

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

ĐỀ THI KẾT THÚC HỌC KỲ I
NĂM HỌC 2019 - 2020

Môn thi: THỐNG KÊ ỨNG DỤNG

Mã môn học: **MAT2406**

Số tín chỉ: **4**

Đề số: **1**

Dành cho sinh viên hệ: Chính quy *Ngành:* Máy tính và khoa học thông tin

Thời gian làm bài: **120 phút** (không kể thời gian phát đề)

Câu 1. Vì sao trung bình mẫu ngẫu nhiên độc lập là ước lượng không chệch của kỳ vọng?

Câu 2. Cho mẫu quan sát gồm 50 cặp giá trị về tốc độ (đơn vị: km/h) và khoảng cách dừng (đơn vị: mét) của một loại ô tô. Khoảng cách dừng là quãng đường xe ô tô di chuyển được tính từ thời điểm bắt đầu phanh đến khi xe dừng hẳn.

Gọi "speed" và "dist" lần lượt là dữ liệu về tốc độ và khoảng cách dừng.

Sử dụng phần mềm thống kê R, ta thu được một số kết quả.

a) Giải thích các kết quả chạy ra bên dưới. Phát biểu bài toán kiểm định giả thuyết và kết luận.

One Sample t-test

```
data: dist
t = 1.9153, df = 49, p-value = 0.0613
alternative hypothesis: true mean is not equal to 36
90 percent confidence interval:
 36.87008 49.08992
sample estimates:
mean of x
 42.98
```

b) Giải thích các kết quả chạy ra bên dưới. Kết luận gì về mối quan hệ giữa tốc độ và khoảng cách dừng của xe?

Pearson's product-moment correlation

```
data: speed and dist
t = 9.464, df = 48, p-value = 1.49e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6816422 0.8862036
sample estimates:
cor
0.8068949
```

c) Giải thích các kết quả chạy ra bên dưới. Kết luận gì về mô hình biểu diễn quan hệ giữa khoảng cách dừng và tốc độ của xe?

```
call:
lm(formula = dist ~ speed)

Residuals:
    Min       1Q   Median       3Q      Max
-29.069  -9.525  -2.272   9.215  43.201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.5791     6.7584  -2.601   0.0123 *
speed         3.9324     0.4155   9.464 1.49e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.38 on 48 degrees of freedom
Multiple R-squared:  0.6511,    Adjusted R-squared:  0.6438
F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

Câu 3. Cho mẫu quan sát về lượng đường trong máu (đơn vị: mg/dL) của 40 người được chia thành 3 nhóm: nhóm 1 gồm 9 bệnh nhân mắc bệnh crohn; nhóm 2 gồm 11 bệnh nhân mắc bệnh colitis và nhóm 3 gồm 20 đối tượng không mắc bệnh.

Gọi "group1", "group2", "group3" và "all" lần lượt là dữ liệu về lượng đường trong máu của nhóm 1, nhóm 2, nhóm 3 và cả ba nhóm.

Nhân tố "group" trong dữ liệu "all" có 3 mức giá trị {1, 2, 3} tương ứng với 3 nhóm.

Sử dụng phần mềm thống kê R, ta thu được một số kết quả.

a) Giải thích các kết quả chạy ra bên dưới. Phát biểu bài toán kiểm định giả thuyết và kết luận.

```
welch Two sample t-test

data: group2 and group3
t = -2.4227, df = 16.581, p-value = 0.02718
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 -993.5575 -162.3516
sample estimates:
mean of x mean of y
 2226.545  2804.500
```

b) Giải thích các kết quả chạy ra bên dưới. Ba nhóm bệnh nhân có lượng đường trong máu trung bình khác nhau hay không?

Analysis of Variance Table

Response: all

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	5683620	2841810	8.6655	0.0008191 ***
Residuals	37	12133923	327944		

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Câu 4. Tính tỉ số tương quan của Y đối với X và hệ số xác định của tập số liệu sau đây:

X	25	20	10	15	20	30	25	10	30	15
Y	18	25	6	13	21	10	22	10	14	17

Có kết luận gì về mối tương quan giữa X và Y (phi tuyến hay tuyến tính)?

—————Hết—————

Ghi chú: Sinh viên được sử dụng tài liệu, cán bộ coi thi không giải thích gì thêm.