**МОДЕЛИРАНЕ НА ДИНАМИЧНИ РЕДОВЕ – АНАЛИЗ НА ОСТАТЪЦИТЕ**

Общият вид на модела на динамичен ред е

, където

*  (тренд),  (сезонна компонента) и  (циклична компонента) са детерминирани;
*  (остатъците) е случайна компонента, която сама по себе си представлява стационарен ред.

Условията, на които трябва да отговаря  са

* Да има нормално разпределение;
* Математическото очакване да е равно на 0;
* Да липсва тенденция на средната и на дисперсиятата;
* Да отсъства автокорелация.

За проверката на тези условия следва да се използват подходящи статистически тестове.

Проверката за нормалност може да се извърши с теста на Shapiro-Wilk. Статистиката на теста е

,

Където  са стойностите на остатъците, подредени по възходящ ред (наредени статистики), а  са коефициенти, които се вземат от таблица, генерирана с метода Монте-Карло. Критичните стойности на статистиката също са определени с метода Монте-Карло.

Класическият тест на Shapiro-Wilk се прилага за извадки с обем до 50. Royston е създал разширение на теста за извадки с обем до 2000.

Тестът на Shapiro-Wilk (с разширението на Royston) може да се приложи в средата на R с вградената функция **shapiro.test** от пакета **stats**.

Проверката за нулева стойност на математическото очакване може да се извърши с проерка за равенство на средната на 0 при неизвестна дисперсия. Нулевата хипотеза е , т.е., че средната е равна на 0. Алтернативната хипотеза е , т.е., че средната не е равна на 0. Определя се двустранна критична област.

Оценява се средната по формулата

.

Намира се неизместената оценка на дисперсията по формулата

.

Статистиката



има разпределение на Стюдънт с  степени на свобода.

Двустранната критична област се определя от неравенството

.

Ако неравенството не е изпълнено, няма основание за отхвърляне на нулевата хипотеза. Ако неравенството е изпълнено, има основание за отхвърляне на нулевата хипотеза.

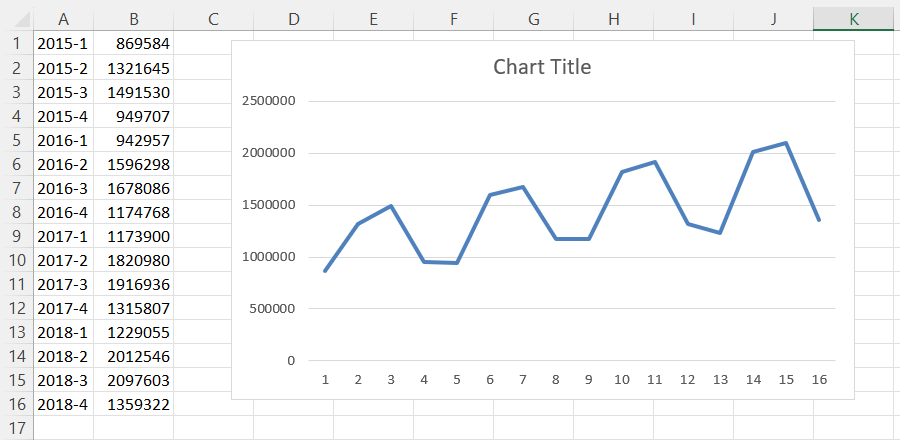
Отсъствието на тенденция на средната и дисперсията може да се се провери с теста на Фостър-Стюарт, като се използва вградената в R функция **foster.test** от пакета **RecordTest.**

Отсъствието на автокорелация може да се провери с теста на Дърбин-Уотсън с помощта на вградената в R функция **dwtest** от пакета **lmtest.**

**Пример:** Даден е броят на българските граждани, пътували зад граница през периода 2015-2018 година по тримесечия (файл TimeSer1.xlsx, лист Danni). Необходимо е да се моделира реда и да се анализира модела.

Ще използваме всички тестове с ниво на значимост .

Построява се диаграма. Ясно се вижда, че редът притежава тренд и сезонна компонента.

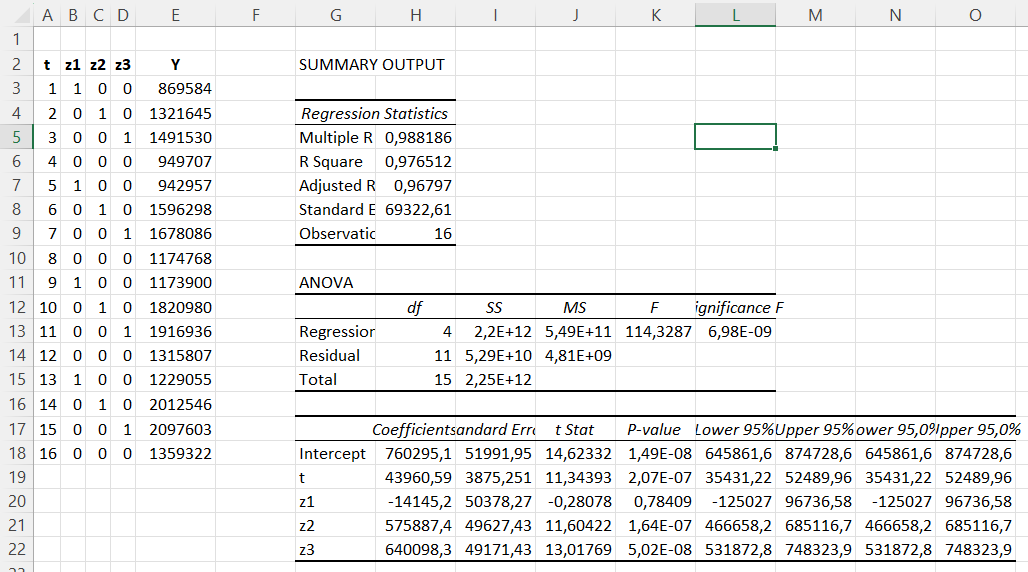


Един от възможните начини за моделиране е чрез използване на фиктивни променливи, като моделът има вида

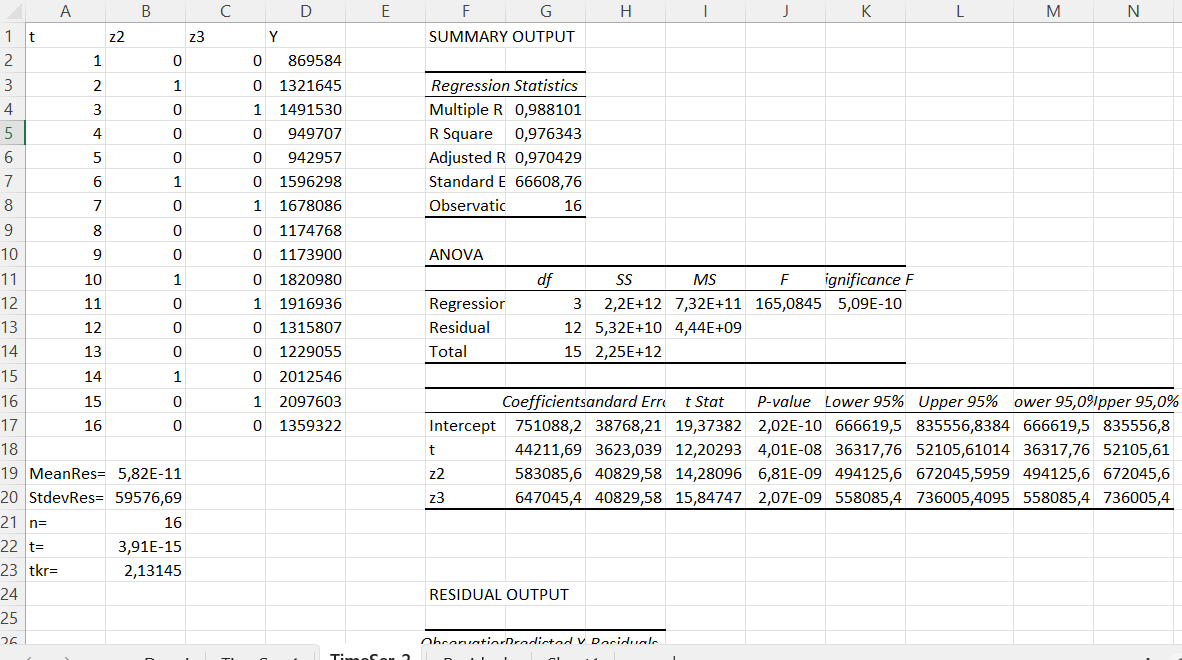
, където

 са независими променливи, които имат стойност 1 през -тото тримесечие и 0 през останалите тримесечия.

За построяването на реда може да се използва инструментът Data Analysis в Excel (стр. **TimeSer-1**).



От резултатите се вижда, че променливата  не е статистически зна-чима. Отстраняваме променливата и създаваме нов модел (стр. **TimeSer-2**), като задаваме извеждане на остатъците (Residuals).



Проверяваме дали средната на остатъците се отличава значително от 0. Стойността на статистиката . Критичната стойност е , следователно няма основание да считаме, че средната е различна от 0.

За да използваме R, създаваме листове **Residuals** и **Model DW**. В лист **Residuals** записваме стойностите на остатъците, В лист **Model DW** се записват началните данни на модела (свързано е с особеностите на вградената функция **dwtest**)**.**

В средата на R инсталираме пакетите **xlsx**, **RecordTest** и **lmtest** и ги асоциираме с текущата сесия:

**install.packages(“openxlsx”)**

**install.packages(“RecordTest”)**

**install.packages(“lmtest”)**

**library(openxlsx)**

**library(RecordTest)**

**library(lmtest)**

Прочитаме данните за остатъците:

**res<-read.xlsx("TimeSer1.xlsx",4)**

Аргументът 4 указва от кой поред лист се прочитат данните.

Прочитаме данните за модела

**Mod1<-read.xlsx("TimeSer1.xlsx",5)**

Проверяваме остатъците за нормалност с теста на Shapiro-Wilk:

**> shapiro.test(res$res)**

**Shapiro-Wilk normality test**

**data: res$res**

**W = 0.89229, p-value = 0.06055**

Вероятността , следователно няма основание за отхвърляне на нулевата хипотеза, че разпределението е нормално.

С теста на Фостър-Стюарт проверяваме за отсъствие на тренд в

А) средната

**> foster.test(res$res,statistic="d")**

**Foster-Stuart d-statistic test**

**data: res$res**

**Z = -0.45828, p-value = 0.6766**

**alternative hypothesis: true 'statistic' is greater than 0**

**sample estimates:**

**statistic E VAR**

**-1.000000 0.000000 4.761458**

Вероятността , следователно няма основание за отхвърляне на нулевата хипотеза, че липсва тренд w средната

Б) дисперсията

**> foster.test(res$res,statistic="s")**

**Foster-Stuart s-statistic test**

**data: res$res**

**Z = -2.4159, p-value = 0.9922**

**alternative hypothesis: true 'statistic' is greater than 6.761458**

**sample estimates:**

**statistic E VAR**

**3.000000 6.761458 2.424072**

Вероятността , следователно няма основание за отхвърляне на нулевата хипотеза, че липсва тренд в дисперсията

Отсъствието на автокорелация проверяваме с теста на Дърбин-Уотсън.

**> dwtest(Y~t+z2+z3,data=mod1)**

**Durbin-Watson test**

**data: Y ~ t + z2 + z3**

**DW = 2.1622, p-value = 0.5563**

**alternative hypothesis: true autocorrelation is greater than 0**

Тази вградена функция изисква като вход изходните данни за модела и сама определя коефициентите на модела, изислява остатъците и ги проверява за автокорелация.

Изчислената статистика има стойност DW=2,1622. От таблицата за извадка с обем 16 и модел с 4 параметъра( вкл. свободния член) определяме стойностите Dl=0,85718 и Du=1,72773.

Изчислената статистика попада в интервала

, следователно няма основание за отхвърляне на нулевата хипотеза, че в реда от остатъците липсва автокорелация.

Тестът на Дурбин-Уотсън може да се изпълни и в средата на Excel. За целта се пресмята статистиката

.

За целта се създава отделен лист

