

## Індивідуальне завдання: лінійна регресія.

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`. Побудуйте просту лінійну регресію з залежною змінною `mpg` і незалежною – `weight`. Виведіть та прокоментуйте результати з допомогою `summary()`. Зобразіть графічно предиктор та залежну змінну. Використовуючи функцію `plot()`, прокоментуйте які ви бачите проблеми.

2. Модифікуйте дані `Carseats`, аналогічно до завдання 1. Побудуйте модель множинної регресії для прогнозування `Sales` використовуючи `Price`, `Urban`, та `US`. Яку зі змінних можна вилучити з моделі? Побудуйте модель, яка використовує лише ті предиктори, для яких зв'язок з залежною змінною є істотним. Дослідіть отриману модель на наявність викидів та спостережень з високим рівнем леввереджу.

3. Виконайте наступні команди в R:

```
> set.seed(variant)
> x1=runif(100)
> x2=(variant/2)* x1+rnorm(100)*variant/10
> y=(2*variant)+variant* x1 +(variant/3)* x2+rnorm(100)
```

Яка кореляція між `x1` та `x2`? Оцініть регресію методом найменших квадратів, щоб передбачити `y`, використовуючи `x1` та `x2`. Чи можна відхилити нульову гіпотезу  $H_0: \beta_1 = 0$ ? Як щодо гіпотези  $H_0: \beta_2 = 0$ ? Побудуйте регресію `y` на `x1`. Чи можна відхилити нульову гіпотезу  $H_0: \beta_1 = 0$ ? Побудуйте регресію `y` на `x2`. Чи можна відхилити нульову гіпотезу  $H_0: \beta_2 = 0$ ? Поясніть отримані результати. Нехай одне додаткове спостереження було неправильно виміряно.

```
> x1 = c(x1, 0.1)
> x2 = c(x2, (variant/2)* 0.9)
> y = c(y, 5*variant)
```

Переоцініть попередні лінійні моделі, використовуючи ці нові дані. Дослідіть отриману модель на наявність викидів та спостережень з високим рівнем леввереджу.