

# BÀI TẬP THÊM 1

Hạn nộp bài: hết ngày 05/05/2024

- Làm bài trên **R script**, lưu lại với tên có dạng: "LTTK\_MSSV\_HoTen\_BT1.R".
- Kiểm tra kĩ file, chỉ nộp **một lần duy nhất**. Link nộp bài: [link google form](#).
- Bài làm cần trình bày như sau:

```
##
## Bai tap them 1 - Thuc hanh Ly thuyet Thong ke
## Lop ... ca ...
##
## Ho ten: ..... - MSSV: .....
##
##*****
## Bai 1:
(phan bai lam cua Bai 1)
##-----
## Bai 2:
(phan bai lam cua Bai 2)
##-----
.....
##*****
## Ket thuc
```

- Cách tính điểm cộng:
  - Có 2 bài tập, tương ứng tối đa 4 điểm. Làm hết 1 bài mới được cộng điểm. Làm nửa bài không tính.
  - Chỉ cộng tối đa 50% số điểm còn thiếu để được 10. Ví dụ: 4 được cộng tối đa thành 7, 5 được cộng tối đa thành 7.5 ...

## Bài 1 (2đ)

Cho biến ngẫu nhiên  $X$  nhận các giá trị  $\{0; 1; 2; \dots; n\}$  (với  $n \in \mathbb{N}$ ,  $n \geq 1$ ) và có hàm trọng lượng xác suất được xác định bởi

$$\mathbb{P}(X = k) = C_n^k \cdot p^k \cdot (1 - p)^{n-k},$$

với  $k = 0, 1, 2, \dots, n$  và  $0 < p < 1$ .

**1.1.** Viết hàm `mass.prob(k,n,p)` để tính hàm trọng lượng xác suất của  $X$ , với  $n, k, p$  tổng quát.

- 1.2.** Viết hàm `cdf(k,n,p)` tính phân phối xác suất tích lũy của  $X$ , tức là ứng với hàm  $F$  được xác định bởi, với  $x \in \mathbb{R}$ :

$$F_X(x) \equiv F(x) := \mathbb{P}(X \leq x).$$

- 1.3.** Áp dụng cho  $n = 10$  và  $p = 0.25$ .

- Dùng hàm `cdf(k,n,p)` để tính  $F_X(5)$  và  $F_X(8.5)$ .
- Tính kỳ vọng  $\mathbb{E}(X)$  và phương sai  $Var(X) := \mathbb{E}[(X - \mathbb{E}X)^2]$ .
- Vẽ đồ thị phù hợp thể hiện hàm trọng lượng xác suất của biến ngẫu nhiên  $X$ .
- Vẽ đồ thị phù hợp thể hiện hàm phân phối xác suất  $F_X$  của biến ngẫu nhiên  $X$ .

## Bài 2 (2đ)

File `diesel_engine.csv` và `diesel_time.csv` chứa số liệu về hoạt động của các động cơ chạy bằng dầu diesel. Thực hiện:

- 2.1.** Đọc số liệu từ hai file này, gán vào hai dataframe, đặt tên hai dataframe cùng tên với file. Liệt kê tên các biến có trong hai dataframe vừa nhập.
- 2.2.** Xác định có bao nhiêu dữ liệu bị khuyết (missing data) trong `diesel_engine`. Thay thế các giá trị khuyết trong biến `speed` bằng 1500, biến `load` bằng 20.
- 2.3.** Tính: trung bình, phương sai, độ lệch tiêu chuẩn, giá trị lớn nhất, nhỏ nhất của biến `alcohol` trong dataframe `diesel_engine`.
- 2.4.** Ghép hai dataframe lại thành một dataframe có tên là `diesel`.
- 2.5.** Trích giá trị của biến `run` (số thứ tự các động cơ) mà có thời gian trễ (biến `delay`) dưới 1.000.
- 2.6.** Đếm xem có bao nhiêu động cơ có `timing` bằng 30.
- 2.7.** Vẽ biểu đồ boxplot cho các biến `speed`, `timing` và `delay`.
- 2.8.** Vẽ biểu đồ phân tán cho các cặp biến (`timing`, `speed`), (`temp`, `press`).
- 2.9.** Chia phạm vi giá trị của biến `delay` thành 4 đoạn đều nhau và đếm số giá trị nằm trong các đoạn đó. Tạo bảng thống kê và vẽ biểu đồ cột.
- 2.10.** Chia phạm vi giá trị của biến `delay` thành 4 đoạn như sau:  $(0.283, 0.7]$ ,  $(0.7, 0.95]$ ,  $(0.95, 1.2]$ ,  $(1.2, 1.56]$ . Tạo bảng thống kê và vẽ biểu đồ cột