

Bài tập thực hành thống kê

Bài tập số 3

NHỮNG ĐIỀU LƯU Ý TRƯỚC KHI LÀM BÀI VÀ NỘP BÀI:

- Làm bài trên 1 file script duy nhất (trình bày theo file mẫu - link bên dưới), đặt tên như sau: **btth.mssv.R**. Ví dụ, sinh viên có mã số là 1712171 sẽ đặt tên file như sau: **btth.1712171.R** (lưu ý là viết thường các chữ btth).
- Hạn nộp bài số 3: nộp bài **trước 23g59 ngày 26/12/2021**. Nộp bài sau thời điểm này: 0 đ. Các bài sao chép nhau: 0 đ. Những trường hợp nộp đúng hạn nhưng nộp lại vì một lý do nào đó (ví dụ: phải chỉnh sửa) sẽ bị trừ 1 điểm trên một lần nộp lại.
- Link nộp bài sẽ đăng kèm với bài tập trên google classroom.
- Với một số lệnh mới trong bài tập, để hiểu rõ cách sử dụng, sinh viên có thể sử dụng help hoặc tìm hướng dẫn trên google.
- File dữ liệu các bạn lấy ở “Data cho môn học” tuần 1.
- File mẫu tải về tại đây:
<https://drive.google.com/file/d/1pAasHiCqZPxi3IBiCieZbmM-SgRtJBN6/view?usp=sharing>

Đề

Bài 1. Doanh số bán hàng (triệu đồng) của hai chi nhánh trong hệ thống siêu thị Co-op Mart đặt tại quận 1 và quận 3 trong 200 ngày được cho trong file **Profit.csv**.

- Hãy xác định khoảng tin cậy 95% cho sự sai khác về doanh số bán hàng trung bình của hai chi nhánh này?
- Một ngày được gọi là có doanh số cao nếu doanh số bán hàng trong ngày đó trên 600 triệu đồng. Có nhận xét rằng tỉ lệ các ngày có doanh số cao tại chi nhánh quận 1 cao hơn tỉ lệ các ngày có doanh số bán hàng cao tại chi nhánh quận 3. Hãy kiểm chứng nhận xét này với $\alpha = 0.05$. Hãy cho biết p-giá trị của kiểm định vừa rồi.
- Viết hàm **prop.test.leq**(x, y, α) để kiểm định giả thuyết $H_0 : p_x = p_y$ và đối thuyết $H_1 : p_x < p_y$, trong đó p_x, p_y lần lượt là tỉ lệ ngày có doanh số cao (trên 600 triệu) tại chi nhánh quận 1 và quận 3; x, y lần lượt là hai vec-tơ chứa doanh số bán hàng tại hai chi nhánh của Co-op Mart đặt tại quận 1 và quận 3 trong một số ngày nào đó; α là mức ý nghĩa. Hàm xuất ra kết quả chấp nhận hoặc bác bỏ và cho biết p-giá trị. Áp dụng cho câu b) với đối thuyết $H_1 : p_x < p_y$. (Lưu ý: không gọi hàm có sẵn trong **R**).

Bài 2. Khi độ cao tăng lên thì nhiệt độ không khí giảm xuống. Theo kinh nghiệm của một số người, nhiệt độ sẽ giảm $9.8^\circ C/km$. Để kiểm tra xem nhận định này có đúng hay không, một nhóm sinh viên mang theo nhiệt kế và thiết bị đo độ cao và ghi lại được dữ liệu sau trong một chuyến leo núi của họ.

Độ cao (ft)	600	1000	1250	1600	1800	2100	2500	2900
Nhiệt độ không khí (F)	56	54	56	50	47	49	47	45

- Vẽ đồ thị phân tán và đường hồi quy trên cùng hệ trục tọa độ.
- Ước lượng đường thẳng hồi quy cho dữ liệu trên. Kinh nghiệm trên đề bài có đúng không?

Lưu ý: Khi làm bài các em cần nhận xét, kết luận, giải thích (nếu cần) thông qua dạng ghi chú (# ghi chú) như trong code hướng dẫn những bài kiểm định thầy đã làm trong slide bài giảng.