

## Bài ôn tập buổi 5&6: Canxi

Một sinh viên học khoa dinh dưỡng muốn xem xét mối quan hệ giữa lượng canxi (calcium\_intake g/day) và kiến thức về canxi knowledge\_score (thang đo 0 đến 5) ở các sinh viên trong lớp học. Và cô ấy muốn biết liệu kiến thức về canxi có thể được sử dụng để dự đoán lượng canxi của sinh viên hay không. Cô ấy đã thu thập dữ liệu và lưu trong tập tin canxi.xlsx.

**Hãy thực hiện các yêu cầu được liệt kê sau đây:**

1. Đọc dữ liệu. Xem thông tin dữ liệu với head(), tail(), str(), summary().
2. Vẽ biểu đồ phân phối tần suất của knowledge\_score. Nhận xét.
3. Thực hiện các thống kê cơ bản cho knowledge\_score và calcium\_intake (mean, median, mode, max, min, range)
4. Cho biết các giá trị ở phân vị thứ 20, 35, 65 và 90 của calcium\_intake. Biểu diễn phân vị và giá trị tương ứng trên biểu đồ.
5. Vẽ boxplot cho knowledge\_score và cho calcium\_intake. knowledge\_score có outlier hay không? calcium\_intake có outlier hay không? Nếu có thì hãy thay thế outlier này bằng giá trị phù hợp. Giá trị bạn chọn để thay thế là giá trị nào (mean/median/...)? Giải thích.
6. Tìm phương sai (variance) của knowledge\_score và calcium\_intake.
7. Tìm độ lệch chuẩn (Standard deviation) của knowledge\_score và calcium\_intake.
8. Tìm độ xiên (Skewness) và độ nhọn (Kurtosis) của knowledge\_score. Nhận xét kết quả.
9. Vẽ biểu đồ thể hiện mối quan hệ giữa knowledge\_score và calcium\_intake. Nhận xét.
10. Tính giá trị covariance, correlation giữa knowledge\_score và calcium\_intake.

```
In [ ]: library("xlsx")
library("ggplot2")
library("ggExtra")
library("tidyverse")
library("mlbench")
library("e1071")
```