

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC BÁCH KHOA  
KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH



## CẤU TRÚC RỜI RẠC CHO KHMT (CO1007)

---

Ứng dụng thống kê  
khảo sát kết quả của bài tập online cho phép nộp bài nhiều lần

GVHD: Huỳnh Tường Nguyên  
Nguyễn Ngọc Lễ  
SV thực hiện: Nguyễn Văn A – 22102134  
Trần Văn B – 88471475  
Lê Thị C – 36811334  
Phạm Ngọc D – 97501334  
Kiều Thị E – 12341334

Tp. Hồ Chí Minh, Tháng 05/2020



## Mục lục

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Động cơ nghiên cứu</b>              | <b>2</b> |
| <b>2</b> | <b>Mục tiêu</b>                        | <b>2</b> |
| <b>3</b> | <b>Mô tả dữ liệu</b>                   | <b>2</b> |
| <b>4</b> | <b>Nhiệm vụ</b>                        | <b>3</b> |
| <b>5</b> | <b>Hướng dẫn và yêu cầu</b>            | <b>5</b> |
| 5.1      | Hướng dẫn . . . . .                    | 5        |
| 5.2      | Yêu cầu . . . . .                      | 5        |
| 5.3      | Nộp bài . . . . .                      | 6        |
| <b>6</b> | <b>Cách đánh giá và xử lý gian lận</b> | <b>6</b> |
| 6.1      | Đánh giá . . . . .                     | 6        |
| 6.2      | Xử lý gian lận . . . . .               | 6        |
|          | <b>Tài liệu</b>                        | <b>6</b> |

## 1 Động cơ nghiên cứu

Khi bạn đang đi học, có những quyết định khó khăn để đưa ra; thật khó để chọn con đường sự nghiệp phù hợp. Bạn có thể không nhận thức được tất cả các lựa chọn của mình, nơi bạn sẽ thành công và cuối cùng, bạn sẽ thích làm gì. Từ đó Trường Đại học Bách Khoa, ĐHQG-HCM kết hợp cùng các công ty trong cũng như ngoài nước để tạo ra môi trường công nghiệp thực sự để giúp các bạn sinh viên thực tập. Bạn sẽ có cơ hội khám phá những gì bạn yêu thích, kết nối với đồng nghiệp, người cố vấn và chuyên gia trong ngành. Đó là cách tốt nhất để có được một chỗ ngồi tại công ty hay thực hiện bước đầu tiên để bắt đầu một sự nghiệp tuyệt vời. Sau khi chương trình kết thúc công ty sẽ cung cấp bảng nhận xét đánh giá các sinh viên qua các tiêu chí cụ thể.

Phân tích & thống kê dữ liệu qua bảng đánh giá sinh viên của công ty không những giúp khoa máy tính có những hướng đúng trong việc phát hiện ra những kiến thức mà sinh viên chưa chắc chắn hay thiếu sót trong đào tạo, cũng như có hướng để cải thiện bổ sung phần học liệu, đào tạo trong tương lai để phù hợp với hơn người học.

## 2 Mục tiêu

Với tập dữ liệu được thu thập, thực hiện phân tích dữ liệu một cách hợp lý và diễn giải cũng như đưa ra kết luận từ những phân tích đó. Trong bài tập lớn này, mục tiêu trọng tâm là trang bị cho sinh viên những kỹ năng định lượng cần thiết mà họ có thể sử dụng và xây dựng dựa trên những cách thức linh hoạt. Các sinh viên chuyên ngành được yêu cầu học các khái niệm và công cụ để làm việc với dữ liệu.

Sinh viên phải hiểu [1] các nguyên tắc cơ bản của lý thuyết xác suất, [2] lý luận thống kê và phương pháp suy luận, [3] tính toán thống kê, [4] phân tích mô tả, diễn giải và khám phá dữ liệu bằng đồ họa và các phương tiện khác; [5] sinh viên hoàn tất bài tập cũng phải học cách trình bày bài hiệu quả.

## 3 Mô tả dữ liệu

Dữ liệu mẫu là 10 sheet dữ liệu từ "E1" đến "E10", mỗi sheet là một bảng tính gồm nhiều dòng, nhiều cột. Mỗi dòng mô tả một tiêu chí đánh giá sinh viên gồm có 26 tiêu chí. Mỗi cột thể hiện kết quả của một sinh viên tương ứng với các tiêu chí. Mỗi sinh viên sẽ được đánh giá bởi 5 mức điểm từ 0 đến 5 thể hiện ở bên dưới (**Bảng 1**).

| Mức chọn: 0. Không có thông tin để đánh giá; 1. Rất yếu; 2. Yếu; 3. Trung bình; 4. Tốt              |     |     |
|---|-----|-----|
| TIÊU CHÍ  | P.1 | P.2 |
| 01. Khả năng ứng dụng kiến thức tin học trong công việc   | 3   | 2   |
| 02. Khả năng suy nghĩ logic và/hoặc vận dụng kiến thức toán học trong công việc                     | 3   | 3   |
| 03. Khả năng sử dụng công cụ và công nghệ thích hợp tại đơn vị                                      | 3   | 3   |
| 04. Khả năng thiết kế ứng dụng dựa trên yêu cầu của khách hàng                                      | 0   | 0   |
| 05. Khả năng kiểm tra một ứng dụng thông qua việc xây dựng một tập kiểm tra mẫu                     | 0   | 0   |
| 06. Khả năng trao đổi ý tưởng và chia sẻ công việc  | 3   | 3   |
| 07. Nhận thức về trách nhiệm của sinh viên thực tập   | 2   | 4   |
| 08. Khả năng viết báo cáo   | 3   | 3   |
| 09. Khả năng chuẩn bị một bài thuyết trình  | 4   | 4   |
| 10. Kỹ năng nói khi thuyết trình  | 2   | 4   |
| 11. Khả năng ước lượng chi phí của một dự án phần mềm   | 0   | 0   |
| 12. Khả năng tự học   | 3   | 3   |
| 13. Các chứng chỉ chuyên môn mà sinh viên tự trang bị   | 3   | 3   |
| 14. Khả năng sử dụng Internet để phục vụ công việc  | 4   | 4   |
| 15. Khả năng sử dụng các ứng dụng xử lý dữ liệu đặc thù trong công việc                             | 0   | 0   |
| 16. Sinh viên làm chủ ít nhất một ngôn ngữ lập trình  | 0   | 0   |
| 17. Sinh viên thấy được sự trả giá giữa thời gian thực thi và không gian lưu trữ của một giải thuật | 0   | 0   |
| 18. Sinh viên có khả năng phân tích ưu và khuyết điểm của một giải pháp kỹ thuật                    | 2   | 3   |
| 19. Sinh viên có khả năng đọc hiểu một phần thiết kế của một dự án thực tế                          | 2   | 2   |
| 20. Khả năng xác định được vai trò và trách nhiệm của mình khi làm việc nhóm                        | 4   | 4   |
| 21. Giúp đỡ các bạn cùng nhóm trong công việc   | 0   | 4   |
| 22. Đóng góp kỹ thuật vào dự án chung của nhóm  | 2   | 2   |
| 23. Kỹ năng thu thập và tổng hợp thông tin  | 4   | 4   |
| 24. Kỹ năng tổ chức thực hiện   | 3   | 3   |
| 25. Tính kỷ luật trong công việc  | 4   | 3   |
| 26. Tham gia vào các hoạt động sinh hoạt chung của đơn vị   | 4   | 4   |

**Bảng 1:** Bảng mẫu đánh giá

## 4 Nhiệm vụ

Gọi **MD** là mã đề riêng cho mỗi nhóm (gồm 6 ký số) không trùng nhau, nhóm sinh viên sẽ thực hiện các yêu cầu dưới đây với các giá trị xác định như sau:

- Mỗi nhóm sẽ dùng R để thao tác trên 6 sheet dữ liệu được chọn theo cách:
  - **Sheet i:** { "ký số thứ j của MD" + 1},  $i \in \{1, 2, 3, 4, 5, 6\}$

### Câu hỏi

- Xác định số lượng sinh viên trong tập mẫu
- Tính điểm trung bình cho mỗi tiêu chí, làm tròn 2 số thập phân
- Tính điểm trung bình cho mỗi SV, làm tròn 2 số thập phân
- Liệt kê danh sách sinh viên có tổng điểm các tiêu chí là nhỏ nhất
- Liệt kê danh sách sinh viên có tổng điểm các tiêu chí là lớn nhất
- Liệt kê danh sách tiêu chí mà tổng điểm của các sinh viên là nhỏ nhất
- Liệt kê danh sách tiêu chí mà tổng điểm của các sinh viên là lớn nhất

8. Liệt kê tổng điểm các tiêu chí và số lượng sinh viên mà tổng điểm các tiêu chí là bằng nhau

Mỗi nhóm sẽ lấy 2 tiêu chí bằng cách sau đây:

- **Tiêu chí 1:**  $tc1 = MD \bmod 26 + 1$
  - **Tiêu chí 2:**  $tc2 = (MD + tc1) \bmod 26 + 1$
9. Tính minimum, maximum, median, tứ phân vị (quartile) thứ nhất (Q1) và thứ ba (Q3) cho 2 tiêu chí trên của mẫu.
10. Tính tổng điểm trung bình của tập các tiêu chí về kỹ thuật, làm tròn 2 số thập phân
11. Tính tổng điểm trung bình của tập các tiêu chí về kỹ năng mềm, làm tròn 2 số thập phân
12. Tính tổng điểm trung bình cho các tiêu chí về tính kỷ luật, làm tròn 2 số thập phân
13. Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về kỹ thuật là nhỏ nhất
14. Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về kỹ thuật là lớn nhất
15. Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về kỹ năng mềm là nhỏ nhất
16. Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về kỹ năng mềm là lớn nhất
17. Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về tính kỷ luật là nhỏ nhất
18. Liệt kê danh sách sinh viên mà có tổng điểm các tiêu chí về tính kỷ luật là lớn nhất
19. Kiểm tra các cặp tiêu chí sau xem liệu các công ty đã đánh giá tương thích chưa. Bằng cách liệt kê danh sách các sinh viên mà có sai lệch lớn về đánh giá các tiêu chí có liên quan với nhau cho từng sheets riêng lẻ:
- Khả năng ứng dụng kiến thức tin học trong công việc so với sử dụng Internet để phục vụ công việc
  - Nhận thức về trách nhiệm của sinh viên thực tập so với tính kỷ luật trong công việc
  - Khả năng trao đổi ý tưởng và chia sẻ công việc so với Giúp đỡ các bạn cùng nhóm trong công việc
  - Khả năng trao đổi ý tưởng và chia sẻ công việc so với đóng góp kỹ thuật vào dự án chung của nhóm
  - Khả năng sử dụng công cụ và công nghệ thích hợp tại đơn vị so với khả năng tự học
  - Khả năng thiết kế ứng dụng dựa trên yêu cầu của khách hàng so với làm chủ ít nhất một ngôn ngữ lập trình
  - Khả năng suy nghĩ logic và/hoặc vận dụng kiến thức toán học trong công việc so với sinh viên có thể đọc hiểu một phần thiết kế của một dự án thực tế

## Biểu đồ

20. Hãy vẽ biểu đồ hiển thị phân phối tổng điểm các tiêu chí của sinh viên
21. Vẽ phổ tổng điểm các tiêu chí theo sinh viên
22. Vẽ phổ tổng điểm các sinh viên theo tiêu chí
23. Hãy vẽ đồ thị dotplot cho tiêu chí '01'. Khả năng ứng dụng kiến thức tin học trong công việc" và cho giải thích về hình dạng (shape), tâm (center), độ biến thiên (variability), outlier. của đồ thị cho từng sheets riêng lẻ,
- (a) shape: symmetric, skewed

- (b) center: median
  - (c) variability: dùng giá trị nhỏ nhất và lớn nhất khi mô tả sự thay đổi
  - (d) outlier: quan sát nằm phía ngoài tổng thể mẫu
24. Thử so sánh phân phối của 2 biến tiêu chí bằng cách vẽ đồ thị dotplot và thảo luận shape, outliers, center, và variability. Từ 2 tiêu chí được tính ra cho nhóm cho từng sheets riêng lẻ.
  25. Vẽ box-and-whisker cho 2 tiêu chí ở câu trên của nhóm
  26. Vẽ scatterplot cho tổng điểm trung bình gồm các sheets của tập các tiêu chí về kỹ thuật, làm tròn 2 số thập phân
  27. Vẽ scatterplot cho tổng điểm trung bình gồm các sheets của tập các tiêu chí về kỹ năng mềm, làm tròn 2 số thập phân
  28. Vẽ scatterplot cho tổng điểm trung bình gồm các sheets của tập các tiêu chí về tính kỷ luật, làm tròn 2 số thập phân.
  29. Vẽ biểu đồ tần số tương đối tích lũy của điểm của sinh viên trên tập các tiêu chí về kỹ thuật
  30. Vẽ biểu đồ tần số tương đối tích lũy của điểm của sinh viên trên tập các tiêu chí về kỹ năng mềm
  31. Vẽ biểu đồ tần số tương đối tích lũy của điểm của sinh viên trên tập các tiêu chí về tính kỷ luật
  32. Vẽ biểu đồ cho thấy số lượng sinh viên bị sai lệnh lớn trong đánh giá các cặp tiêu chí tương thích cho từng trường hợp ở câu hỏi trên cho 6 sheets

## 5 Hướng dẫn và yêu cầu

### 5.1 Hướng dẫn

- Cài đặt đồng thời cả R và Rstudio.
- Đọc kỹ và xử lý lại tất cả những thí dụ đã có trong file mẫu.
- Tìm hiểu kỹ cách soạn thảo văn bản bằng LaTeX và cách sử dụng phần mềm R trong các file hướng dẫn và tìm hiểu thêm trong các tài liệu khác.
- Tạo một folder chung chứa mọi thứ cần thiết để share giữa các thành viên trong nhóm trên các cloud services như [Google Drive](#) hay [Dropbox](#),...
- Dùng Doodle để lên kế hoạch họp nhóm.
- Dùng Trello để quản lý project.

### 5.2 Yêu cầu

Mỗi nhóm, từ 3 đến 6 sinh viên, đề xuất giải pháp. Nhóm cần nộp báo cáo trình bày về lời giải cho các câu hỏi và kết quả thực nghiệm. Đồng thời, nhóm cũng cần nộp source code, và trình bày các kết quả của nhóm trong khoảng 5 minutes.

Báo cáo và slide trình bày cần được viết dưới dạng LaTeX.

- Thời gian làm bài: **Từ ngày 09/11/2020 – 18g00 ngày 04/12/2020.**

Đối với mỗi bài toán, yêu cầu sinh viên trình bày lời giải theo lối truyền thống, sử dụng các công thức, kết quả lý thuyết trong phần kiến thức chuẩn bị. Đồng thời, sau đó trình bày kết quả tính toán và biểu đồ minh họa bằng R.

- Trình bày cả code R và kết quả tính toán trong R giống như file mẫu.
- Viết báo cáo theo đúng **bố cục như trong file mẫu** bằng LaTeX.
- Mỗi nhóm khi nộp bài **cần phải nộp theo file log (nhật ký)** ghi rõ: tiến độ công việc, phân công nhiệm vụ, trao đổi của các thành viên,...

### 5.3 Nộp bài

- SV chỉ nộp bài qua hệ thống Sakai: nén tất cả các file cần thiết (file .tex, file .R, ...) thành một file tên là “*BTL-CO1007-MT161-Nhom-n.zip*” và nộp trong mục Assignment.
- Lưu ý: mỗi nhóm **chỉ cần một thành viên là nhóm trưởng nộp bài**.

## 6 Cách đánh giá và xử lý gian lận

### 6.1 Đánh giá

Mỗi bài làm sẽ được đánh giá như sau.

| Nội dung  | Tỉ lệ điểm (%) |
|---|----------------|
| Giải đúng các bài toán bằng công thức và lập luận | 30%            |
| Các lệnh (hàm) R được sử dụng đúng đắn và hợp lý  | 30%            |
| Trình bày kiến thức chuẩn bị rõ ràng, phù hợp     | 20%            |
| Trình bày văn bản đẹp, đúng chuẩn                 | 20%            |

### 6.2 Xử lý gian lận

Bài tập lớn phải được sinh viên (nhóm) TỰ LÀM. Sinh viên (nhóm) sẽ bị coi là gian lận nếu:

- Có sự giống nhau bất thường giữa các bài thu hoạch (nhất là phần kiến thức chuẩn bị). Trong trường hợp này, **TẤT CẢ** các bài nộp có sự giống nhau đều bị coi là gian lận. Do vậy sinh viên (nhóm) phải bảo vệ bài làm của mình.
- Sinh viên (nhóm) không hiểu bài làm do chính mình viết. Sinh viên (nhóm) có thể tham khảo từ bất kỳ nguồn tài liệu nào, tuy nhiên phải đảm bảo rằng mình hiểu rõ ý nghĩa của tất cả những gì mình viết.

Bài bị phát hiện gian lận thì sinh viên sẽ bị xử lý theo quy định của nhà trường.

## Tài liệu

- [Dal] Dalgaard, P. *Introductory Statistics with R*. Springer 2008.
- [K-Z] Kenett, R. S. and Zacks, S. *Modern Industrial Statistics: with applications in R, MINITAB and JMP*, 2nd ed., John Wiley and Sons, 2014.
- [Ker] Kerns, G. J. *Introduction to Probability and Statistics Using R*, 2nd ed., CRC 2015.