

0.1 Lập trình cơ bản trong R

Bài tập 0.1.1. Cho điểm của 40 sinh viên khoa Toán ngẫu nhiên từ $0 \rightarrow 10$. Dùng vòng lặp **for** tính kì vọng và phương sai của mẫu vừa cho.

Bài tập 0.1.2. Trong bài tập này ta mô phỏng việc tung ngẫu nhiên một con xúc sắc. Giả sử rằng trong trò chơi tung xúc sắc, bạn cược rằng có ít nhất một mặt sáu khi gieo ngẫu nhiên n con xúc sắc cân đối đồng chất trong một lần chơi

- Hãy viết hàm `sixes(n)` để mô phỏng kết quả một lần chơi, trong đó kết quả xuất ra là `TRUE` nếu có một mặt sáu trong số n con xúc sắc được gieo và `FALSE` nếu ngược lại.
- Hãy dùng hàm `sixes` bạn viết ở ý (a) để viết hàm `play(n,N)` để mô phỏng kết quả N lần chơi (trong đó mỗi lần bạn đánh cược là có ít nhất một mặt sáu trong số tổng số n con xúc sắc được gieo). Kết quả xuất ra của bạn là tỉ lệ thắng cuộc của bạn trong số N lần chơi.

Chạy 2 hàm vừa tạo với $n = 5, N = 20$.

Bài tập 0.1.3. Bộ dữ liệu `giamcan.csv` cho thấy sự thay đổi cân nặng trước và sau khi thực hiện giảm cân. Đọc dữ liệu và dùng vòng lặp `for` để tính các yêu cầu sau:

- Đếm số người có cân nặng dưới 65kg trước và sau khi giảm cân. Tính tỉ lệ người có cân nặng dưới 65kg trước và sau khi giảm cân?
- Đếm số người có cân nặng từ 65kg đến 75kg sau khi giảm cân. Tính tỉ lệ người có cân nặng từ 65kg đến 75kg ?

Bài tập 0.1.4. File `buswaitingtime.xls` chứa số liệu về thời gian phải đợi của khách đi xe buýt. Thực hiện các bước sau:

- Đọc số liệu từ `buswaitingtime.xls` và gán vào 1 dataframe có tên `wtime`. Sau đó, chiết dữ liệu sang biến có tên `time`.
- Viết hàm xuất ra các giá trị sau: thời gian đợi ngắn nhất, dài nhất, thời gian đợi trung bình.
- Biết rằng, nếu thời gian đợi từ 15 phút trở lên, khách hành sẽ phàn nàn về dịch vụ đi lại bằng xe buýt. Dùng lệnh **if** và vòng lặp **for** để đếm số khách có khả năng phàn nàn trên.

Bài tập 0.1.5. Cho bộ dữ liệu sau (có tổng cộng 20 số)

$X = (1.3975068, 1.1253621, 0.7472591, 0.8144085, 1.8202663,$
 $0.8671346, 1.4210106, 1.67840051, 1.3697046, 0.6182268,$
 $1.9064036, 0.9360782, 1.1175578, 1.2527916, 1.9584229,$
 $1.3689436, 1.5685716, 1.0691046, 1.6625395, 1.7949509)$

Tính các giá trị sau của bộ dữ liệu: $\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$

Bài tập 0.1.6. Tung một đồng xu cân đối đồng chất, giả sử bạn tung được mặt sấp sẽ trả ra kết quả là 1 và ngược lại tung được mặt ngửa thì kết quả là 0.

- (a) Viết hàm mô phỏng kết quả của n lần tung một đồng xu cân đối đồng chất,
- (b) Nếu bạn tung đồng xu được mặt sấp bạn sẽ nhận được 10\$, nếu được mặt ngửa bạn bị mất 5\$. Viết hàm tính số tiền bạn được khi thực hiện m lần tung. Giả sử xác suất nhận được mặt sấp khi tung đồng xu của nhà cái là 0.35.
- Bài tập 0.1.7.** (a) Hãy sử dụng vòng lặp **for** để viết hàm `meanN` theo cấu trúc `meanN = function(x, n)` để tìm trung bình của n phần tử đầu tiên của bộ dữ liệu (vector) x . Hàm cần kiểm tra điều kiện nhằm đảm bảo rằng n phải không vượt quá kích thước bộ dữ liệu.
- (b) Tạo ra vector x là điểm Thực hành xác suất thống kê của 20 sinh viên trường ĐH KHTN (được phát sinh ngẫu nhiên) với thang điểm từ 0 đến 10.
Sử dụng dữ liệu này để thực hiện các yêu cầu bên dưới.
- (c) Tìm điểm trung bình của 20 sinh viên trên.
- (d) Dùng “toán tử” \wedge để tìm x^2 . Từ đó, dùng lệnh `meanN` để tìm phương sai (chưa hiệu chỉnh) theo công thức $s_x^2 = \overline{x^2} - (\bar{x})^2$.
- (e) So sánh kết quả ở (c-d) với kết quả từ lệnh `mean`, `var` của phần mềm **R**; và đưa ra nhận xét. Lý giải cho nhận xét.

0.2 Biến ngẫu nhiên

Bài tập 0.2.1. Cho vecto X chứa n giá trị quan sát, phân vị thứ p được xác định

- Sắp xếp dữ liệu theo thứ tự tăng dần (từ nhỏ đến lớn).
- Tính chỉ số i : $i = \left(\frac{p}{100}\right)n$
- Nếu i không phải là số nguyên, làm tròn i . Phân vị thứ p chính là giá trị nằm ở vị trí thứ i đã được làm tròn.
- Nếu i nguyên, phân vị thứ i chính là giá trị trung bình của 2 giá trị nằm ở vị trí thứ i và thứ $i + 1$.

Mô phỏng 1 ví dụ cho vecto X và tìm mức phân vị p của vecto X .

Bài tập 0.2.2. Chiều cao trung bình hiện nay ở phụ nữ Việt Nam tuân theo phân phối chuẩn là 156cm, với độ lệch chuẩn là 4.6cm

- Vẽ hàm mật độ của phụ nữ Việt Nam trong khoảng $[130, 190]$
- Tính xác suất 1 phụ nữ Việt Nam có chiều cao 160cm
- Vẽ hàm phân phối tích lũy của phụ nữ Việt Nam trong khoảng $[130, 190]$
- Tính xác suất chiều cao phụ nữ Việt nam thấp hơn hay bằng 150cm

Bài tập 0.2.3. Các nghiên cứu chứng tỏ rằng việc sử dụng nhiên liệu cho các xe hơi cỡ nhỏ bán tại Hoa Kỳ là có phân phối chuẩn, với mức sử dụng trung bình là 30,5 dặm/galông nhiên liệu(mpg) và một độ lệch chuẩn là 4,5 mpg,

- Vẽ hàm phân phối xác suất của việc sử dụng nhiên liệu cho các xe hơi cỡ nhỏ bán tại Hoa Kỳ [20, 45]
- Tính xác suất nhiên liệu một xe hơi sử dụng là 29 dặm/galông.
- Vẽ hàm phân phối tích lũy của việc sử dụng nhiên liệu cho các xe hơi cỡ nhỏ bán tại Hoa Kỳ [20, 45]
- Tính xác suất nhiên liệu xe hơi sử dụng thấp hơn hay bằng 32 dặm/galông.

Bài tập 0.2.4. Số khách hàng phải đợi khi đến giao dịch tại một cây ATM trong một giờ tuân theo phân phối Poisson với tham $\lambda = 4$. Gọi X là số khách hàng phải đợi, khi đó $X \sim P(4)$. Tính xác suất để trong một giờ số người đợi không quá 2. Biết rằng, biến ngẫu nhiên $X \sim P(\lambda)$ thì

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

- Viết hàm để tính xác suất $\mathbb{P}(X \leq 2)$.
- Vẽ đồ thị hàm phân phối tích lũy.
- Tính phân vị mức 0.75 của X .

Bài tập 0.2.5. Cho biến ngẫu nhiên có hàm mật độ như sau:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}, \quad 0 < x < \infty$$

với α và β là hai tham số.

- Viết hàm mật độ xác suất của X với $\alpha = 2, \beta = 5$ và kiểm tra xem tích phân có bằng 1 hay không? Tính giá trị $f(x)$ tại điểm $x = 4$.
- Vẽ hàm mật độ xác suất của X
- Tính $F_x(2.75)$
- Vẽ hàm phân phối xác suất của X
- Tính phân vị ở mức 0.75 của X

Bài tập 0.2.6. Cho hàm mật độ của biến ngẫu nhiên X , $f_X(x) = 2e^{-2x}$ với $x \geq 0$.

- Vẽ hàm mật độ của A
- Viết hàm phân phối tích lũy (c.d.f) của X , và vẽ đồ thị của hàm c.d.f
- Tính $P(X \leq 5)$

Bài tập 0.2.7. Cho $X \sim B(n = 20, p = 0.1)$.

- Tìm các xác suất $\mathbb{P}(X = 2)$, $\mathbb{P}(X < 3)$, và $\mathbb{P}(X \geq 4)$.
- Vẽ đồ thị hàm mật độ f_X .
- Vẽ đồ thị hàm phân phối tích lũy F_X .