

Індивідуальна 1. Вступ.

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. З допомогою функції `read.csv()` завантажте дані в R з файлу `College.csv`. Модифікуйте завантажені дані `College` наступним чином: встановивши `seed`, що дорівнює значенню змінної `variant`, видаліть `redundant %` спостережень з допомогою функції `sample`

```
(College_new =  
College[-  
sample(1:length(College[,1]), round((redundant/100)*length(College[,1])), )], )
```

Перегляньте дані з допомогою функції `fix()`. Оскільки перший стовпець це просто назви університетів, виконайте наступні команди:

```
rownames=College_new[,1]  
College_new=College_new[,-1]  
fix(College_new)
```

Що тепер сталося з першим стовпцем? Для чого змінна `rownames`?

Використовуючи функції `summary()` та `pairs()`, отримайте певні підсумки відносно даних. Використайте `pairs()` для побудови матриці графіків довільних 3 стовпців даних.

Побудуйте діаграму `Outstate vs Private`, використовуючи `plot()` та прокоментуйте результати.

Створіть новий якісний показник `Elite`, використовуючи `Top10perc`, тобто поділивши всі університети на дві групи в залежності чи перевищує відсоток студентів з топ 10% шкіл 50% чи ні.

```
> Elite=rep("No", nrow(College_new))  
> Elite[College_new$Top10perc>50]="Yes"  
> Elite=as.factor(Elite)  
> College_new=data.frame(College_new, Elite)
```

Для чого дві останні команди? Чи багато таких університетів? Побудуйте діаграму `Outstate vs Elite` та прокоментуйте результати.

2. Завантажте дані `Boston`, які є частиною бібліотеки `MASS`.

```
> library(MASS)
```

Дані знаходяться в об'єкті `Boston`.

```
> Boston
```

Перегляньте інформацію про дані:

> ?Boston

Скільки рядків і стовпців міститься в множині?

Модифікуйте завантажені дані Boston наступним чином: встановивши seed, що дорівнює значенню змінної variant, видаліть redundant % спостережень з допомогою функції sample().

Скільки кварталів в даній множині межують з річкою Charles?

Обчисліть медіану для відношення учні-вчителі для міста загалом?

Які квартали міста мають найменше та найбільше відношення учні-вчителі?

В яких кварталах в середньому є більше 7 кімнат в помешканні? Більше 8?

Більше 4, але менше 7?