МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

ЗВІТ

до індивідуального завдання №4

з дисципліни «Моделі статистичного навчання»

Виконав

студент групи ПМіМ-12:

Зелінський Олександр

Перевірив:

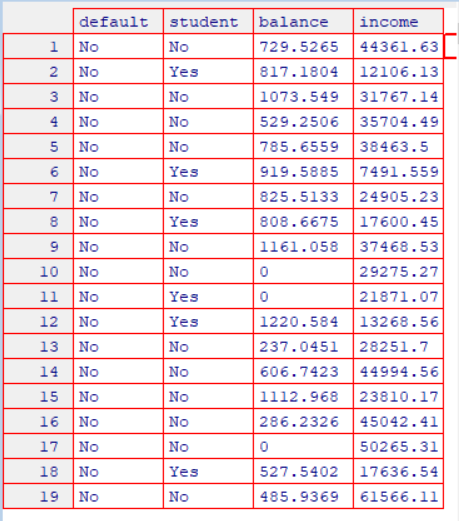
Проф. Заболоцький Т. М.

Львів – 2021

**Хід виконання**

**1. Метод валідаційного набору**

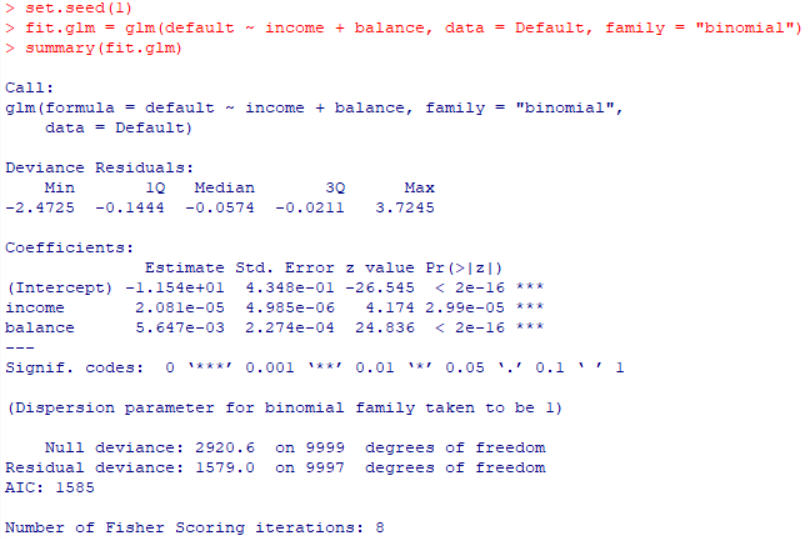
Приклад даних з Default та результат функції summary для них.



Зображення, що містить текст

Автоматично згенерований опис

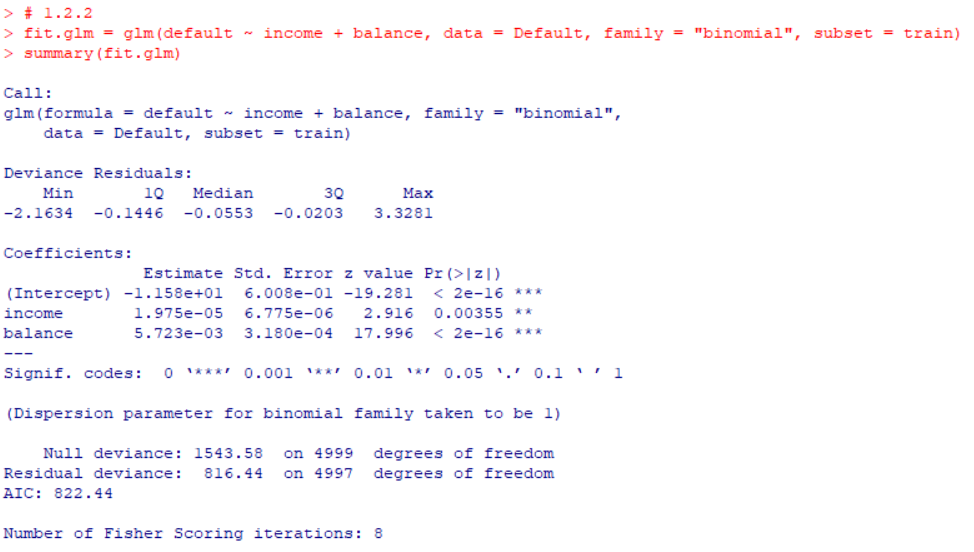
**1.1 Логістична регресія**



**1.2**



**1.2.2**

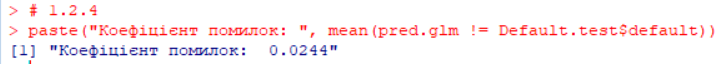


**1.2.3**

Зображення, що містить текст

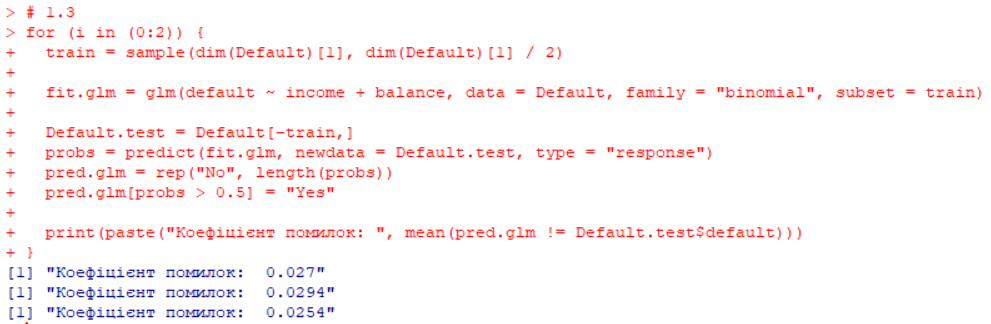
Автоматично згенерований опис

**1.2.4**



Можемо бачити що коефіцієнт тестової помилки буде 2,4% що є хорошим результатом.

**1.3**



Видно, що коефіцієнт тестової помилки змінюється в залежності від того, які спостереження потрапляють в навчальний набір, проте в цілому ці значення не сильно відрізняються

**1.4**

Зображення, що містить текст

Автоматично згенерований опис

Додавання фіктивної змінної student не покращило коефіцієнт тестової помилки, можна стверджувати, що він залишився в тих самих межах.

**2. Бутстрап**

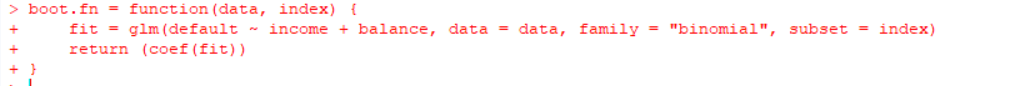
**2.1**

Зображення, що містить текст

Автоматично згенерований опис

Оцінки стандартних похибок за допомогою glm() для коефіцієнтів:

**2.2**



**2.3**

Зображення, що містить текст

Автоматично згенерований опис

Оцінки стандартних похибок за допомогою bootstrap для коефіцієнтів:

**2.4**

Як можемо побачити стандартні похибки досить близькі проте в bootstrap незначно нижчі.

**3. LOOCV**

**3.1**

Зображення, що містить текст

Автоматично згенерований опис

**3.2**

Зображення, що містить текст

Автоматично згенерований опис

**3.3**



З цього можемо зробити висновок, що перше спостережння “Up” було не правильно класифіковане, оскільки справжній напрямок “Down”

**3.4**

Зображення, що містить текст

Автоматично згенерований опис

**3.5**



Отже, бачимо, що оцінка LOOCV коефіцієнта тестової помилки рівна 41%, що є достатньо хорошим результатом.

**4.**

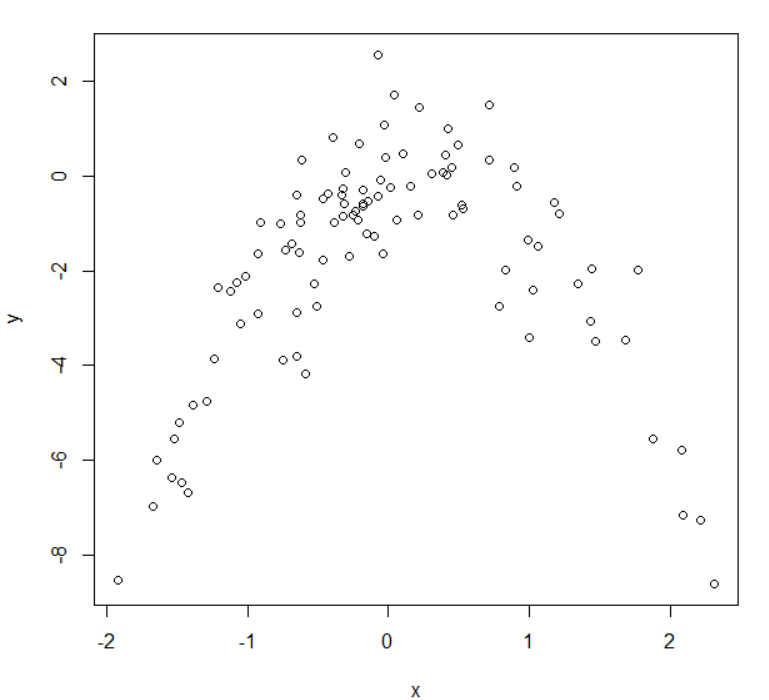
**4.1**

Зображення, що містить текст

Автоматично згенерований опис

У нашому випадку , а модель у формі рівняння матиме наступний вигляд:

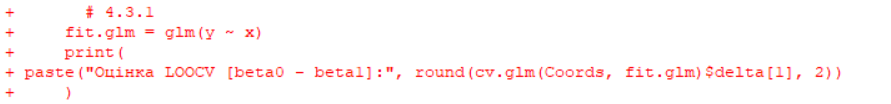
**4.2**



З графіка, очевидно, що в нас не лінійна залежність.

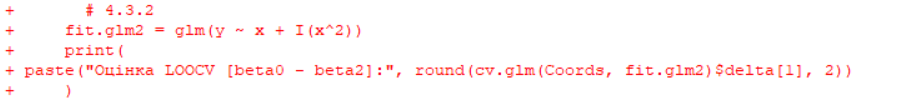
**4.3**

**4.3.1**



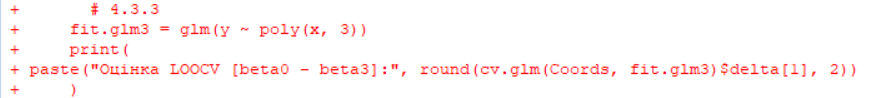


**4.3.2**



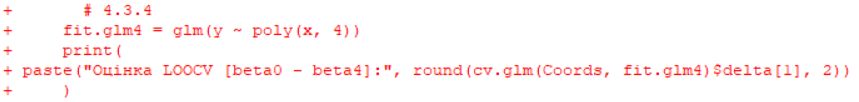


**4.3.3**





**4.3.4**





**4.4**

Зображення, що містить текст

Автоматично згенерований опис

Так результати такі самі, це пояснюється тим, що LOOCV оцінює n згорток одного спостереження.

**4.5**

Видно, що оцінка LOOCV для тестового мінімального квадратичного відхилення є найменшою для 4.3.2, що легко пояснюється тим, що наше відношення з 4.1 – квадратичне.

**4.6**

Зображення, що містить текст

Автоматично згенерований опис

Значення p вказують на те, що лінійний і квадратичний члени є статистично значущими, а кубічний і 4-го ступеня не є. Це, ясна річ, узгоджується з нашими результатами LOOCV, які були мінімальними для квадратичної моделі.

**5. Boston**

**5.1**

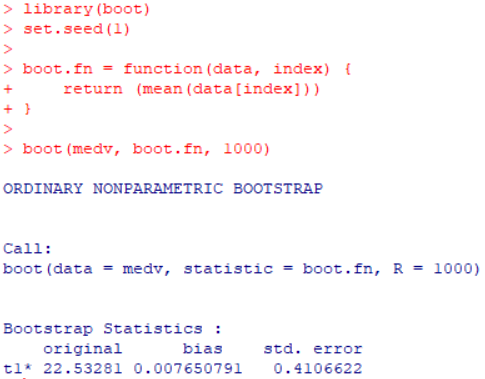


**5.2**



З результатів видно, що стандартна похибка 41%.

**5.3**



Зважаючи на отримані результати можна сказати, що стандартні похибки практично не відрізняються.

**5.4**

Зображення, що містить текст

Автоматично згенерований опис

Зображення, що містить текст

Автоматично згенерований опис

З результатів, можна сказати, що інтервал довіри для bootstrap є дуже близьким до того який видає функція t.test.

**5.5**



**5.6**

Зображення, що містить текст

Автоматично згенерований опис

Ми отримали таке саме значення як і медіана тобто 21,2 із відносно невеликою стандартною похибкою 0.377, що є невеликим в порівнянні із значенням медіани.

**5.7**

Зображення, що містить текст

Автоматично згенерований опис

**5.8**

Зображення, що містить текст

Автоматично згенерований опис

Ми отримали таке саме значення як і десятий процентиль тобто 12,75 із відносно невеликою стандартною похибкою 0.492, що є відносно невеликим в порівнянні зі значенням процентиля.