

## Індивідуальне завдання: нелінійні моделі.

Встановіть значення змінної `variant`: сума номера групи помноженого на 25 і порядкового номеру студента в списку групи (групі ПМОм-11 відповідає номер 0, групі ПМІм-11 відповідає номер 1, групі ПМІм-12 відповідає номер 2, групі ПМІм-13 відповідає номер 3). Далі встановіть `set.seed(variant)` та згенеруйте значення змінної `redundant` як заокруглене до цілого (для заокруглення можна використати функції `floor` або `round`) випадкове число з рівномірного на інтервалі (номер групи + 5, 25 – номер групи) розподілу (функція `runif`).

1. Модифікуйте дані `Auto` наступним чином: встановивши `seed`, що дорівнює значенню змінної `variant`, видаліть `redundant %` спостережень з допомогою функції `sample`. Розглянемо `horsepower` як предиктор та `mpg` як залежну змінну. Використовуючи функцію `poly()`, встановіть кубічну поліноміальну регресію для передбачення `mpg` за допомогою `horsepower`, опишіть результати регресії та побудуйте графік даних та поліноміальної регресії. Побудуйте поліноміальні моделі для різних степенів (наприклад, від 1 до 15), і наведіть їхні `RSS`. На основі перехресної перевірки виберіть «оптимальний» ступінь для поліноміальної регресії та поясніть отримані результати. Використовуючи функцію `bs()`, пристосуйте сплайн регресію для прогнозування `mpg` за допомогою `horsepower`. Опишіть результати отримані з використанням трьох вузлів та побудуйте графік даних та отриманої сплайн регресії. Пристосуйте сплайн регресію для діапазону ступенів свободи та наведіть відповідні `RSS`. На основі перехресної перевірки виберіть найкращий ступінь свободи для сплайн регресії на цих даних. Порівняйте найкращу поліноміальну регресію та найкращу сплайн регресію.

2. Застосуйте метод підгонки для моделі множинної лінійної регресії: зафіксуйте всі оцінки коефіцієнтів, крім одного на поточному значенні та оновлюємо лише одну оцінку коефіцієнта за допомогою простої лінійної регресії. Процес продовжується поки не досягнеться збіжність, тобто поки оцінки коефіцієнтів не перестануть змінюватися. Згенеруйте два предиктори  $X_1$  і  $X_2$  як вибірки з  $t$ -розподілу з  $[\text{variant}/15]$  ступенів свободи з  $n = 100$ , попередньо встановивши `seed`, що дорівнює значенню змінної `variant`, та обчисліть

$$Y = \text{variant} * X_1 + 2 * (\text{variant} + 1) * X_2 + \varepsilon,$$

де  $\varepsilon$  нормально розподілена випадкова величина з середнім 0 та дисперсією 1. Ініціалізуйте оцінку  $\beta_1$  випадковим значенням з  $\chi^2$  розподілу з  $[\text{variant}/\text{redundant}]$  ступенями свободи за заданого `seed`, що дорівнює значенню змінної `variant`. Не змінюючи  $\beta_1$  оцініть модель

$$Y - \beta_1 * X_1 = \beta_2 * X_2 + \varepsilon.$$

Зафіксувавши оцінку  $\beta_2$ , оцініть модель

$$Y - \beta_2 * X_2 = \beta_1 * X_1 + \varepsilon.$$

Використайте `for` для організації циклу з повторень попередніх кроків 100 разів. Наведіть оцінки параметрів  $\beta_1$  і  $\beta_2$  на кожній ітерації. Побудуйте графіки на яких відображено ці значення для  $\beta_1$  і  $\beta_2$  різними кольорами. Порівняйте отримані методом підгонки оцінки параметрів моделі з результатами оцінки множинної регресії  $Y$  на  $X_1$  і  $X_2$ .