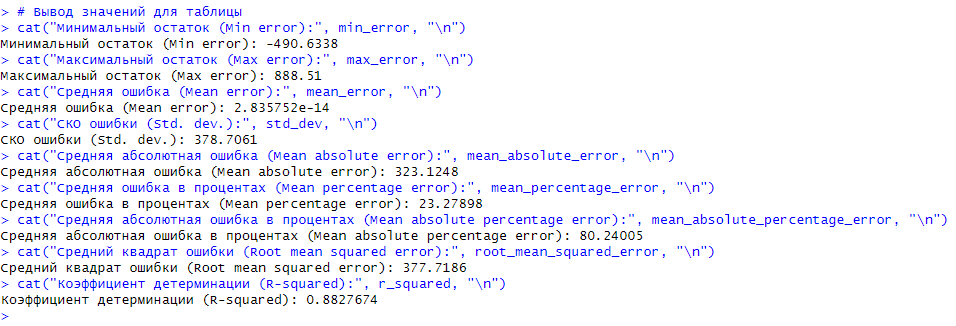


Рисунок 8. Значения для таблицы

Рисунок 9. График исходного ряда с наложением тренда

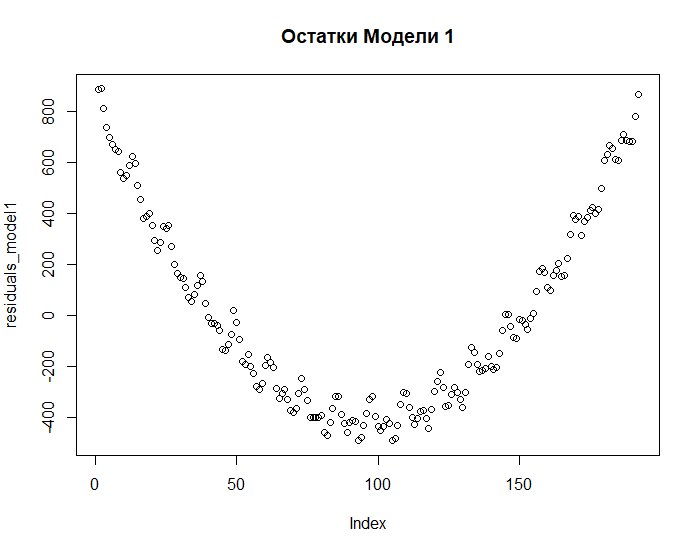


Рисунок 10. Остатки модели

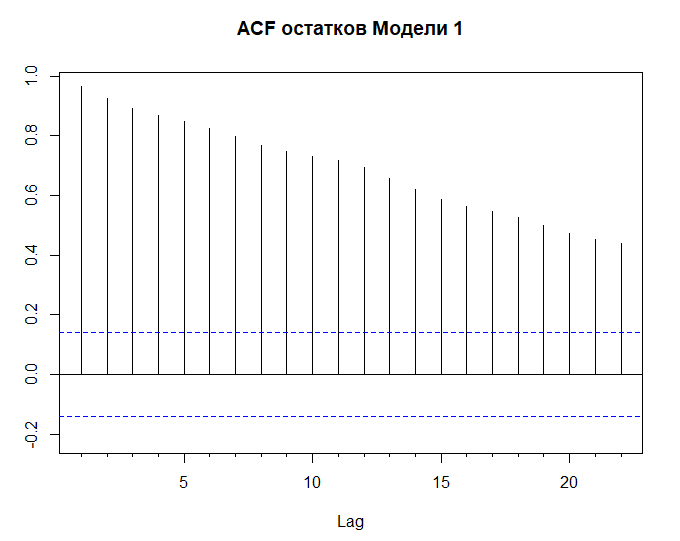


Рисунок 11. ACF остатков модели

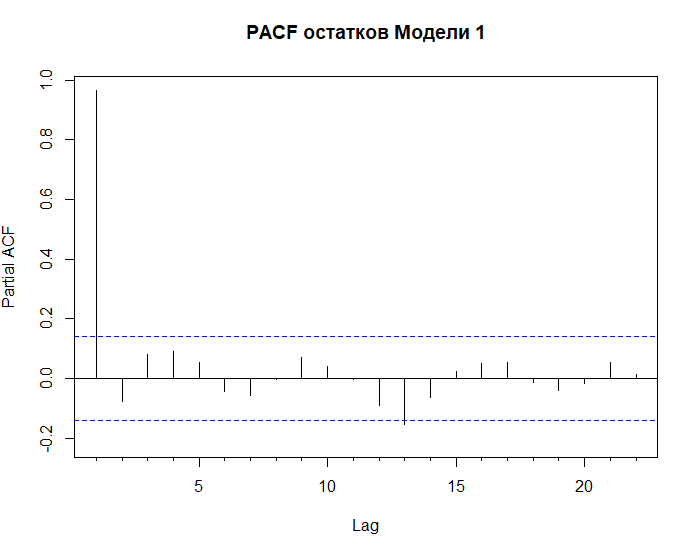


Рисунок 12. PACF остатков модели

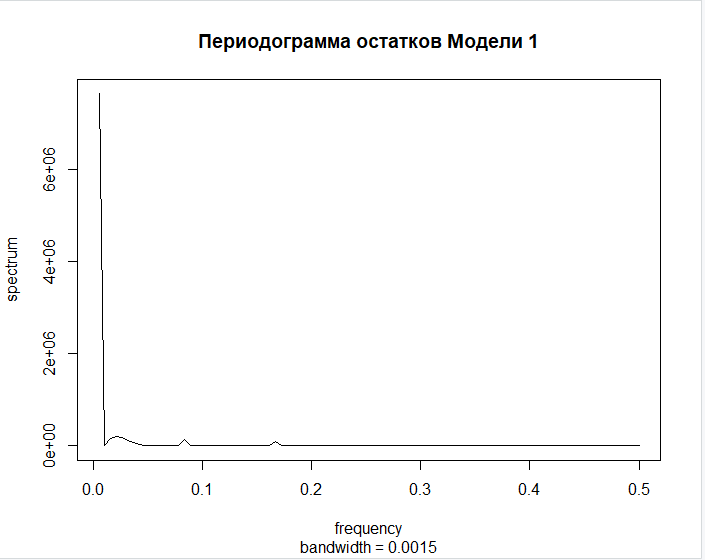


Рисунок 13. Периодограмма остатков модели

Линейная модель тренда в целом хорошо справляется с описанием данных, так как в остатках не наблюдается выраженных сезонных или циклических структур. Однако низкочастотные компоненты могут указывать на более сложные долгосрочные тенденции, которые не учтены линейной моделью.

1. **Идентификация сезонной составляющей временного ряда**



Рисунок 14. Идентификация сезонной составляющей

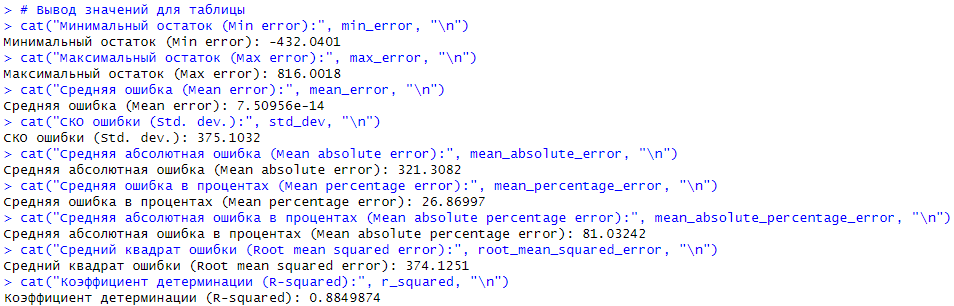


Рисунок 15. Значения для таблицы

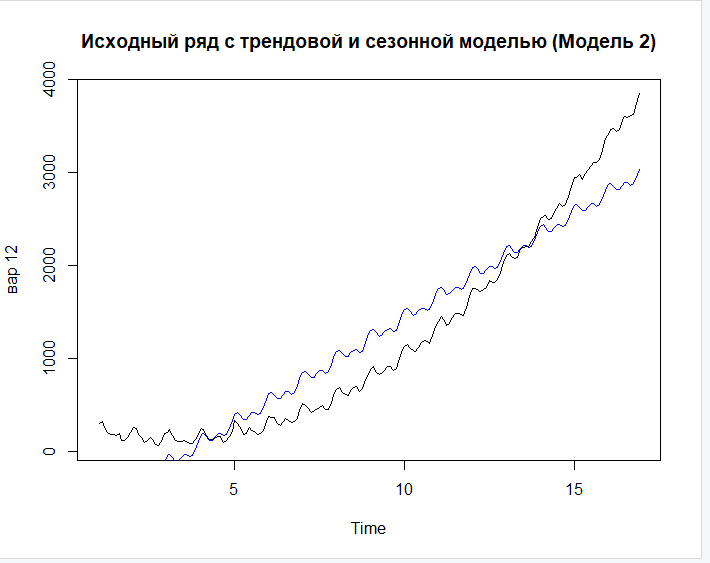


Рисунок 16. График исходного ряда с наложением тренда и сезонной составляющей

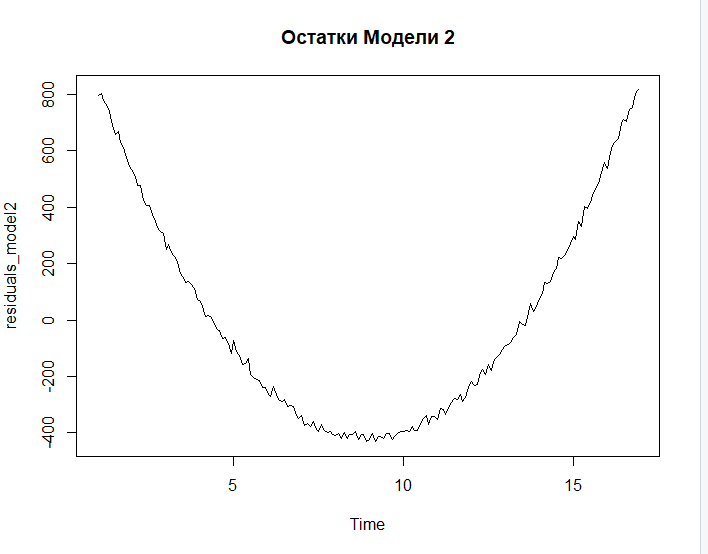


Рисунок 17. Остатки модели

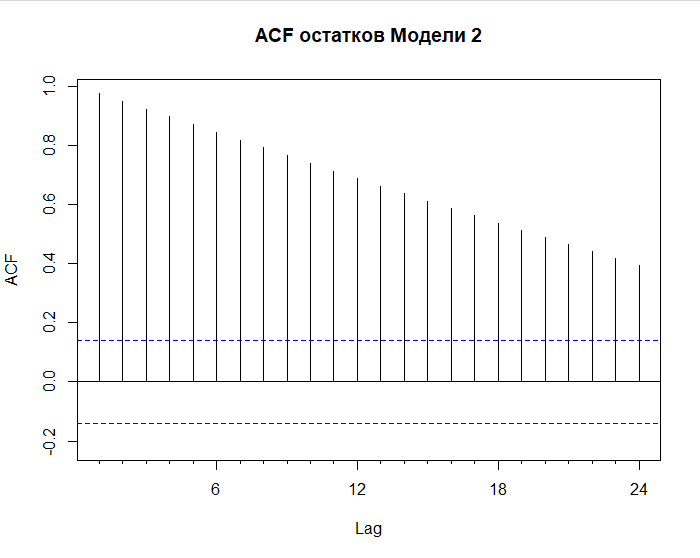


Рисунок 18. ACF остатки модели

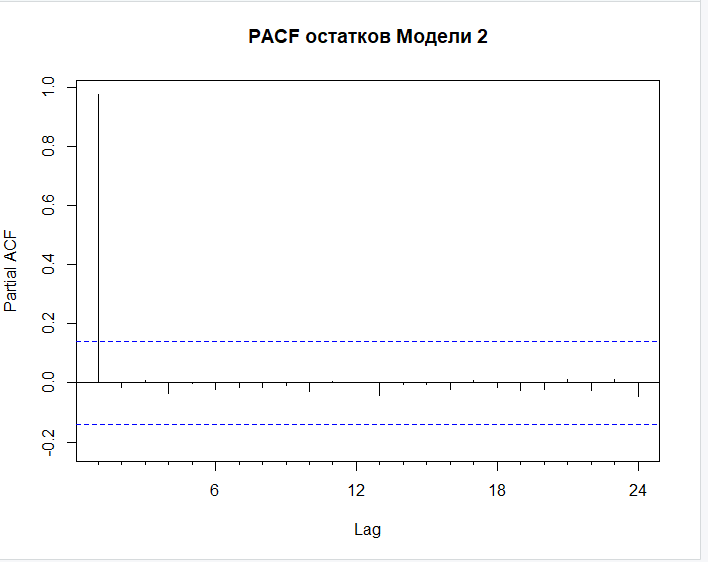


Рисунок 19. PACF остатки модели

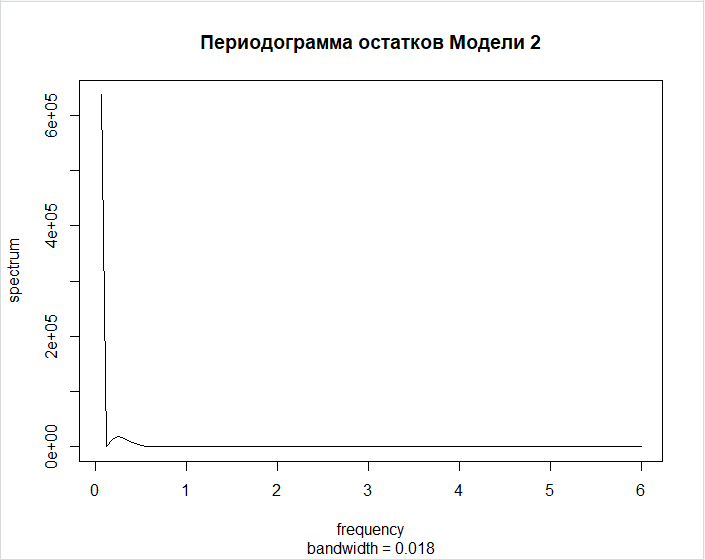


Рисунок 20. Периодограмма модели

По анализу периодограммы остатков для модели 2 можно сделать следующие выводы о структуре остатков и адекватности модели:

* На периодограмме практически отсутствуют выраженные пики на частотах, отличных от нуля, что указывает на отсутствие сильных периодических (циклических) составляющих в остатках. Это предполагает, что модель хорошо справилась с выявлением тренда и сезонности.
* На графике спектральная плотность снижается и близка к нулю на высоких частотах. Это также является положительным признаком, так как остатки не содержат значимых высокочастотных колебаний, что часто свидетельствует об отсутствии систематических ошибок или регулярных колебаний, которые модель не учла.

Таким образом, на основании периодограммы можно предположить, что Модель 2 является адекватной для описания тренда и сезонности, так как остатки не содержат значимых частотных составляющих.

1. **Идентификация авторегрессионной составляющей временного ряда**

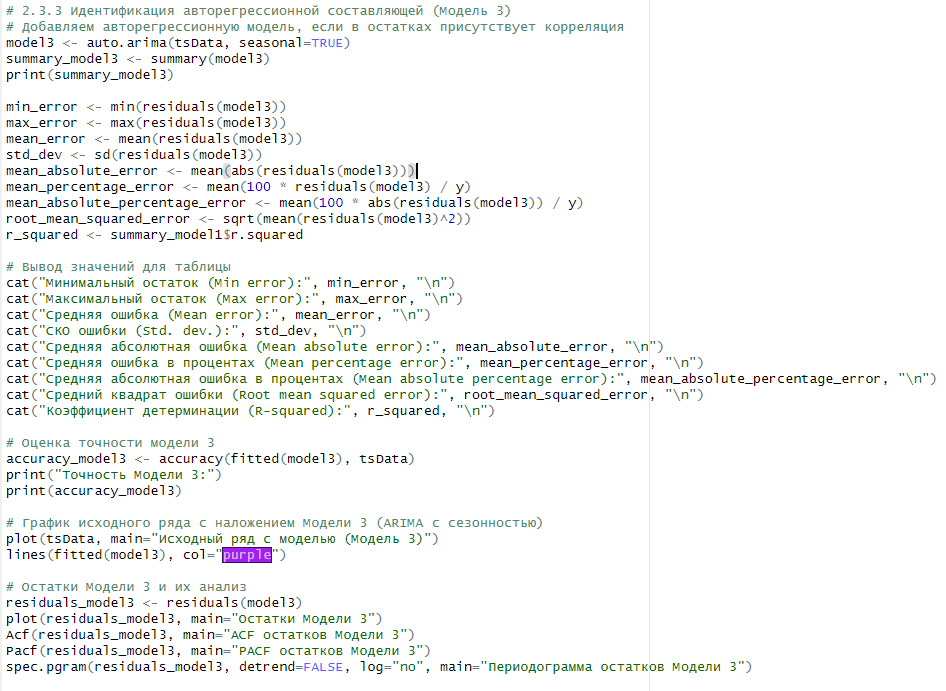


Рисунок 21. Идентификация авторегрессионной составляющей

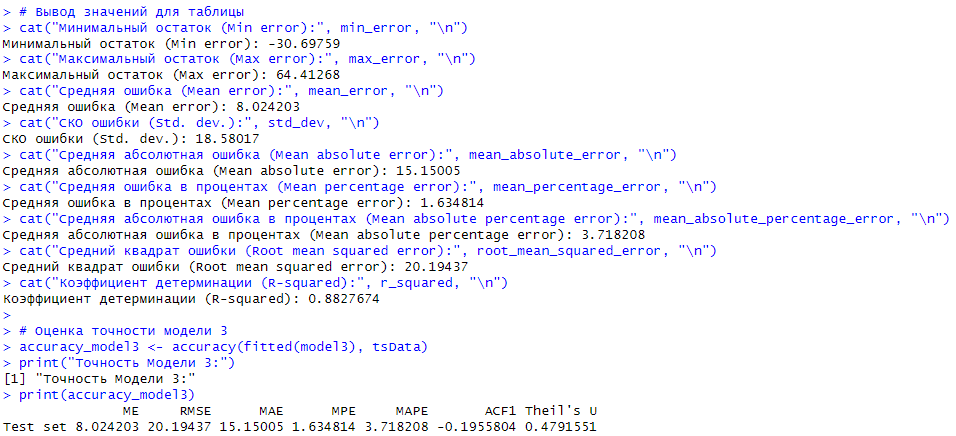


Рисунок 22. Значения для таблицы

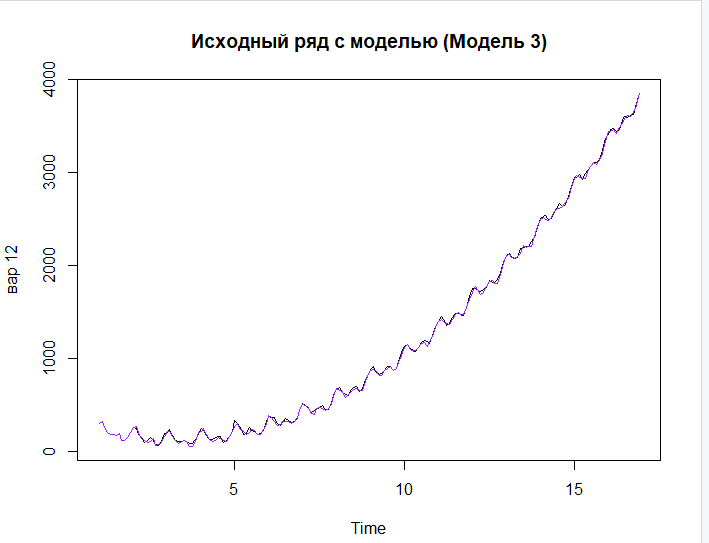
****

Рисунок 23. График исходного ряда с наложением Модели 3 (ARIMA с сезонностью)

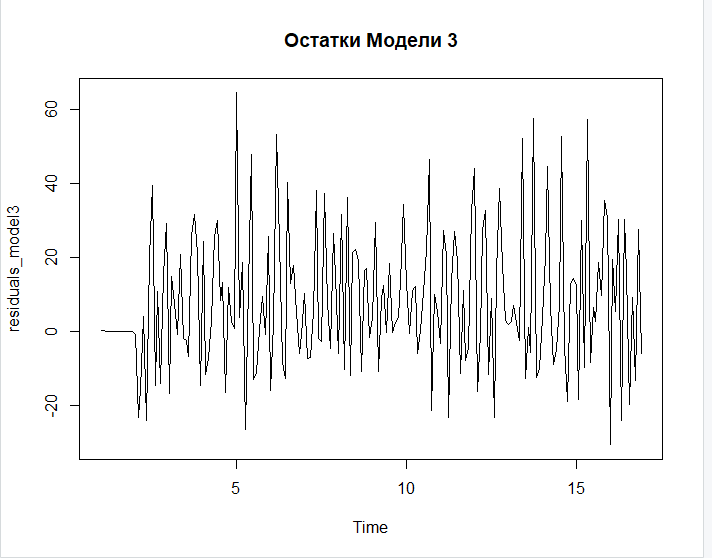


Рисунок 24. Остатки модели

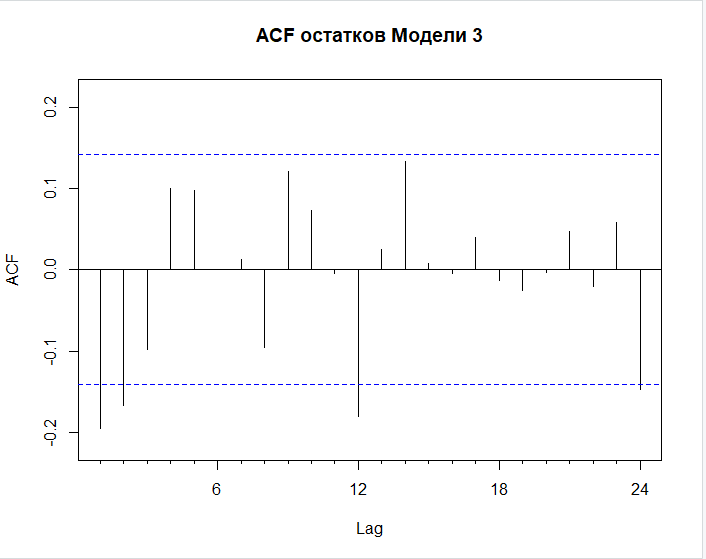
****

Рисунок 24. ACF остатки модели

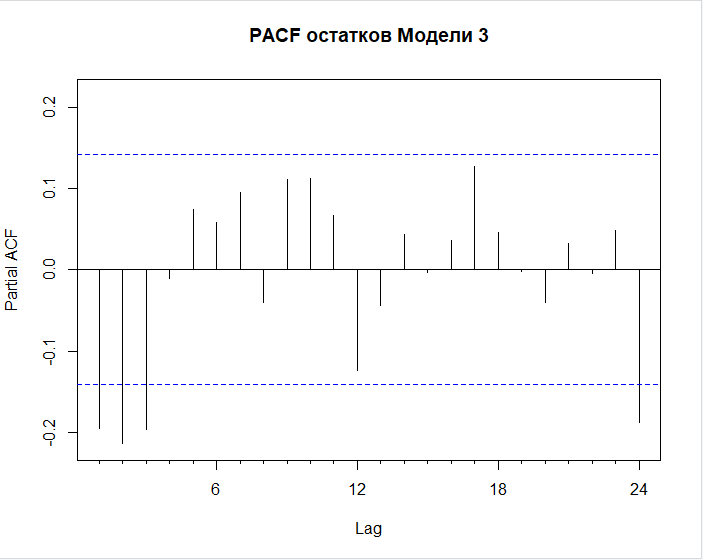


Рисунок 24. PACF остатки модели

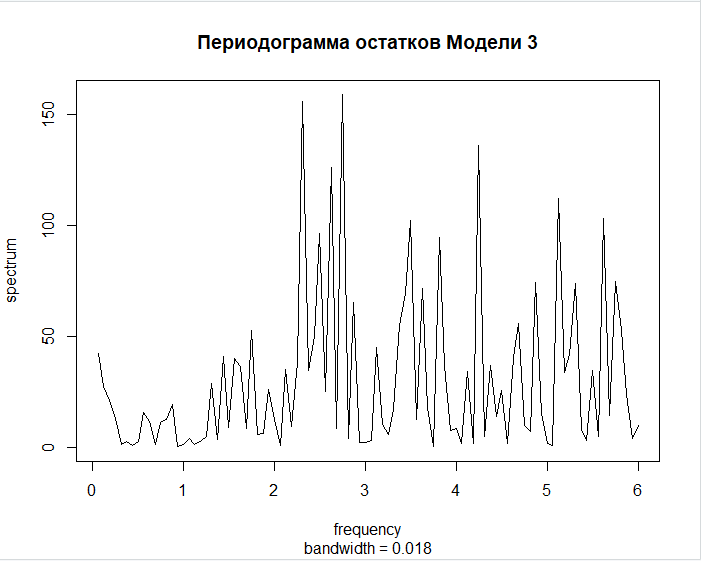


Рисунок 25. Периодограмма модели

По анализу периодограммы остатков для модели 3 можно сделать следующие выводы:

* На периодограмме видны выраженные пики на различных частотах, что указывает на наличие сильных периодических компонент в остатках. Это может свидетельствовать о том, что модель не полностью учла сезонность или цикличность данных, и в остатках присутствуют незамоделированные регулярные колебания.
* Поскольку в остатках есть значительные частотные компоненты, это говорит о том, что модель не является адекватной для описания временного ряда. Она оставляет незамоделированные циклические компоненты, что указывает на возможные упущенные факторы.

Таблица 1. Характеристики точности прогнозных моделей

|  |  |  |  |
| --- | --- | --- | --- |
|  | Модель1 | Модель2 | Модель3 |
| 1.Прогнозная модель | Линейная модель тренда | Линейная модель тренда + сезонная составляющая | Линейная модель + авто регрессивная составляющая |
| 2. Минимальный остаток (Min error) | -490.6338 | -432.0401 | -30.69759 |
| 3.Максимальный остаток (Max error) | 888.51 | 816.0018 | 64.41268 |
| 4. Средняя ошибка (Mean error) | 2.835752e-14 | 7.50956e-14 | 8.024203 |
| 5. СКО ошибки (Std. dev.) | 378.7061 | 375.1032 | 18.58017 |
| 6. Средняя абсолютная ошибка (Mean absolute error) | 323.1248 | 321.3082 | 15.15005 |
| 7. Средняя ошибка в процентах (Mean percentage error) | 23.27898 | 26.86997 | 1.634814 |
| 8. Средняя абсолютная ошибка в процентах (Mean absolute percentage error) | 80.24005 | 81.03242 | 3.718208 |
| 9. Средний квадрат ошибки (Root mean squared error) | 377.7186 | 374.1251 | 20.19437 |
| 10. Коэффициент детерминации | 0.8827674 | 0.8849874 |  |

Исходя из данных в таблице 1, сделаны выводы что, модель 3, которая включает линейную тенденцию, сезонную и авторегрессионную составляющие, является наиболее адекватной и точной из всех трех моделей. Она значительно снижает ошибки и демонстрирует наилучшее соответствие исходным данным, что делает её предпочтительной для прогнозирования в данном случае. Модель 1 с простым трендом имеет наибольшие отклонения и, вероятно, не учитывает значимые компоненты временного ряда, такие как сезонность и авторегрессию. Модель 2 улучшает прогноз за счет добавления сезонности, но всё же уступает модели 3, которая сочетает все ключевые компоненты для получения наиболее точного прогноза.

1. **Сравнение модели по точности. Прогноз значений ВР на 3 шага вперед на основе полученной итоговой модели.**

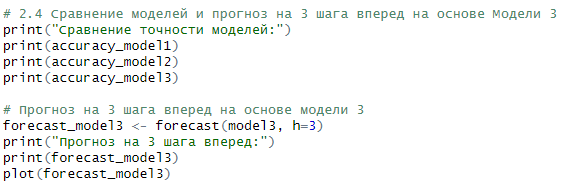


Рисунок 26. Сравнение и прогноз.

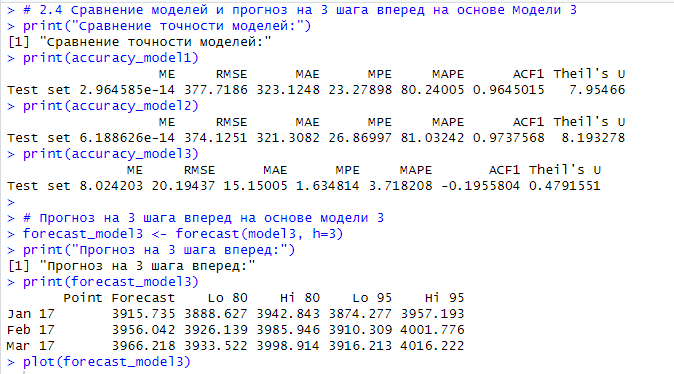


Рисунок 27. Результат сравнения и прогноза.

Сравнение показало, что модель 3 явно превосходит первые две модели по всем метрикам. Включение авторегрессионной составляющей позволило учесть остаточную корреляцию, что привело к значительному снижению ошибки и улучшению прогноза. Модель 3 можно считать наиболее подходящей для прогнозирования данных.

Прогноз показывает, что ожидается рост значений в течение следующих трёх месяцев. Интервалы доверия дают диапазоны, в которых можно ожидать реальные значения с определённой вероятностью (80% и 95%). Прогноз выглядит устойчивым и узкие интервалы доверия подтверждают, что модель является надёжной.

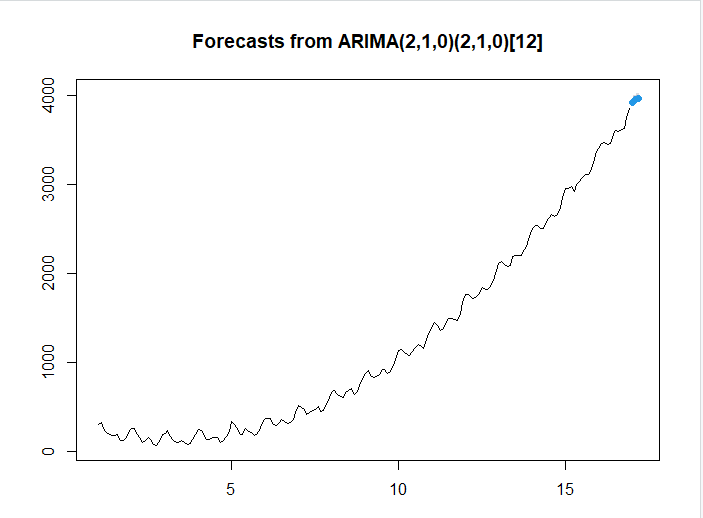


Рисунок 28. Изображение прогноза

**Заключение**

В ходе выполнения лабораторной работы были изучены методы и алгоритмы прогнозирования временных рядов, исследована эффективность прогнозирования временных рядом.

Результаты работы показывают, что для временных рядов с выраженной сезонностью и трендом наилучшие результаты дает использование комбинированных моделей, включающих как сезонные, так и авторегрессионные компоненты. Модель 3, использующая подход ARIMA с сезонной компонентой, позволяет достигнуть наибольшей точности при прогнозировании. Это подтверждается низкими значениями ошибок, что делает ее оптимальным выбором для задач прогнозирования на краткосрочный период.

**Приложение**

# Загрузка библиотек

library(forecast)

library(tseries)

# Загрузка данных

data <- read.csv("data.csv", header=TRUE, sep=";", check.names=F, fileEncoding="Windows-1251")

tsData <- ts(data[15], frequency=12) # Загрузка 12 варианта временного ряда

# 2.2 Построение графика временного ряда и его декомпозиция

plot.ts(tsData, main="Временной ряд")

decomp <- decompose(tsData)

plot(decomp)

# Построение автокорреляционной и частной автокорреляционной функций

Acf(tsData, main="ACF Временного ряда")

Pacf(tsData, main="PACF Временного ряда")

# 2.3.1 Идентификация тренда (Модель 1)

# Используем линейную модель тренда

t <- 1:length(tsData)

y <- as.numeric(tsData)

data1 <- data.frame(y=y, t=t)

model1 <- lm(y ~ t, data=data1) # Линейная регрессия

summary\_model1 <- summary(model1)

print(summary\_model1)

# Оценка точности модели

accuracy\_model1 <- accuracy(fitted(model1), tsData)

print("Точность Модели 1:")

print(accuracy\_model1)

min\_error <- min(residuals(model1))

max\_error <- max(residuals(model1))

mean\_error <- mean(residuals(model1))

std\_dev <- sd(residuals(model1))

mean\_absolute\_error <- mean(abs(residuals(model1)))

mean\_percentage\_error <- mean(100 \* residuals(model1) / y)

mean\_absolute\_percentage\_error <- mean(100 \* abs(residuals(model1)) / y)

root\_mean\_squared\_error <- sqrt(mean(residuals(model1)^2))

r\_squared <- summary\_model1$r.squared

# Вывод значений для таблицы

cat("Минимальный остаток (Min error):", min\_error, "\n")

cat("Максимальный остаток (Max error):", max\_error, "\n")

cat("Средняя ошибка (Mean error):", mean\_error, "\n")

cat("СКО ошибки (Std. dev.):", std\_dev, "\n")

cat("Средняя абсолютная ошибка (Mean absolute error):", mean\_absolute\_error, "\n")

cat("Средняя ошибка в процентах (Mean percentage error):", mean\_percentage\_error, "\n")

cat("Средняя абсолютная ошибка в процентах (Mean absolute percentage error):", mean\_absolute\_percentage\_error, "\n")

cat("Средний квадрат ошибки (Root mean squared error):", root\_mean\_squared\_error, "\n")

cat("Коэффициент детерминации (R-squared):", r\_squared, "\n")

# График исходного ряда с наложением тренда

plot(t, y, main="Исходный ряд с трендовой моделью (Модель 1)")

abline(model1, col="red")

# Остатки Модели 1 и их анализ

residuals\_model1 <- residuals(model1)

plot(residuals\_model1, main="Остатки Модели 1")

Acf(residuals\_model1, main="ACF остатков Модели 1")

Pacf(residuals\_model1, main="PACF остатков Модели 1")

spec.pgram(residuals\_model1, detrend=FALSE, log="no", main="Периодограмма остатков Модели 1")

# 2.3.2 Идентификация сезонной составляющей (Модель 2)

# Добавляем сезонную составляющую

model2 <- tslm(tsData ~ trend + season)

summary\_model2 <- summary(model2)

print(summary\_model2)

min\_error <- min(residuals(model2))

max\_error <- max(residuals(model2))

mean\_error <- mean(residuals(model2))

std\_dev <- sd(residuals(model2))

mean\_absolute\_error <- mean(abs(residuals(model2)))

mean\_percentage\_error <- mean(100 \* residuals(model2) / y)

mean\_absolute\_percentage\_error <- mean(100 \* abs(residuals(model2)) / y)

root\_mean\_squared\_error <- sqrt(mean(residuals(model2)^2))

r\_squared <- summary\_model2$r.squared

# Вывод значений для таблицы

cat("Минимальный остаток (Min error):", min\_error, "\n")

cat("Максимальный остаток (Max error):", max\_error, "\n")

cat("Средняя ошибка (Mean error):", mean\_error, "\n")

cat("СКО ошибки (Std. dev.):", std\_dev, "\n")

cat("Средняя абсолютная ошибка (Mean absolute error):", mean\_absolute\_error, "\n")

cat("Средняя ошибка в процентах (Mean percentage error):", mean\_percentage\_error, "\n")

cat("Средняя абсолютная ошибка в процентах (Mean absolute percentage error):", mean\_absolute\_percentage\_error, "\n")

cat("Средний квадрат ошибки (Root mean squared error):", root\_mean\_squared\_error, "\n")

cat("Коэффициент детерминации (R-squared):", r\_squared, "\n")

# Оценка точности модели 2

accuracy\_model2 <- accuracy(fitted(model2), tsData)

print("Точность Модели 2:")

print(accuracy\_model2)

# График исходного ряда с наложением тренда и сезонной составляющей

plot(tsData, main="Исходный ряд с трендовой и сезонной моделью (Модель 2)")

lines(fitted(model2), col="blue")

# Остатки Модели 2 и их анализ

residuals\_model2 <- residuals(model2)

plot(residuals\_model2, main="Остатки Модели 2")

Acf(residuals\_model2, main="ACF остатков Модели 2")

Pacf(residuals\_model2, main="PACF остатков Модели 2")

spec.pgram(residuals\_model2, detrend=FALSE, log="no", main="Периодограмма остатков Модели 2")

# 2.3.3 Идентификация авторегрессионной составляющей (Модель 3)

# Добавляем авторегрессионную модель, если в остатках присутствует корреляция

model3 <- auto.arima(tsData, seasonal=TRUE)

summary\_model3 <- summary(model3)

print(summary\_model3)

min\_error <- min(residuals(model3))

max\_error <- max(residuals(model3))

mean\_error <- mean(residuals(model3))

std\_dev <- sd(residuals(model3))

mean\_absolute\_error <- mean(abs(residuals(model3)))

mean\_percentage\_error <- mean(100 \* residuals(model3) / y)

mean\_absolute\_percentage\_error <- mean(100 \* abs(residuals(model3)) / y)

root\_mean\_squared\_error <- sqrt(mean(residuals(model3)^2))

r\_squared <- summary\_model3$r.squared

# Вывод значений для таблицы

cat("Минимальный остаток (Min error):", min\_error, "\n")

cat("Максимальный остаток (Max error):", max\_error, "\n")

cat("Средняя ошибка (Mean error):", mean\_error, "\n")

cat("СКО ошибки (Std. dev.):", std\_dev, "\n")

cat("Средняя абсолютная ошибка (Mean absolute error):", mean\_absolute\_error, "\n")

cat("Средняя ошибка в процентах (Mean percentage error):", mean\_percentage\_error, "\n")

cat("Средняя абсолютная ошибка в процентах (Mean absolute percentage error):", mean\_absolute\_percentage\_error, "\n")

cat("Средний квадрат ошибки (Root mean squared error):", root\_mean\_squared\_error, "\n")

cat("Коэффициент детерминации (R-squared):", r\_squared, "\n")

# Оценка точности модели 3

accuracy\_model3 <- accuracy(fitted(model3), tsData)

print("Точность Модели 3:")

print(accuracy\_model3)

# График исходного ряда с наложением Модели 3 (ARIMA с сезонностью)

plot(tsData, main="Исходный ряд с моделью (Модель 3)")

lines(fitted(model3), col="purple")

# Остатки Модели 3 и их анализ

residuals\_model3 <- residuals(model3)

plot(residuals\_model3, main="Остатки Модели 3")

Acf(residuals\_model3, main="ACF остатков Модели 3")

Pacf(residuals\_model3, main="PACF остатков Модели 3")

spec.pgram(residuals\_model3, detrend=FALSE, log="no", main="Периодограмма остатков Модели 3")

# 2.4 Сравнение моделей и прогноз на 3 шага вперед на основе Модели 3

print("Сравнение точности моделей:")

print(accuracy\_model1)

print(accuracy\_model2)

print(accuracy\_model3)

# Прогноз на 3 шага вперед на основе модели 3

forecast\_model3 <- forecast(model3, h=3)

print("Прогноз на 3 шага вперед:")

print(forecast\_model3)

plot(forecast\_model3)