МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

ЗВІТ

до індивідуального завдання №6

з дисципліни «Моделі статистичного навчання»

Виконав

студент групи ПМіМ-12:

Зелінський Олександр

Перевірив:

Проф. Заболоцький Т. М.

Львів – 2021

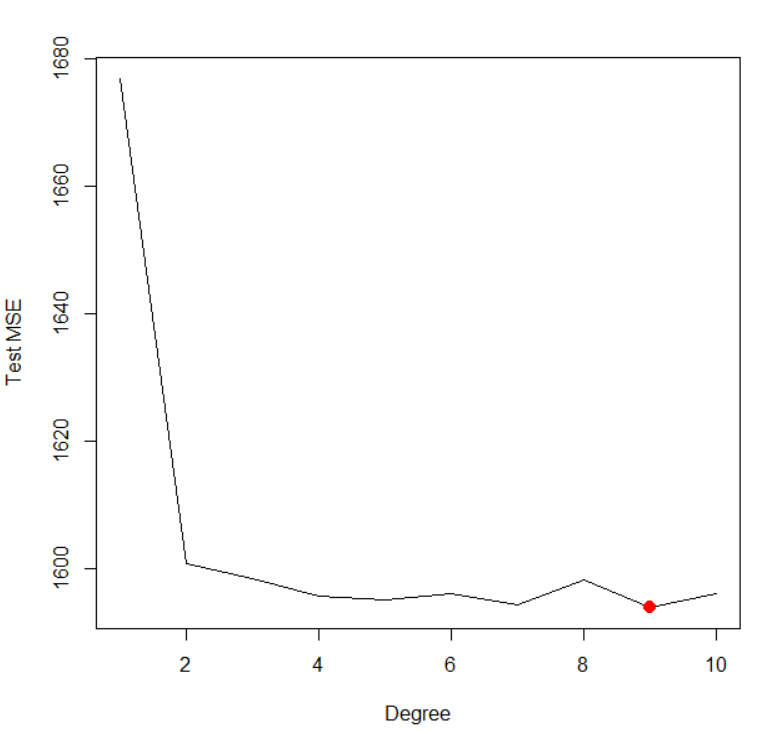
**Хід виконання**

**1. Wage**

**1.1**

Зображення, що містить текст

Автоматично згенерований опис



Зважаючи на результат можемо побачити, що d=9 є оптимальним степенем для полінома.

Зображення, що містить стіл

Автоматично згенерований опис

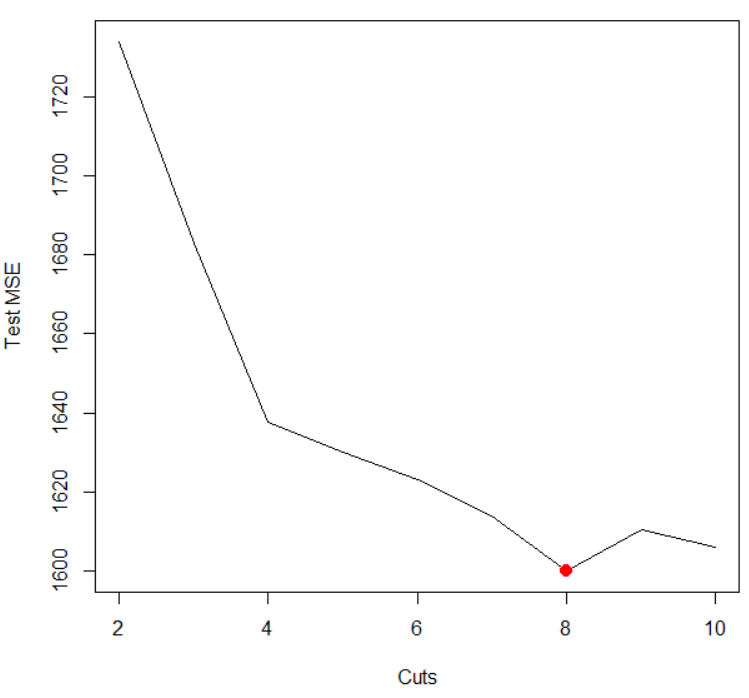
З результатів видно, що зважаючи на p-значення, ми можемо побачити, що поліном 2 і 3 степеня забезпечує найкращий результат. Побудуємо ці два поліноми.

|  |  |
| --- | --- |
|  |  |

**1.2**

Зображення, що містить текст

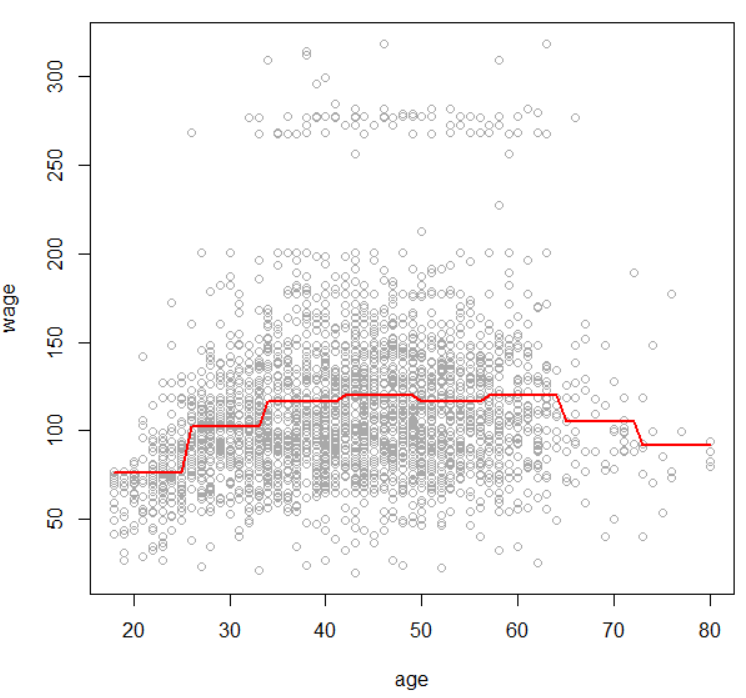
Автоматично згенерований опис



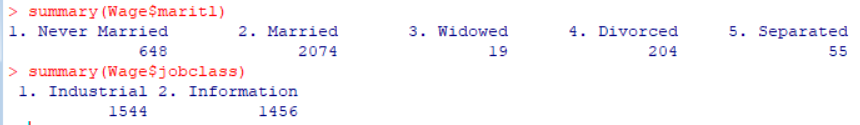
Зважаючи на результат можемо побачити, що помилка буде мінімальною для 8 зрізів.

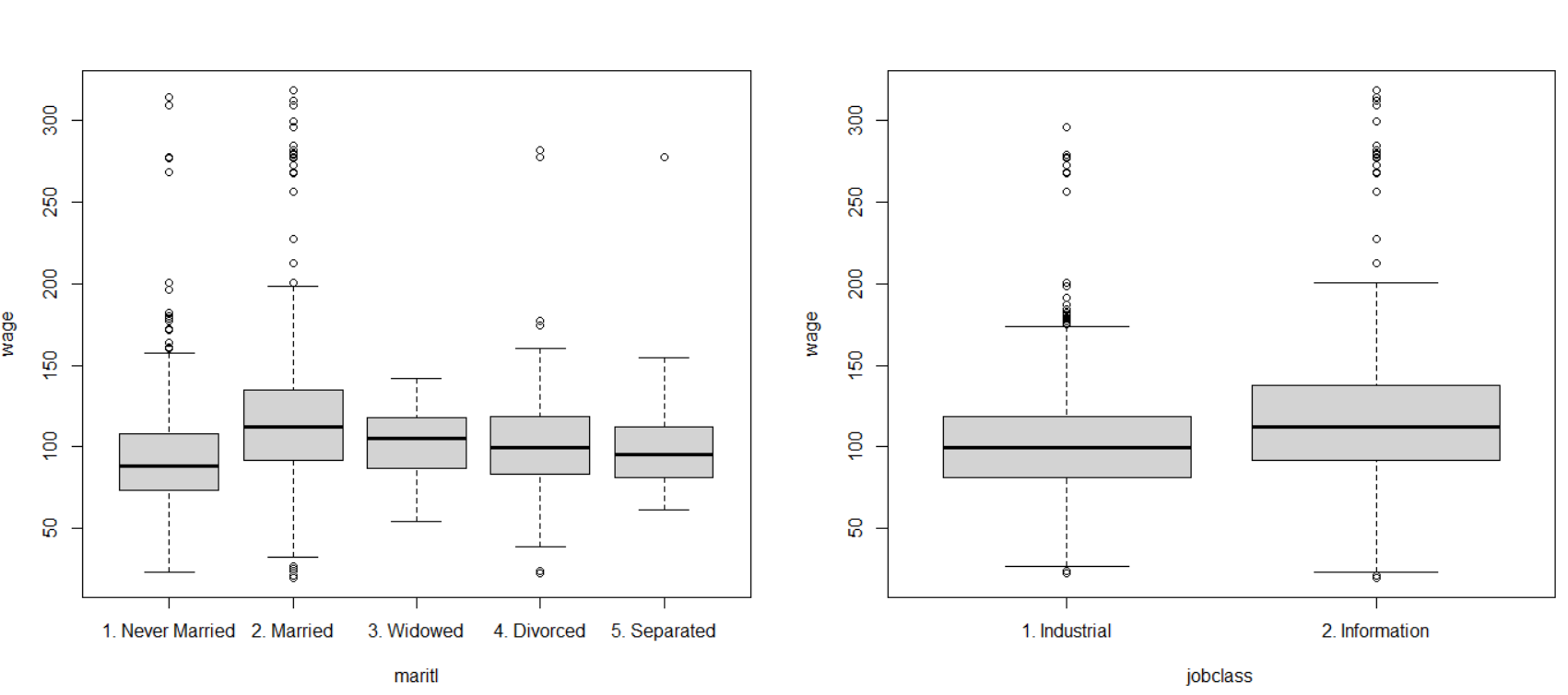
Зображення, що містить текст

Автоматично згенерований опис



**2. Wage**



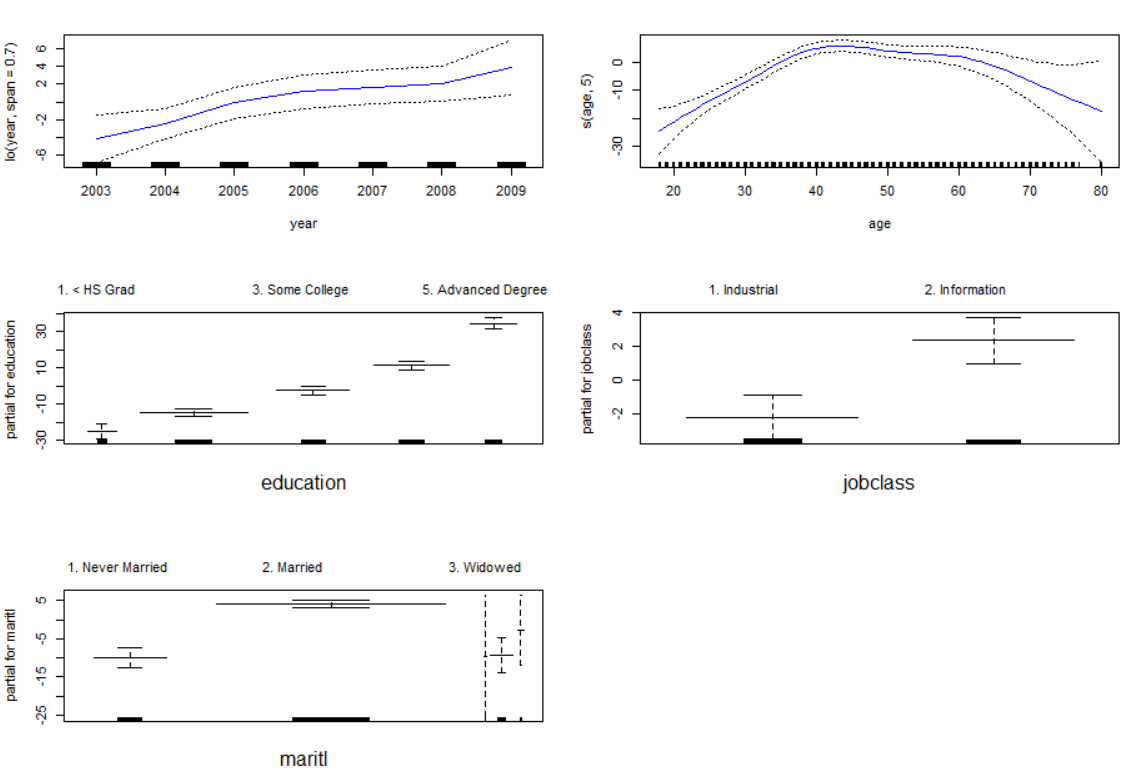


Отже, можна зробити висновок, що подружжя в середньому заробляє більше грошей ніж решта категорій, а також, що працівник в інформаційній сфері в середньому заробляє більше ніж в індустріальній.

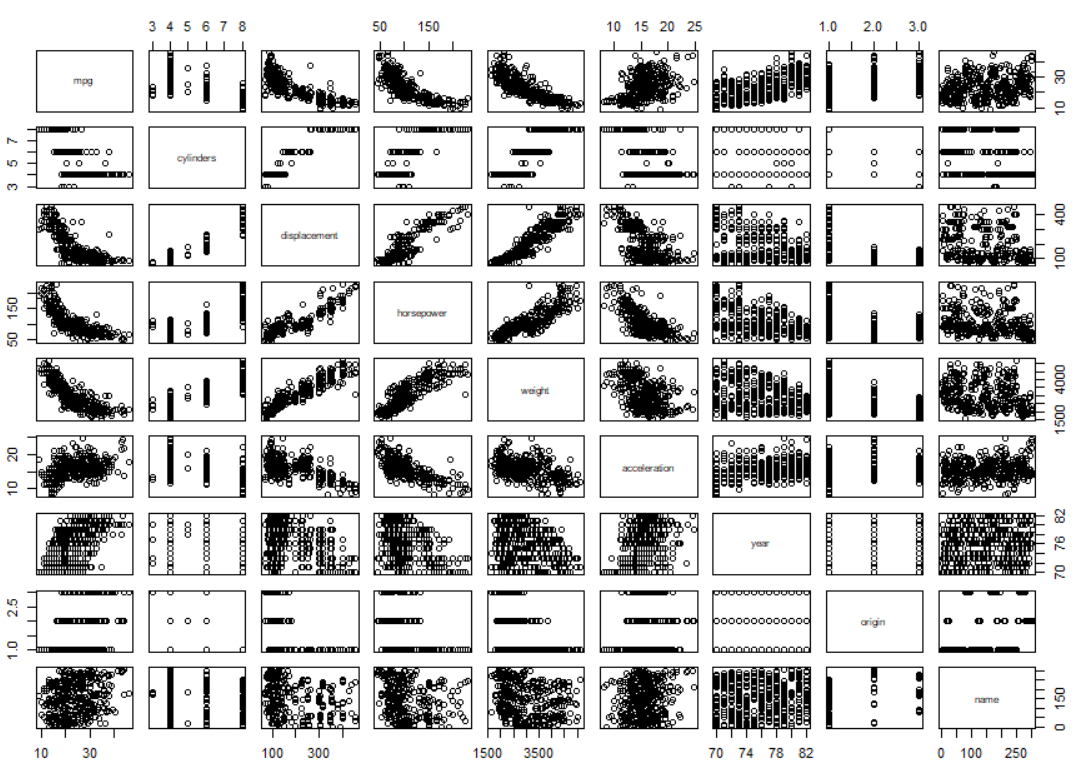
Зображення, що містить стіл

Автоматично згенерований опис

З результатів видно, що найкраще підходить третя модель.



**3. Auto**

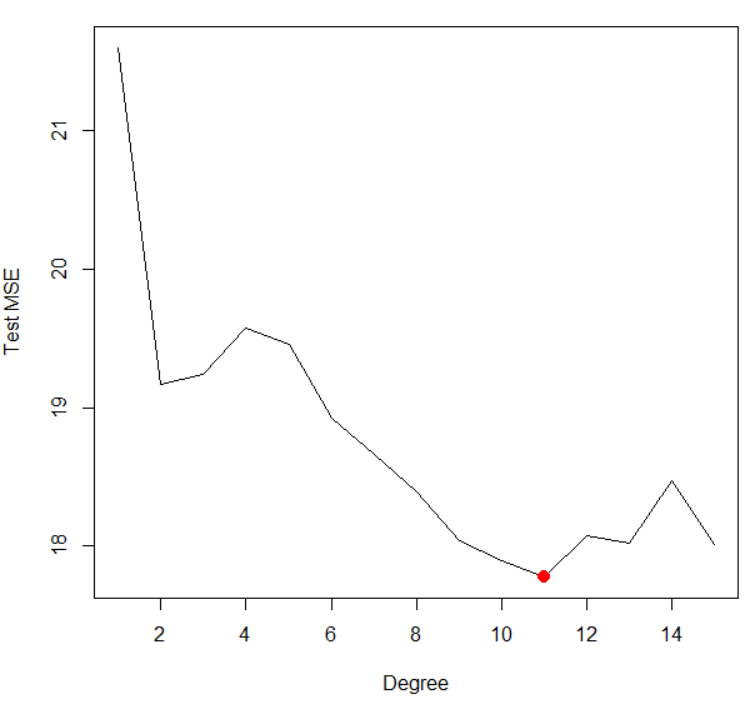


З графіків пар можна легко побачити, що mpg негативно корелює з cylinders, displacement, horsepower та weight.

Зображення, що містить текст

Автоматично згенерований опис

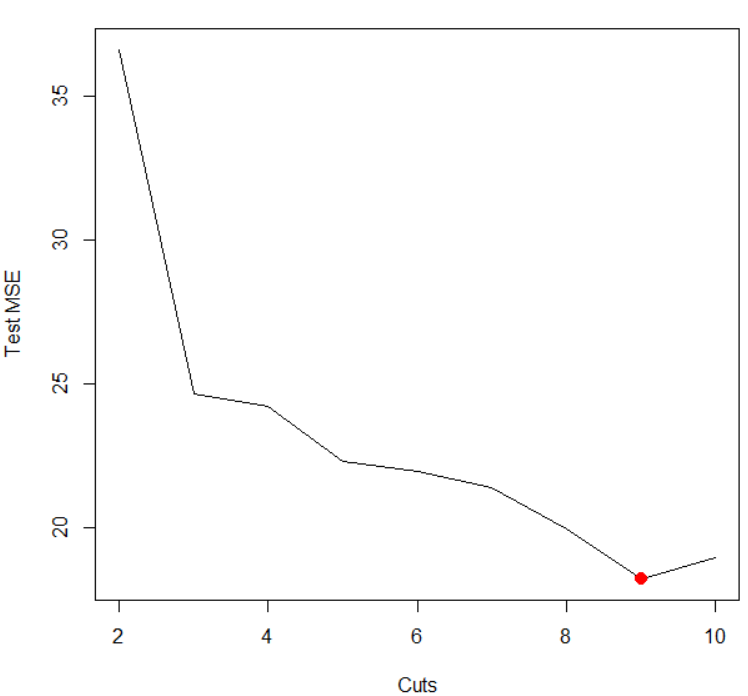
Зважаючи на графік наведений нижче можемо побачити, що d=11 є оптимальним степенем для полінома.



Зображення, що містить текст

Автоматично згенерований опис

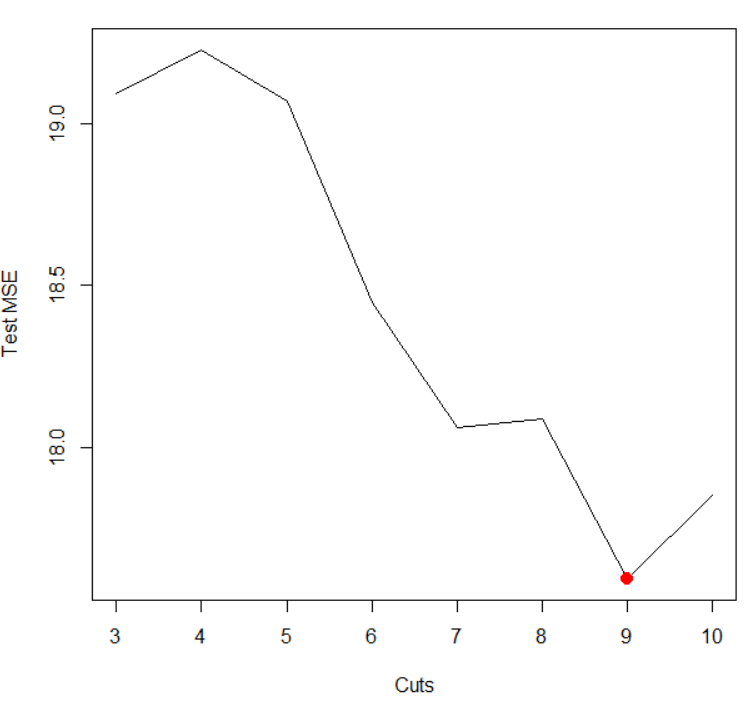
Зважаючи на графік наведений нижче можемо побачити, що помилка буде мінімальною для 9 зрізів.



Зображення, що містить текст

Автоматично згенерований опис

Зважаючи на графік наведений нижче, можемо побачити, що похибка мінімальна для 9 ступенів свободи.

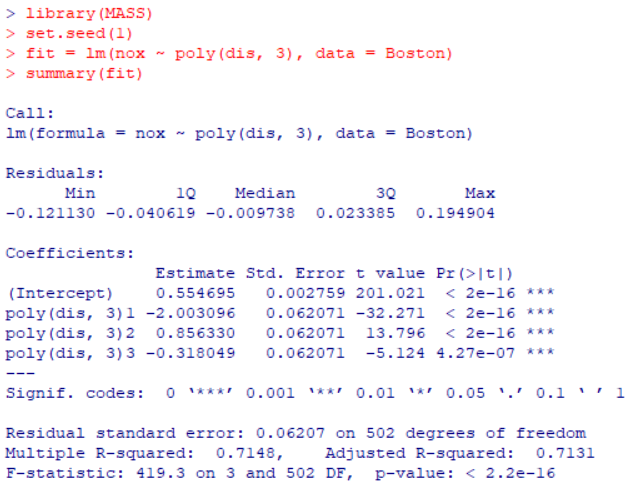


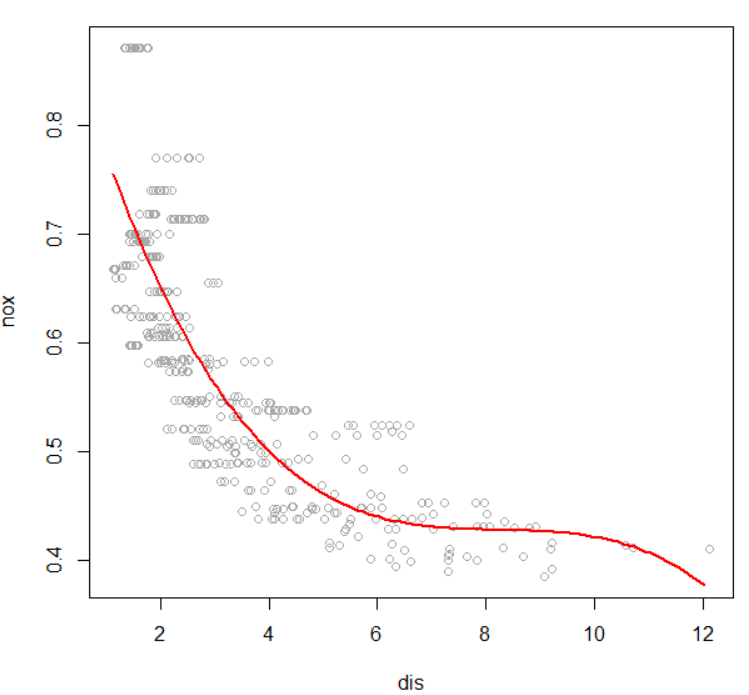
Зображення, що містить текст

Автоматично згенерований опис

**4. Boston**

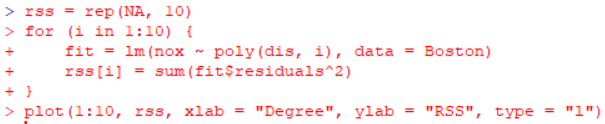
**4.1**

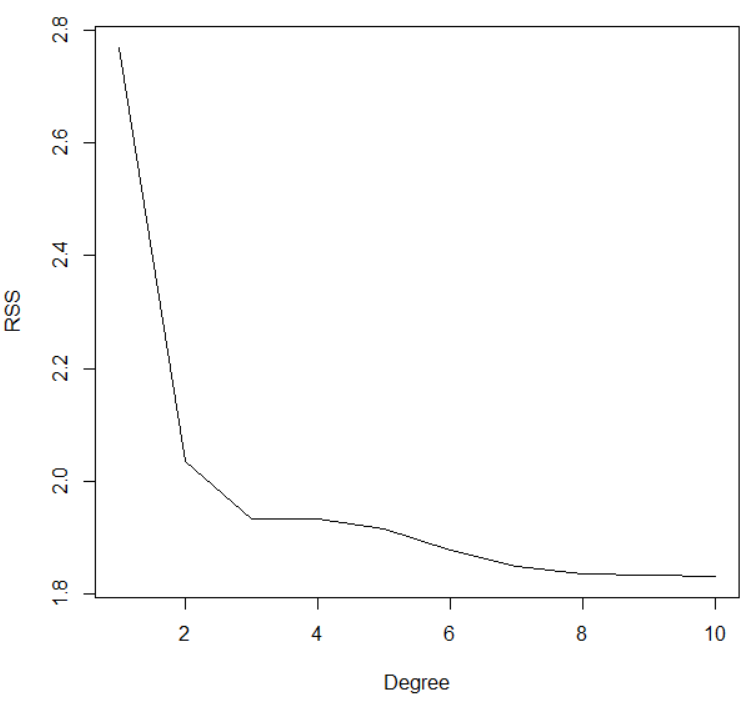




З наведених вище результатів можна сказати, що всі доданки в поліномі є значущими.

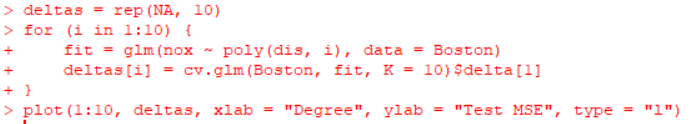
**4.2**

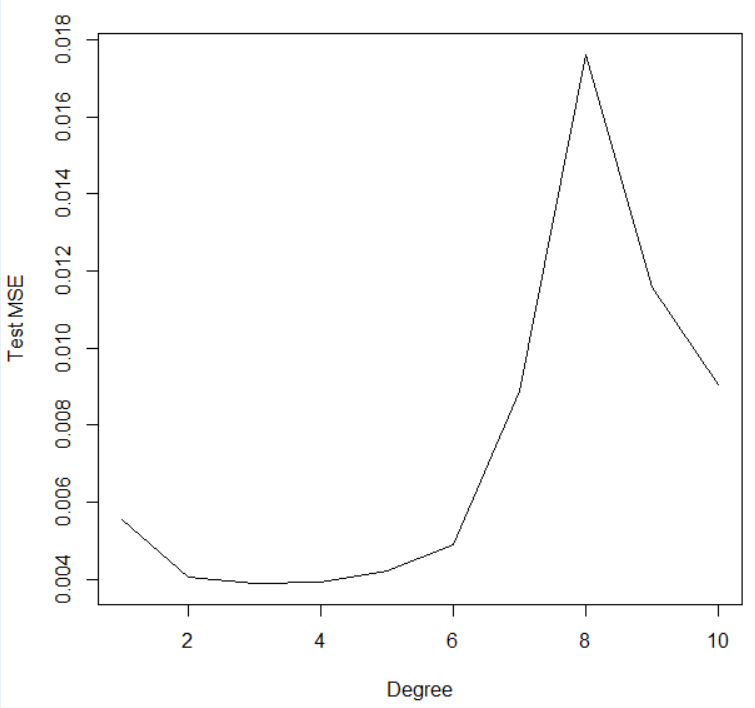




З графіка, можна сказати, що RSS зменшується зі збільшенням степеня полінома.

**4.3**



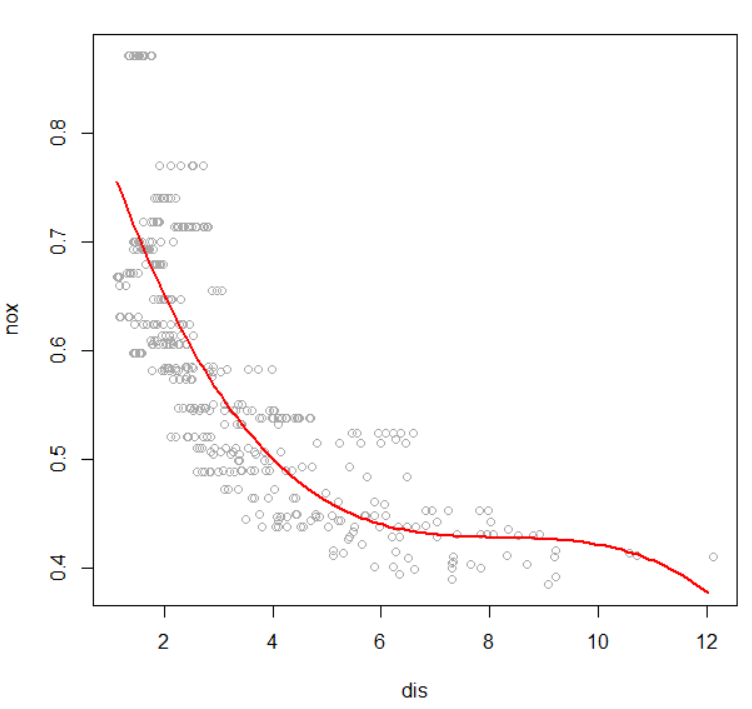


З графіка видно, що найменше MSE при степені полінома чотири.

**4.4**

Зображення, що містить стіл

Автоматично згенерований опис

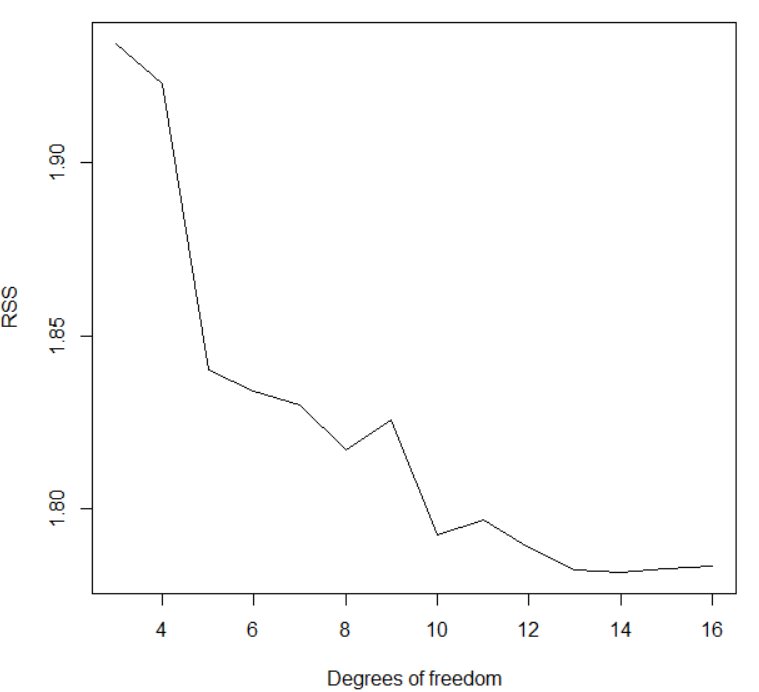


З наведених вище результатів можна сказати, що всі терми у сплайні є значущими.

**4.5**

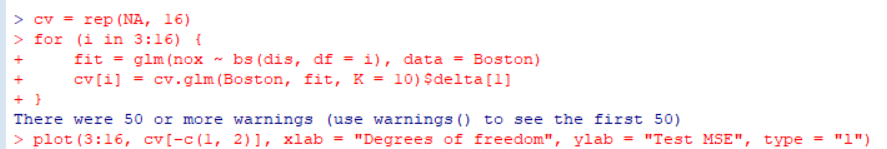
Зображення, що містить текст

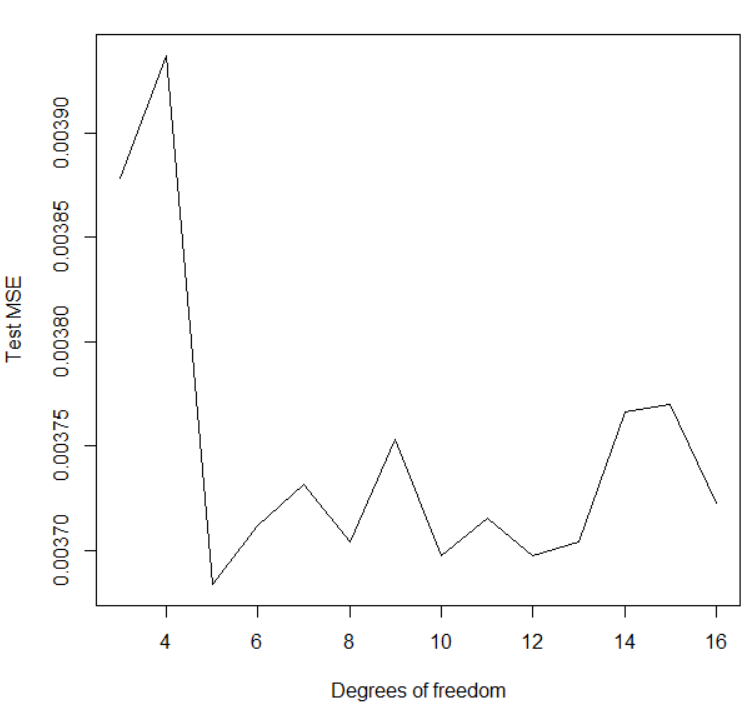
Автоматично згенерований опис



Можемо побачити, що RSS спадає до 14, а потім починає по трохи зростати.

**4.6**

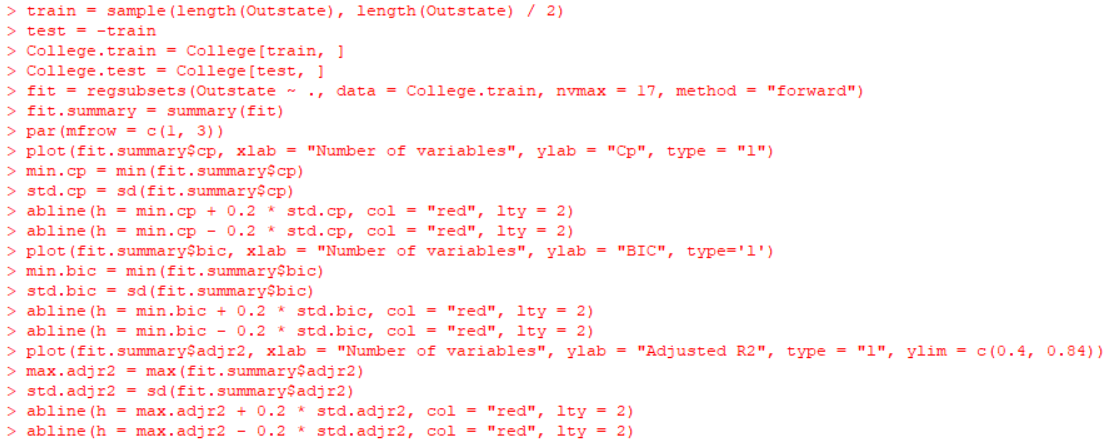


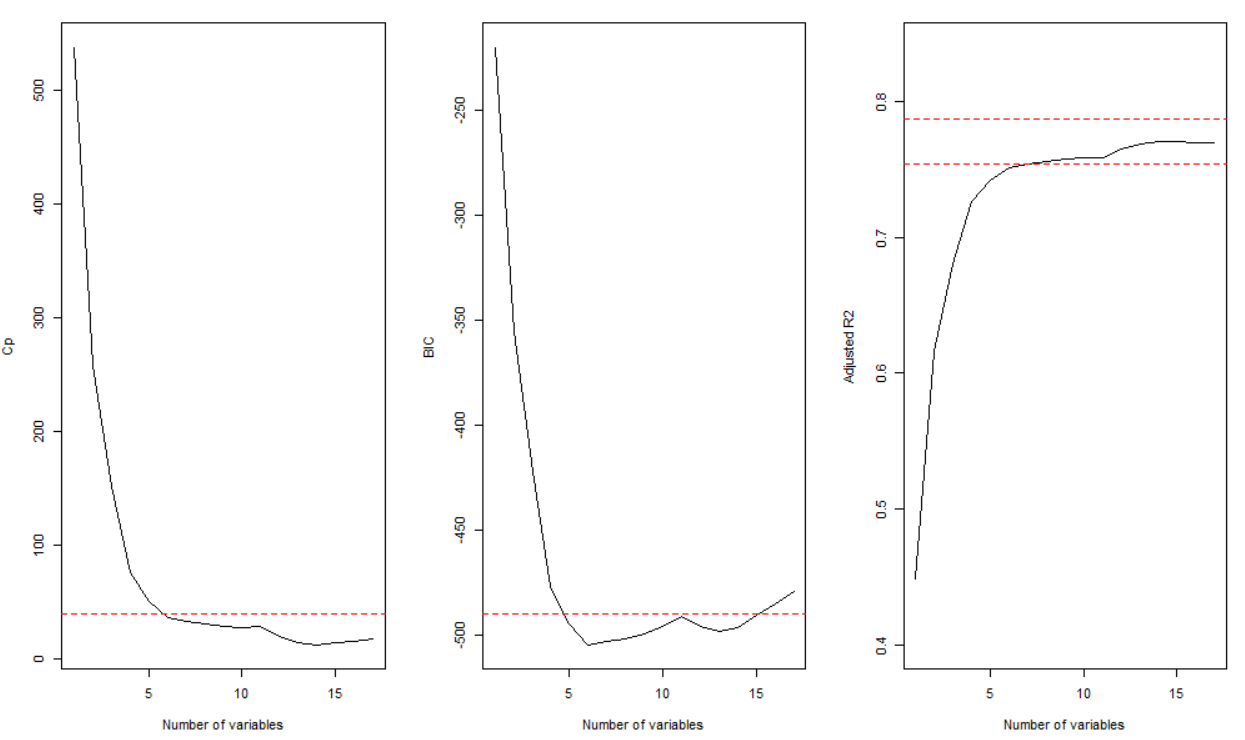


Тестова MSE мінімальна для п’яти ступенів свободи.

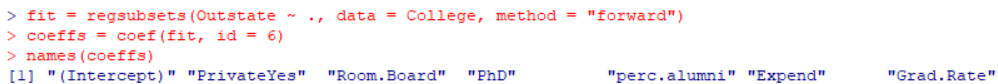
**5. College**

**5.1**

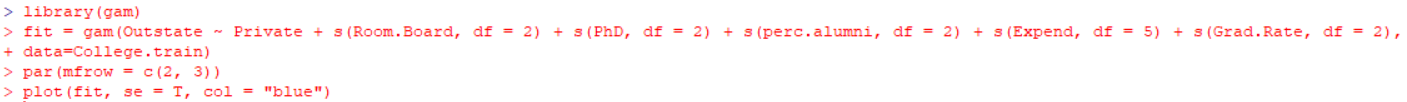


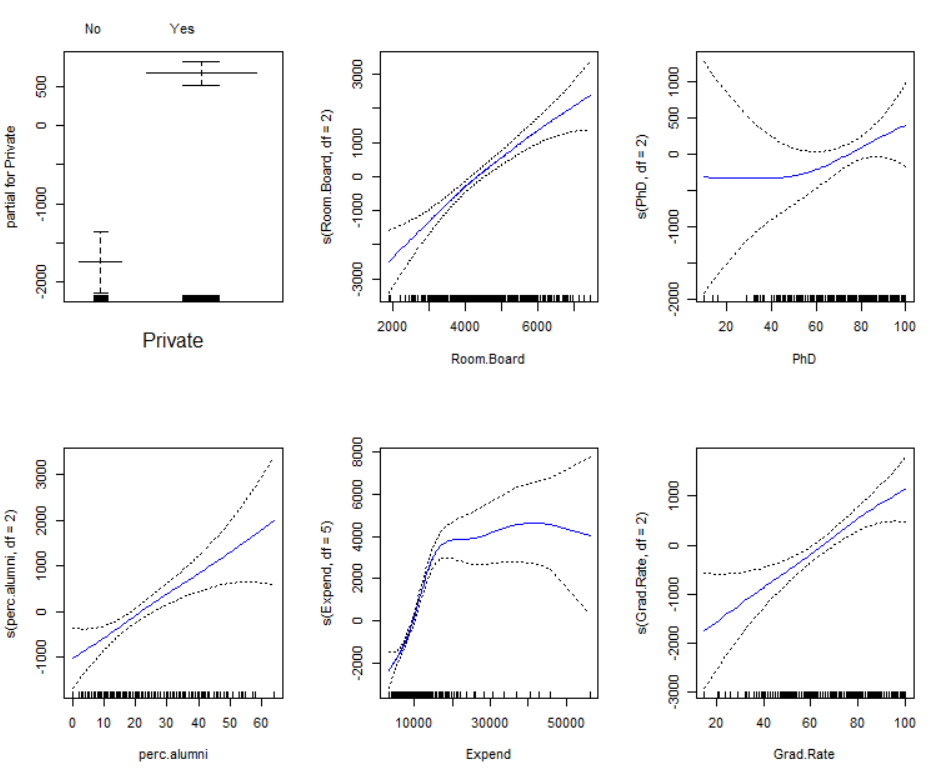


Cp, BIC і скорегований показують, що розмір 6 є мінімальним розміром для підмножини.

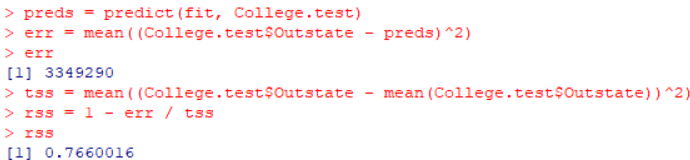


**5.2**





**5.3**



Ми отримуємо тест 0,766, використовуючи GAM з 6 предикторами.

**5.4**

Зображення, що містить текст

Автоматично згенерований опис

ANOVA показує чіткий не лінійний зв’язок між Outstate та Expend.

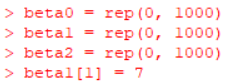
**6. Метод підгонки**

**6.1**

Зображення, що містить текст

Автоматично згенерований опис

**6.2**

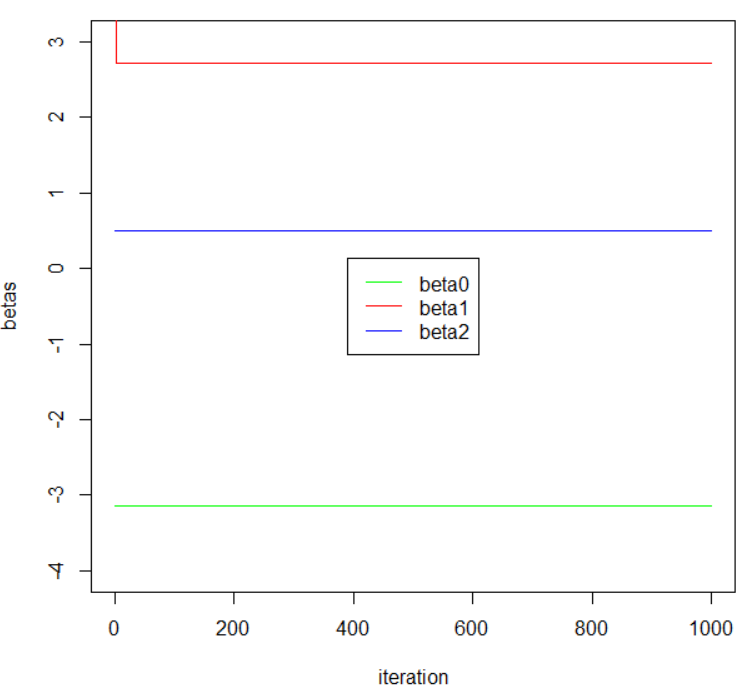


Візьмемо .

**6.3–6.5**

Зображення, що містить текст

Автоматично згенерований опис

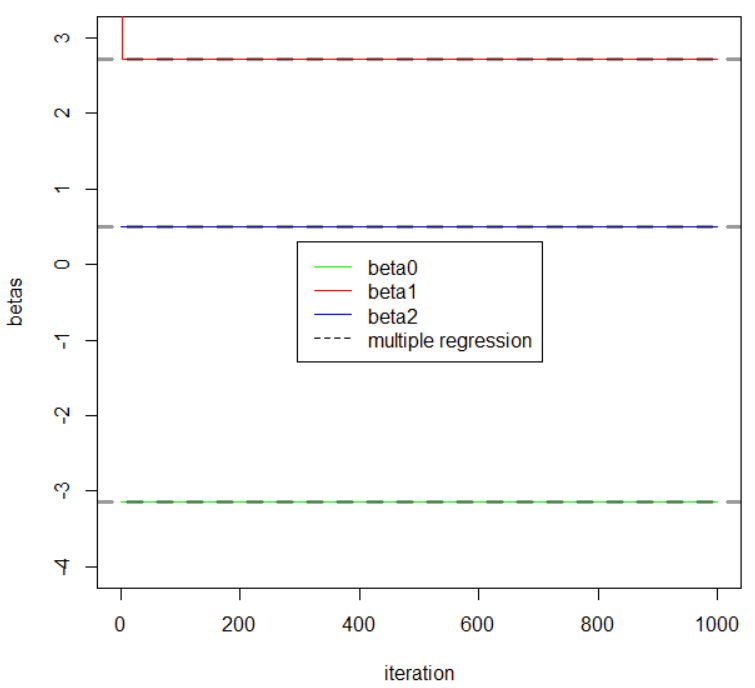


Коефіцієнти швидко досягають значень найменших квадртатів.

**6.6**

Зображення, що містить текст

Автоматично згенерований опис



Пунктирні лінії на графіку показують, що коефіцієнти множинної регресії точно збігаються з коефіцієнтами отриманими за допомогою backfitting-у.

**6.7**

Коли зв’язок між ігреком та іксами є лінійним, однієї ітерації достатньо для досягнення хорошого наближення до істинних коефіцієнтів регресії.

**7.**

Зображення, що містить текст

Автоматично згенерований опис

Десять ітерацій достатньо, щоб отримати хорошу апроксимацію визначену пороговим значенням суми квадратичних помилок між наступними ітераціями. Також видно, що похибка збільшується на 11-й ітерації.

