МИНИСТЕРСТВО НАУКИ И ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»



Институт

ИЯТШ

ЛАБОРАТОРНАЯ РАБОТА № 1

«Прогнозирование временных рядов»

по дисциплине:

**Статистическое моделирование и прогнозирование**

Вариант № 7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Выполнил:** |  | | | | |
| студент группы 0В21 |  |  | Дзебан А.А. |  | 11.06.2025 |
|  |  |  |  |  | Дата сдачи |
| **Проверил:** | Рекундаль О.И. | | | | |
| доцент |  |  |  |  |  |
|  |  |  |  |  |  |

Томск – 2025

**Цели:**

1. Изучить приём вычисления выборочных автоковариаций и автокорреляций и их использование для проверки временных рядов на случайность.
2. Научиться выявлять и количественно описывать трендовую и сезонную компоненты временного ряда.
3. Освоить методы сглаживания (скользящее среднее, экспоненциальное сглаживание) и построение одношагового прогноза.
4. Сравнить точность различных моделей (линейная регрессия, MA, EMA) при помощи метрик MAE и MSE.
5. Закрепить навык построения аддитивной модели (тренд + сезонность) и выполнения прогноза на несколько периодов вперёд.

**Теоретическая часть:**

**1.** Временной ряд и его компоненты

Временной ряд — это упорядоченная последовательность наблюдений , полученных через равные интервалы времени. Классически выделяют четыре структурные составляющие:

* Тренд (T) — долгосрочная направленная динамика.
* Сезонность (S) — регулярные циклические колебания с известным периодом (квартал, месяц и т. д.).
* Циклическая компонента (C) — колебания с переменным периодом, отражающие деловые циклы.
* Случайная (остаточная) компонента (R) — нерегулярные колебания.

В рамках работы используется аддитивная схема декомпозиции, то есть:

2. Функция выборочной автокорреляции вычисляется по следующей формуле:

Для выявления автокорреляции проводится тест Льюинга-Бокса:

Проверяется

H0: отсчеты временного ряда статистически независимы

H1: отсчеты временного ряда не являюются независимыми.

Статистика:

Если , гипотеза H0 принимается.

Для выявления трендовости была проведена оценка значимости параметра b в линейной регрессии. (Если отвергается нулевая гипотеза b = 0 – то модель обладает трендом.)

Построение моделей прогнозирования

А. Трендовая модель

Если ряд содержит только тренд, используется регрессия:

Коэффициенты a, b находятся методом наименьших квадратов (МНК).

Б. Сезонная модель

Если ряд содержит сезонность, применяется аддитивная или мультипликативная модель.

Аддитивная модель:

Этапы построения:

1. Расчет скользящих средних для устранения сезонности.

2. Выделение сезонных компонент.

3. Корректировка сезонности (сумма компонент за период должна быть равна нулю).

4. Построение тренда на десезонализированных данных.

Для оценки качества моделей использовались следующие метрики:

Так же для получения трендов возможно использовать аппроксимации на расчете значения для различных окон:

Где L – длина окна L = 2m+1, где m -длина полуинтервала

**Практическая часть:**

**Задание 1:**

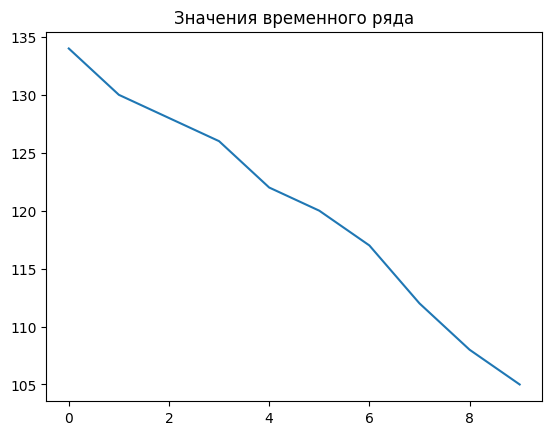
По данным о выпуске продукции за десять лет, которые представлены в табл. 1, в соответствии с номером варианта проверить временной ряд на случайность. Вычислить выборочные автоковариации и автокорреляции до пятого порядка включительно. Выявить наличие тренда и в случае положительного ответа построить трендовую часть модели. Тип трендовой модели выберите самостоятельно.

Выявить присутствие сезонности и в случае положительного ответа построить сезонную часть модели.

Построить линейную регрессию по времени (окончательный вид регрессии выбрать самостоятельно), оценив коэффициенты модели и вычислив дисперсию ошибки построенной модели.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № варианта | Годы выпуска продукции (t) | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 13,5 | 12,7 | 12 | 11,9 | 11,5 | 11,2 | 10,8 | 10,7 | 10,6 | 10,5 |
| 2 | 251 | 249 | 248 | 246 | 242 | 239 | 235 | 230 | 228 | 225 |
| 3 | 0,91 | 0,87 | 0,85 | 0,82 | 0,79 | 0,75 | 0,7 | 0,66 | 0,62 | 0,6 |
| 4 | 2,54 | 2,5 | 2,45 | 2,4 | 2,37 | 2,3 | 2,27 | 2,19 | 2,05 | 2 |
| 5 | 6,3 | 6,21 | 6,15 | 6 | 5,8 | 5,45 | 5,05 | 4,85 | 4,5 | 4,2 |
| 6 | 18,2 | 17,5 | 17,1 | 16,8 | 16,1 | 15,7 | 15,2 | 14,5 | 14,3 | 14 |
| **7** | **134** | **130** | **128** | **126** | **122** | **120** | **117** | **112** | **108** | **105** |
| 8 | 64 | 61 | 58 | 52 | 49 | 45 | 40 | 37 | 34 | 30 |
| 9 | 4,25 | 4,2 | 4,18 | 4,11 | 4,05 | 4 | 3,91 | 3,85 | 3,77 | 3,7 |
| 10 | 1,8 | 1,78 | 1,7 | 1,64 | 1,59 | 1,51 | 1,45 | 1,42 | 1,4 | 1,37 |
| 11 | 9,26 | 9,38 | 12,11 | 10,81 | 9,35 | 9,87 | 8,17 | 9,12 | 5,88 | 6,30 |
| 12 | 13,26 | 10,16 | 13,72 | 12,85 | 10,63 | 9,12 | 25,83 | 23,39 | 14,68 | 10,05 |
| 13 | 166,32 | 92,88 | 158,0 | 93,96 | 173,9 | 162,3 | 88,56 | 101,2 | 166,3 | 140,8 |
| 14 | 128,52 | 177,8 | 114,5 | 93,24 | 126,7 | 91,8 | 69,12 | 66,24 | 67,68 | 50,4 |

Зарисуем график данных:



Здесь заметен стабильный линейный тренд (это понадобится в задании 2).

Для вычисления выборочных автокорреляций воспользуемся следующей формулой:

Где

- выборочная автоковариация порядка k, – выборочное среднее временного ряда,

Таким образом – список вычисленных автоковариаций и автокорреляций выглядит так:

|  |  |  |
| --- | --- | --- |
| Лаг | Автоковариация | Автокорреляция |
| 0 | 84,16 | 1,0 |
| 1 | 64,37 | 0,7648 |
| 2 | 42,11 | 0,5 |
| 3 | 16,268 | 0,1933 |
| 4 | -11,426 | -0,1357 |
| 5 | -39,239 | -0,466 |

Для проверки ряда на случайность проведем тест Льюинга – Бокса (см.выше):

Значение статистики:

таким образом нулевая гипотеза отвергается – есть автокорреляция, а данные не являются случайными.

Для статистической оценки наличия линейного тренда (хотя он и очевиден из графика значений) построим линейную трендовую модель,

а после, на основании t-статистики вычислим значимость коэффициента :

С помощью МНК получим коэффициенты уравнения линейной регрессии:

Зарисуем и вычислим :

Значение t – статистики:

Отвергаем нулевую гипотезу об отсутствии тренда (b 0) - модель имеет выраженный тренд.

Исходя из поведения значений автокорреляции (монотонно убывают) сделаем вывод об отсутствии сезонности.

Аналогично оценим значимость коэффициента c:

Значение t – статистики:

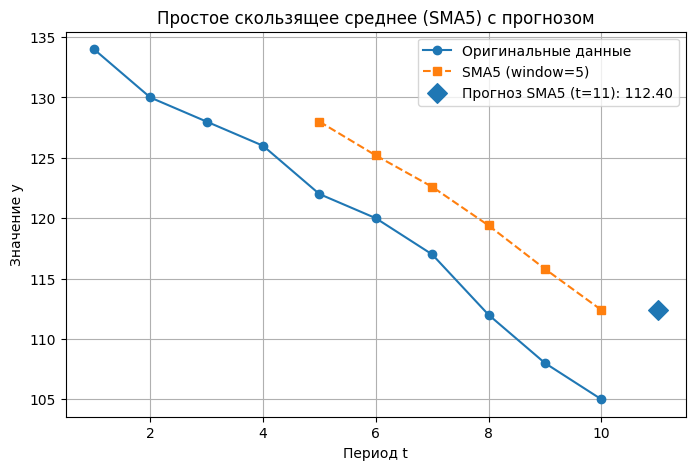
Таким образом коэффициент c значим.

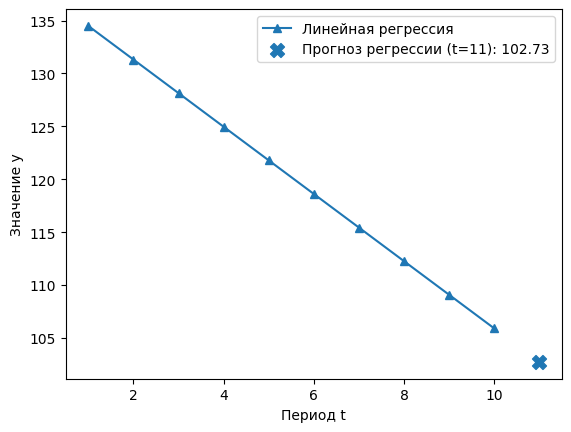
Дисперсия ошибок вычисляется по следующей формуле:

**Задание 2:**

По данным табл. 1 проведите сглаживание данных (методом скользящего среднего или с помощью экспоненциального сглаживания) с длиной полуинтервала *m* = 2 и выполните прогноз на следующий период *t* = 11 для своего варианта задания, используя при необходимости трендовую (сезонную) части модели или их сумму, а так же линейную регрессию. Сравните погрешности различных методов на исходных данных.

Учитывая, что m = 2 -> окно скользящего среднего = 5, определим значения MA5 и составим прогнозы на t =11.





Здесь видно, что MA в таком большом окне относительно количества значений дает неточные прогнозы, учитывая ярко выраженный тренд, а линейная регрессия в свою очередь (учитывая высокие значения ) позволяет получить более логичные предсказания.

Для оценки качества определим MAE и MSE на соответствующих участках

Для

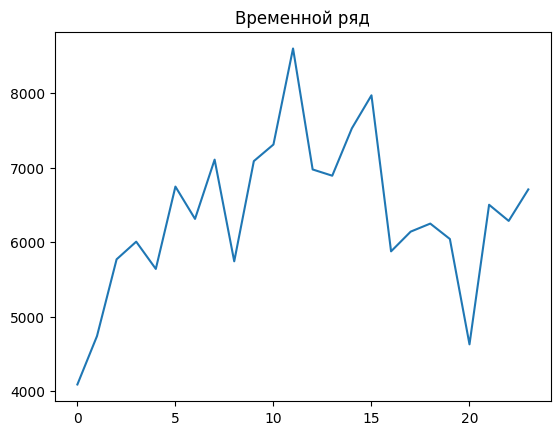
|  |  |  |
| --- | --- | --- |
| Модель\Метрика | SMA5 | Регрессия |
| MAE | 6,567 | 0,903 |
| MSE | 44,127 | 1,081 |

Очевидно, регрессия дает значительно лучшее качество.

**Задание 3:**

В табл. 3 имеются данные об объеме экспорта по кварталам за 2016-2021 гг. Постройте модель временного ряда: трендовую (четные варианты), сезонную (нечетные варианты) или аддитивную (тренд + сезонность), если первоначальная модель неполна или невозможна. Спрогнозируйте экспорт по кварталам на 25 и 26 периоды времени.

|  |  |  |  |
| --- | --- | --- | --- |
| Номер периода | Экспорт, млн руб. | Номер период | Экспорт, млн руб. |
| 1 | 4087 | 13 | 6975 |
| 2 | 4737 | 14 | 6891 |
| 3 | 5768 | 15 | 7527 |
| 4 | 6005 | 16 | 7971 |
| 5 | 5639 | 17 | 5875 |
| 6 | 6745 | 18 | 6140 |
| 7 | 6311 | 19 | 6248 |
| 8 | 7107 | 20 | 6041 |
| 9 | 5741 | 21 | 4626 |
| 10 | 7087 | 22 | 6501 |
| 11 | 7310 | 23 | 6284 |
| 12 | 8600 | 24 | 6707 |



Из графика прослеживается сезонность и трендовость – составим аддитивную модель:

Для расчета сезонных компонент воспользуемся методом центрированной скользящей средней, а именно:

1. Вычислим скользящую сумму для окна длиной 4
2. Среднее за 4 года:
3. Определим центрированное скользящее среднее:

Полученная серия значений CMA отражает только долгосрочный тренд без сезонных колебаний.

Тогда остатки

Эти остатки несут в себе сезонные составляющие

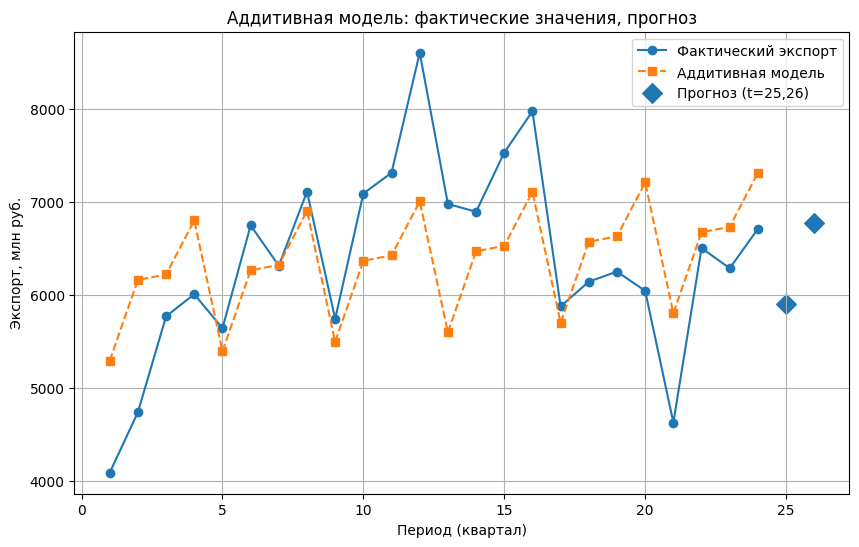
Каждому t будем сопоставлять номер сезона:

Для каждой четверти берем остатки rt и с соответствующим номером qt подсчитаем первичные сезонные индексы

В аддитивной модели суммарное влияние сезонности за полный цикл должно равняться нулю, для этого найдем

Окончательные сезонные индексы считаются по формуле:

Итоговая модель имеет следующий вид:



Здесь заметно, что модель недостаточно хорошо описывает поведение временного ряда – наверняка требуется добавить другие компоненты (например, случайные) для более точной аппроксимации.

**Вывод:**

1. С помощью автокорреляционного анализа и критерия Льюнга–Бокса показано, что исходный ряд (вариант 7) не является случайным; в нём присутствует сильная положительная взаимосвязь между соседними значениями.
2. Линейная модель выявляет устойчивый нисходящий тренд
3. Сезонность для этого ряда отсутствует, что подтверждается формой ACF и отсутствием значимых пиков на кратных лагах.
4. При прогнозировании на шаг вперёд линейная регрессия (MAE ≈ 0.90; MSE ≈ 1.08) заметно превосходит скользящее среднее MA5 (MAE ≈ 6.57; MSE ≈ 44.13).
5. Для квартальных данных экспорта 2016–2021 гг. построена аддитивная модель: линейный тренд и квартальные индексы. Итоговый прогноз 5903:  млн руб., 6773 млн руб.
6. Практическое освоение методов SMA, EMA и регрессионного тренда показало, что выбор модели должен учитывать структуру данных: при ярко выраженном тренде простые усреднения дают смещённые прогнозы.