

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

ĐỀ THI CUỐI KỲ I
NĂM HỌC 2020 - 2021

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. Nêu định nghĩa và ý nghĩa của xác suất sai lầm loại 1 và xác suất sai lầm loại 2 trong bài toán kiểm định giả thuyết thống kê. Cho ví dụ.

Câu 2. Tỷ số tương quan và hệ số xác định giữa hai đại lượng ngẫu nhiên là gì? Nêu mối liên hệ giữa bài toán phân tích phương sai một nhân tố và bài toán kiểm định giả thuyết tương quan phi tuyến giữa hai đại lượng ngẫu nhiên.

Câu 3. Cho mẫu quan sát về chiều cao (đơn vị: cm) của 25 cặp anh em trai. Gọi x và y lần lượt là dữ liệu về chiều cao của người anh và người em. Sử dụng phần mềm thống kê R, ta thu được kết quả dưới đây. Giải thích kết quả. Phát biểu bài toán kiểm định giả thuyết và kết luận.

wilcoxon signed rank test with continuity correction

data: x and y

v = 122, p-value = 0.3172

alternative hypothesis: true location shift is less than 0

Câu 4. An là một vận động viên bóng chày. Trong mỗi trận đấu, anh ta có $n = 4$ lần đánh bóng. Gọi y là số lần đánh bóng trúng đích trong mỗi trận đấu, xác suất đánh trúng trong mỗi lần là $p = 0.312$. Giả sử, An tham gia 70 trận đấu, trong đó có 17 trận không đánh bóng trúng đích, 31 trận đánh bóng trúng đích 1 lần, 17 trận đánh bóng trúng đích 2 lần và 5 trận đánh bóng trúng đích từ 3 lần trở lên.

a) Sử dụng phần mềm thống kê R để chạy đoạn mã như dưới đây. Đoạn mã đã chính xác chưa? Nếu sai, hãy sửa lại.

```
> x = c(17, 17, 31, 5)
> p = pbinom(c(0, 1, 2, 3, 4), 4, 0.312)
> pt = c(p[1:3], 1 - sum(p[1:3]))
> chisq.test(x, p = pt)
```

b) Với mức ý nghĩa 5%, kiểm tra xem số lần đánh bóng trúng đích có tuân theo phân bố nhị thức $B(4, 0.312)$ hay không nếu kết quả kiểm định như sau:

Chi-squared test for given probabilities

```
data: x
X-squared = 12.081, df = 3, p-value = 0.007112
```

Câu 5. Dữ liệu *iris* trong *UsingR* bao gồm 50 quan sát về bốn phép đo: "Sepal.Length" (chiều dài đài hoa), "Sepal.Width" (chiều rộng đài hoa), "Petal.Length" (chiều dài cánh hoa), "Petal.Width" (chiều rộng cánh hoa) tính bằng centimet của ba loại hoa "Species" thuộc họ Diên vĩ: *setosa*, *versicolor*, và *virginica*.

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

a) Giải thích các kết quả chạy ra bên dưới. Có sự khác biệt về chiều dài đài hoa của ba loại hoa không?

One-way analysis of means

```
data: Sepal.Length and Species
F = 119.26, num df = 2, denom df = 147, p-value < 2.2e-16
```

b) Giải thích các kết quả chạy ra bên dưới. Kết luận gì về mối quan hệ giữa chiều dài và chiều rộng của cánh hoa?

Pearson's product-moment correlation

```
data: Petal.Length and Petal.Width
t = 43.387, df = 148, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9490525 0.9729853
sample estimates:
      cor
0.9628654
```

c) Đưa ra mô hình biểu diễn mối quan hệ giữa chiều dài và chiều rộng cánh hoa. Với chiều dài cánh hoa là 3.9 cm, đưa ra dự đoán về chiều rộng cánh hoa. Có cơ sở để nói rằng các hệ số hồi quy trong mô hình khác 0 không? Vì sao?

```
Call:
lm(formula = Petal.Length ~ Petal.Width)

Residuals:
    Min       1Q   Median       3Q      Max
-1.33542 -0.30347 -0.02955  0.25776  1.39453

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.08356    0.07297   14.85  <2e-16 ***
Petal.Width  2.22994    0.05140   43.39  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4782 on 148 degrees of freedom
Multiple R-squared:  0.9271,    Adjusted R-squared:  0.9266
F-statistic: 1882 on 1 and 148 DF,  p-value: < 2.2e-16
```

—————Hết—————

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