Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский Томский политехнический Университет»



Институт

ИЯТШ

01.03.02 «Прикладная математика и информатика»

**Лабораторная работа №1**

Прогнозирование временных рядов

по дисциплине:

**Статистическое моделирование**

**2 вариант**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Выполнил:** |  | | | | |
| студент группы | 0В01 |  | Белясов А.А |  | 12.03.2023 |
|  |  |  |  |  | Дата сдачи |
| **Проверил:** | Крицкий О.Л. | | | | |
| преподаватель |  |  |  |  |  |
|  |  |  |  |  |  |

Томск – 2023

**Теоретическая справка.**

Рассмотрим временной ряд *xt* вида

*xt*=*τt*+*vt*+ε*t*,

где *τt* – трендовая составляющая или медленное изменение временного ряда в некотором направлении, которое сохраняется в течение длительного промежутка времени, *vt* – сезонная составляющая или изменения, которые происходят регулярно на ежегодной, ежемесячной, еженедельно и т.п. основе, например, выходные дни каждой недели или Новый год,

t ~ *N*0,1.

С экономической точки зрения очень часто происходит так, что вместо аддитивного влияния тренда, сезонности и шума на значения временного ряда *xt* оно мультипликативное:

*xt*=*τtvtexp*(ε*t*),

которое можно записать в аддитивном виде, прологарифмировав.

14

12

10

8

6

4

2

0

1

2

3

4

5

6

7

8

9

10

Рисунок 1. Сезонность и тренд временного ряда

Существует три вида трендов:

1. Тренд среднего – временной ряд выглядит как колебания около медленно возрастающей или убывающей величины;
2. Тренд дисперсии – временной ряд имеет изменяющиеся во времени амплитуды колебаний (гетероскедастичность процесса);
3. Тренд автоковариации (автокорреляции) – временной ряд обладает изменчивостью корреляции между текущим и предшествующим значением ряда.

Тренды среднего бывают:

* полиномиальный тренд:

*τt*=*a*0+ *a*1*t*+…+ *antn*, который при *n*=1 превращается в линейный тренд;

* экспоненциальный тренд:
* гармонический тренд

*τt*=exp(*a*0+ *a*1*t*+…+ *antn*),

*τt*=*R* cos(ω*t*+φ),

где *R–* амплитуда колебаний, ω – частота, φ – фаза колебаний.

* тренд, описываемый логистической функцией

Для оценки коэффициентов полиномиального или экспоненциального тренда нужно

использовать обычный МНК после введения новой переменной *z*  *t j* , j=1,2,…, *n*.

j

Гармонический тренд стоит использовать, если в структуре исходных данных отчетливо прослеживаются периодические колебания. Фиксируя ω или задавая ее оценку, гармонический тренд можно представить в виде *τt*=*a* cos(ω*t*)+ *b sin*(ω*t*) и использовать для оценки неизвестных коэффициентов *a*, *b* МНК.

**Замечание**: можно показать, что наименьшая дисперсия ошибки регрессионной модели достигается на ортогональных многочленах. Заменяя систему функций {1*, t*,…, *tn*} на такие многочлены, например, многочлены Чебышева, Эрмита, Лаггера, мы добиваемся наилучшей точности.

Далее, для моделирования сезонной составляющей можно использовать фиктивные переменные:

*vt*=*λ*1δ1*t*+ *λ*2δ2*t*+…+ *λp*δ*pt*, (\*)

где δ*it–*сезонные фиктивные переменные. δ*jt*=1 в сезон *j* и нуль в остальное время*, p –* общее число сезонов*.*

При построении тренда τt вместо оценивания его коэффициентов можно использовать метод скользящих средних, когда значения временного ряда xt заменяются последовательностью вычисленных на перемещаемом отрезке средних его величин. Пусть (2m+1) – длина некоторого отрезка времени. Подберем полином

к группе первых (2m+1) членов ряда. Этот полином будем использовать в дальнейшем при определении значения тренда в точке t=(m+1), т.е. в середине выбранного отрезка. Далее временной отрезок сдвигается на единицу вправо, т.е. рассматриваются моменты 2, 3, …, (2m+2) и вычисления повторяются: находится значение τm+2. Проводим процедуру до тех пор, пока не будет достигнута последняя группа из (2m+1) точек. Поиск коэффициентов a0, a1,…, ap полинома осуществляем с помощью МНК по первым (2m+1) точкам.

Используем несколько выборочных автокорреляций *r*1, *r*2,…*rm*, *m*<*T*, для доказательства гипотезы о случайности значений исходного временного ряда *xt*. Рассмотрим *Q*-статистику **Льюнга-Бокса** в предположении о нормальности распределения значений *xt*

Было показано, что статистический критерий работает даже в условиях отсутствия нормального закона распределения для исходного временного ряда (выполнена ЦПТ, т.е.

дисперсия *D**xt*   ). Нулевая гипотеза состоит в том, что ряд *xt* является винеровским

процессом, т.е. процессом с независимыми приращениями и нулевым средним. Если значение статистики *Q*(*r*) больше критического (табличного) значения функции распределения при заданном уровне значимости и числе степеней свободы, то признается наличие ненулевых автокорреляций до порядка *m* включительно.

Для проверки случайности можно использовать и другие критерии, например, непараметрический **критерий Спирмена**. Если ряд *xt* случайный, то распределение отдельного наблюдения не зависит от того, на каком месте он в этом ряде стоит. Если упорядочить значения *xt*, например, по возрастанию или по убыванию, получив новый ряд *yt*, и сравнить полученные значения с исходной последовательностью, вычислив корреляцию между значениями рядов.

*T*

Пусть для временного ряда *xt*=*τt*+*vt*+ε*t* вычислены автокорреляции до *k*-го порядка включительно. Проверим статистическую гипотезу о величине первой автокорреляции ρ1, вычисляя статистику Дарбина-Уотсона (**DW-test**):

*et* – погрешность модели (разность между наблюдаемым и модельным значением). Значения статистики γ лежат в интервале [0,4]. Распределение статистики известно и имеет два критических значения: *dl* и *du*, их значения приведены в таблицах. Выдвигаем нулевую гипотезу *H*0: ρ1=0, где ρ1 - первая автокорреляция (остальные автокорреляции статистически не проверяются). Нулевая гипотеза о нулевой автокорреляции подтверждается, если *du*<γ<4-*du*. *H*0 отклоняется в пользу альтернативы о наличии положительной автокорреляции, если γ<*dl*. *H*0 отклоняется в пользу альтернативы о наличии отрицательной автокорреляции, если 4-*dl*<γ. Зона неопределенности критерия, когда нельзя ни принять основную гипотезу, ни принять альтернативную, состоит из двух интервалов, описываемых неравенствами: *dl*<γ<*du* и 4-*du*<γ<4-*dl*.

Регрессионный анализ – это статистический метод исследования функциональной связи случайной величины *y* от переменных 𝑥𝑖, 𝑖 = 1 , 𝑛 , рассматриваемых как неслучайные (известные) многомерные случайные величины с произвольной функцией распределения.

Задачей регрессионного анализа является построение зависимости изучаемой случайной величины *y* от факторов *x* по результатам наблюдения:

𝑦 = 𝑓 𝑥, 𝜃 +ξ, (43.1)

где 𝜃 – неизвестные параметры, ξ– случайные ошибки или остатки, 𝑦𝑖 = 𝑦 𝑥𝑖 = 𝑓 𝑥𝑖, 𝜃 + ξ(𝑥𝑖) – независимые наблюдения, 𝑖 = 1 , 𝑛 .

Регрессионные модели бывают линейными и нелинейными. В линейной регрессионной модели функция 𝑓 𝑥, 𝜃 линейна по параметрам. Уравнение линейной регрессии записывается следующим образом:

𝑦^ = 𝑓(𝑥, 𝜃^) = 𝜃^𝑇 𝜑(𝑥), (43.2)

где 𝜑𝑇(𝑥) = 𝜑1 𝑥 … 𝜑𝑚 (𝑥) – известные функции, *m*<*n*, 𝜃^ – оценка параметров регрессии.

Будем рассматривать линейные оценки истинных параметров. Можно найти оценки 𝜃^, которые являются состоятельными, несмещенными и обладают наименьшими дисперсиями среди множества всех линейных несмещенных оценок. Такие оценки называются

наилучшими линейными оценками, и в случае независимых и распределенных с одинаковыми дисперсиями случайными ошибками ξ вычисляются по формуле:

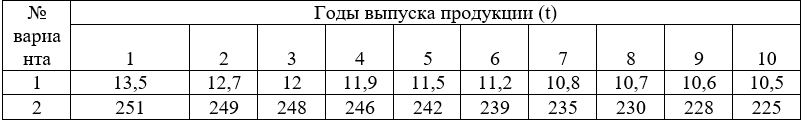
𝜃^ = 𝐼−1𝑍,

где матрица 𝐼 размерности *m*х*m* равна (умножаем вектор-столбец на вектор-строку размерности 1x*m*)

**Задание**

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. Проверить статистику Дарбина-Уотсона на нулевое значение первой автокорреляции. 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 или нелинейную регрессию по времени (окончательный вид и тип регрессии выбрать самостоятельно), оценив коэффициенты модели и вычислив дисперсию ошибки построенной модели.



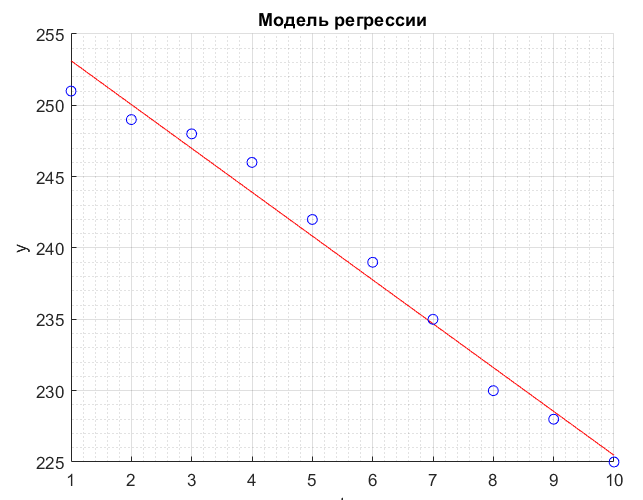
По данным табл. 1 проведите сглаживание данных (методом скользящего среднего или с помощью экспоненциального сглаживания) с длиной полуинтервала *m*=2 и выполните прогноз на следующий период *t*=11 для своего варианта задания, используя при необходимости трендовую (сезонную) части модели или их сумму,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

В табл. 2 имеются данные об объеме экспорта по кварталам за 2011-2016 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 |

**Практическая часть**

Построим линейную регрессию и определим ее дисперсию ошибки.



Уравнение модели:   
Дисперсия ошибки равна 2.1455

Проведем сглаживание данных методом скользящего среднего с длиной полуинтервала *m*=2 и выполним прогноз на следующий период *t*=11. Получим предсказанное значение:

Проверим статистику Дарбина-Уотсона на нулевое значение первой автокорреляции.  
Получим Мы оказались в зоне неопределенности…

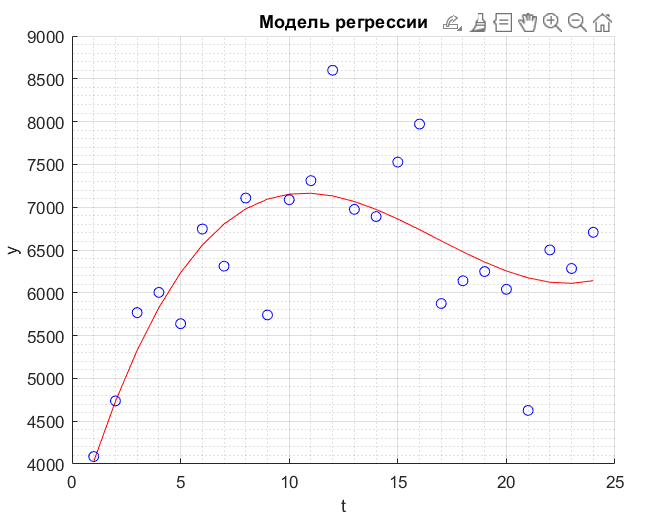
Найдем выборочные автоковариации до 5го порядка:

Также найдем выборочные автокорреляции до 5го порядка:

Проверим случайность данных по статистике Льюнга-Бокса, получим

Таким образом данные распределены случайно

Построим трендовую кубическую модель временного ряда



Уравнение модели имеет вид

Спрогнозируем на следующие 2 периода  
При t=25 y=6226.67 t=26 y=6368.481

**Вывод**: в ходе лабораторной работы мы научились определять случайность данных, находить выборочные автоковариации и автокорреляции. Так же мы построили регрессионные модели и предсказали с помощью них будущие значения.

**Приложение**

**Main**

clc, clearvars, close all, format compact

y\_1 = [251,249,248,246,242,239,235,230,228,225];

y\_2 = [4087, 4737, 5768, 6005, 5639, 6745, 6311, 7107, 5741, 7087, 7310, 8600, 6975, 6891, 7527, 7971, 5875, 6140, 6248, 6041, 4626, 6501, 6284, 6707];

n\_1 = length(y\_1);

n\_2 = length(y\_2);

t\_1 = 1:n\_1;

t\_2 = 1:n\_2;

m = 2;

X\_1 = horzcat(ones(n\_1, 1), transpose(t\_1));

X\_2 = horzcat(ones(n\_2, 1), transpose(t\_2),transpose(t\_2.^2),transpose(t\_2.^3));

theta\_1 = find\_parametrs(X\_1, y\_1);

create\_regression\_model(t\_1, y\_1, theta\_1, 1);

[disp, c, r, gamma] = find\_characteristics(n\_1, y\_1, t\_1, theta\_1);

y\_pred = predict(y\_1);

create\_plot(y\_2);

theta\_2 = find\_parametrs(X\_2, y\_2);

create\_regression\_model(t\_2, y\_2, theta\_2, 0);

**predict**

function y\_pred = predict(y)

a = zeros(6, 3);

t = 3;

for i = 1:6

a(i,1) = 1/35 \* (-3 \* y(i) + 12 \* y(i+1) + 17 \* y(i+2) + 12 \* y(i+3) - 3 \* y(i+4));

a(i,2) = 1/10 \* (-2 \* y(i) - y(i+1) + y(i+3) + 2 \* y(i+4));

a(i,3) = 1/14 \* (2 \* y(i) - y(i+1) - 2 \* y(i+2) - y(i+3) + 2 \* y(i+4));

end

y\_pred = a(6,1) + a(6,2) \* t + a(6,3) \* t^2;

disp('прогноз')

disp(y\_pred)

**create\_plot**

function create\_plot(data)

figure('Color', 'w')

hold on

plot(data, Color='red')

hold off

grid on

grid minor

title('Модель данных');

xlabel('Номер периода ')

ylabel('Экспорт, млн руб. ')

**create\_regression\_model**

function create\_regression\_model(t, y, params, flag)

if flag == 1

f = @(x) params(1) + params(2)\*x;

else

f = @(x) params(1) + params(2)\*x + params(3)\*x.^2+params(4)\*x.^3;

end

figure('Color', 'w')

hold on

plot(t, f(t), Color='red')

plot(t, y, 'bo')

hold off

grid on

grid minor

title('Модель регрессии');

xlabel('t')

ylabel('y')

**find\_characteristics**

function [disp, c , r, gamma,Q] = find\_characteristics(n, y, t, params)

model = @(x) params(1) + params(2) \* x;

m = 5;

Y = ones(n, 1); disp = 0;

c = zeros(6, 1);

r = zeros(6, 1);

e = zeros(n, 1);

for i = 1:n

disp = disp + (model(t(i)) - y(i))^2;

Y(i) = model(t(i));

end

disp = disp / (n-2);

y\_mean = mean(Y);

for i = 1:m+1

sum = 0;

for j = 1:n-(i-1)

sum = sum + (Y(j) - y\_mean) \* (Y(j+(i-1)) - y\_mean);

end

c(i) = sum / n;

end

for i = 2:m+1

r(i) = c(i) / c(1);

end

for i = 1:n

e(i) = model(t(i)) - y(i);

end

gamma = 0;

sum\_gamma = 0;

for i = 2:n

gamma = gamma + (e(i) - e(i-1))^2;

sum\_gamma = sum\_gamma + e(i)^2;

end

gamma = gamma / sum\_gamma;

Q=0;

for i=1:6

Q=Q+r(i)\*r(i)\*n\*(n+2)/(n-i+1);

end

**find\_parametrs**

function params = find\_parametrs(X, y)

params = inv(transpose(X) \* X) \* transpose(X) \* transpose(y);