

PROJECT MÔN PHÂN TÍCH THỐNG KÊ

Giảng viên: Tiến sĩ Bùi Thanh Hùng
Trưởng Lab Khoa học Phân tích dữ liệu và Trí tuệ nhân tạo
Giám đốc chương trình Hệ thống thông tin
Đại học Thủ Dầu Một
hungbt.cntt@tdmu.edu.vn
Website: <https://sites.google.com/site/hungthanhbui1980>
Nộp bài: E-learning (code+dữ liệu+report), deadline: **24h Thứ Năm ngày 23/07/2020**
Chấm bài: **8h 30 Thứ Sáu, 24/07/2020**, tại I1.105 (In đen trắng 1 cuốn báo cáo theo mẫu)

PHẦN 1 (3 điểm):

Hãy giải các bài toán sau:

Bài 1 (1 điểm):

Một cửa hàng thực phẩm nhận thấy thời gian vừa qua trung bình một khách hàng mua 99 nghìn đồng thực phẩm trong ngày. Nay cửa hàng chọn ngẫu nhiên 15 khách hàng thấy trung bình một khách hàng mua xy nghìn đồng trong ngày và phương sai mẫu điều chỉnh là $s^2 = (2 \text{ nghìn đồng})^2$. Với mức ý nghĩa là $y\%$, thử xem có phải sức mua của khách hàng hiện nay thực sự giảm sút.

Với xy : là 2 số cuối mã sinh viên của em

y : là số cuối mã sinh viên của em

Bài 2 (1 điểm):

Nhà máy sản xuất 100.000 sản phẩm trong đó có 30.000 sản phẩm loại 2, còn lại là sản phẩm loại 1. KCS đến kiểm tra và lấy ra 150 sản phẩm để thử. Trong 2 trường hợp chọn lặt và chọn không lặt. Hãy tính xác suất để số sản phẩm loại 2 mà KCS phát hiện ra:

a) Từ xyz đến $xyz + z \cdot 10$

b) Ít hơn $xyz + 10$

Với xyz : là 3 số cuối mã sinh viên của em

Bài 3 (1 điểm):

Một gia đình có $x+y$ người con. Giả sử xác suất sinh con trai, con gái như nhau. Tính xác suất:

- a) Không có con trai.
- b) Có x con trai và y con gái.
- c) Số trai từ x đến y (nếu $y > x$) (hoặc từ y đến x (nếu $x > y$)).

Với x, y là 2 số cuối mã sinh viên của em

PHẦN 2 (7 điểm)**Project**

Thực hiện các yêu cầu sau và viết báo cáo theo mẫu gửi kèm:

1. Thu thập dữ liệu (1 điểm)
2. Phân tích thống kê dữ liệu: Lớn nhất, nhỏ nhất, trung bình, phần tử xuất hiện nhiều nhất trong mỗi cột, viết code để sinh ra kết quả tự động và lưu kết quả vào file excel. (2 điểm)
3. Phân tích độ tương quan giữa các đại lượng bằng 2 độ đo tự viết code: Pearson (ra đồ thị Heatmap) và Spearman (2 điểm)
4. Dự đoán bằng phương pháp Bayes tự viết và đánh giá bằng 3 độ đo (Accuracy, Confusion Matrix score và F1 score) (2 điểm)