

TRƯỜNG ĐẠI HỌC VĂN LANG
Khoa Công nghệ ứng dụng

ĐỀ THI, ĐÁP ÁN/RUBRIC VÀ THANG ĐIỂM
THI KẾT THÚC HỌC PHẦN
Học kỳ 2, năm học 2023-2024

I. Thông tin chung

Tên học phần:	Thống kê sinh học		
Mã học phần:	71GENE40222	Số tin chỉ:	2
Mã nhóm lớp học phần:	232_71GENE40222_01		
Hình thức thi: Tự luận	Thời gian làm bài:	90	phút
<i>Thí sinh được tham khảo tài liệu:</i>	<input checked="" type="checkbox"/> Có	<input type="checkbox"/> Không	

Yêu cầu: Sinh viên sử dụng máy tính cá nhân trong quá trình làm bài. Sau khi đăng nhập CTE để tải đề thi, sinh viên thực hiện bài làm trên máy tính và nộp bài làm (file word hoặc pdf) vào trang thi.

II. Các yêu cầu của đề thi nhằm đáp ứng CLO

(Phần này phải phối hợp với thông tin từ đề cương chi tiết của học phần)

Ký hiệu CLO	Nội dung CLO	Hình thức đánh giá	Trọng số CLO trong thành phần đánh giá (%)	Câu hỏi thi số	Điểm số tối đa	Lấy dữ liệu đo lường mức đạt PLO/PI
(1)	(2)	(3)	(4)	(5)	(6)	(7)
CLO1	Vận dụng các kiến thức về toán học và sinh học trong thống kê để giải quyết các vấn đề xét nghiệm phân tử và liệu pháp sinh học phù hợp trong y dược	Tự luận	20%	1	2.5	2.1
CLO2	Áp dụng thành thạo các nguyên tắc thống kê nhằm phân tích cơ sở khoa học của các phương pháp xét nghiệm phân tử và liệu pháp sinh học	Tự luận	20%	2	2.5	3.1

CLO3	Sử dụng thành thạo các công cụ thống kê và xử lý số liệu nhằm phục vụ trong công việc nghiên cứu và phân tích sản phẩm CNSH Y Dược	Tự luận	20%	3	2.5	4.2
CLO4	Lựa chọn thành thạo các phần mềm thống kê với từng loại dữ liệu sinh học và diễn giải kết quả thống kê	Tự luận	20%	4	2.5	5.2
CLO5	Ý thức về nhu cầu học tập suốt đời để phát triển sự nghiệp	Tự luận	20%	4	2.5	7.3

III. Nội dung câu hỏi thi

Câu hỏi 1: (2.5 điểm)

Nhập dữ liệu có tên igf.txt vào R

- Dữ liệu trên có bao nhiêu cột (biến) và dòng (bệnh nhân), phân loại các biến thuộc loại biến nào (biến nhị phân, biến liên tục, biến phân loại, biến xếp hạng)
- Cho biết giá trị trung bình, trung vị, độ lệch chuẩn, sai số chuẩn về độ tuổi của các bệnh nhân tham gia vào nghiên cứu
- Cho biết độ tuổi trung bình của nam và nữ tham gia vào nghiên cứu
- Cho biết nồng độ trung bình của hormone igf1 ở nam và nữ là bao nhiêu?

Hướng dẫn:

#Đọc dữ liệu igf.txt vào R

```
igfdata <- read.csv("C:/Users/phuc.bui/Downloads/igf.txt", sep=";", header = TRUE)
```

Lưu ý: Thay đổi đường link để truy cập vào dữ liệu **igf.txt**

Truy cập các đối tượng trong khung dữ liệu

```
attach(igfdata)
```

xem xét các cột số trong dữ liệu

```
names(igfdata)
```

#Kiểm tra dữ liệu

```
head(igfdata)
```

Số bệnh nhân và số biến

```
dim(igfdata)
```

#Tính trung bình, trung vị, phương sai, độ lệch chuẩn của độ tuổi

```
round(mean(age),2)
```

```
round(median(age),2)
```

```
round(var(age),2)
```

```
round(sd(age),2)
```

#Cho biết độ tuổi trung bình của nam và nữ tham gia vào nghiên cứu

```
data.frame(round(tapply(igfdata$age, list(igfdata$sex), mean), 2))
```

Câu hỏi 2: (2.5 điểm)

Sử dụng dữ liệu igf.txt

- Vẽ biểu đồ về phân bố nồng độ hormone igfi ở nam và nữ. Copy hình vào bài làm
- Sử dụng kiểm định t để so sánh sự khác biệt về nồng độ hormone igfi ở nam và nữ có ý nghĩa thống kê hay không? Dán kết quả so sánh vào bài làm
- Cần thu thập bao nhiêu bệnh nhân nam và nữ cho nghiên cứu trong trường hợp kết quả so sánh hormone igfi không có ý nghĩa thống kê?

Hướng dẫn:

```
# Cài phần mềm vẽ biểu đồ ggplot2
install.packages("ggplot2")
library(ggplot2)
#Vẽ biểu đồ cột về phân bố nồng độ hormone igfi ở nam và nữ
ggplot(data=igfdata,aes(sex,igfi,colour=sex)) + geom_violin() + geom_point() +
xlab("Sex") + ylab("igfi concentration")
#So sánh sự khác biệt về nồng độ hormone igfi ở nam và nữ có ý nghĩa thống kê hay
không?
t.test(igfi~ sex, alternative = "two.sided")
#Tính độ lệch chuẩn
sd(igfi)
#Cài đặt package "tidyverse"
install.packages("tidyverse")
library(tidyverse)
# Tính giá trị trung bình igfi của 2 nhóm
igfdata %>% group_by(sex) %>% summarise(mean=mean(igfi))
# Cần thu thập bao nhiêu bệnh nhân?
power.t.test(delta = 167 - 160, sd = 50, sig.level = 0.05, power = 0.8, type = "two.sample",
alternative = "two.sided")
```

Câu hỏi 3: (2.5 điểm)

Một nhóm bệnh nhân gồm 10 người được điều trị bằng một thuốc nhằm giảm huyết áp. Huyết áp của bệnh nhân được đo lúc khởi đầu nghiên cứu (lúc chưa điều trị), và sau khi điều trị. Số liệu huyết áp của 10 bệnh nhân như sau:

Trước khi điều trị (x_0)	180, 140, 160, 160, 220, 185, 145, 160, 160, 170
Sau khi điều trị (x_1)	170, 145, 145, 125, 205, 185, 150, 150, 145, 155

- Hãy cho biết độ biến chuyển huyết áp trên có đủ để kết luận rằng thuốc điều trị có hiệu quả giảm áp huyết hay không?
- Trình bày cách làm và giải thích kết quả kiểm định. Cho biết sự khác biệt trung bình về chỉ số huyết áp giữa hai nhóm trước và sau điều trị là bao nhiêu?

Hướng dẫn:

```
# Nhập dữ liệu
before <- c(180, 140, 160, 160, 220, 185, 145, 160, 160, 170)
after <- c(170, 145, 145, 125, 205, 185, 150, 150, 145, 155)
bp <- data.frame(before, after)
bp
# Kiểm định Student t-test cho biến theo cặp
t.test(before, after, paired=TRUE)
```

Câu hỏi 4: (2.5 điểm)

Một nghiên cứu đo lường độ cholestrol trong máu của 18 đối tượng nam. Tỉ trọng cơ thể (body mass index) cũng được ước tính cho mỗi đối tượng bằng công thức tính BMI là lấy trọng lượng (tính bằng kg) chia cho chiều cao bình phương (m²). Kết quả đo lường như sau: Độ tuổi, tỉ trọng cơ thể và cholesterol

Mã số ID (id)	Độ tuổi (age)	BMI (bmi)	Cholesterol (chol)
1	46	25.4	3.5
2	20	20.6	1.9
3	52	26.2	4.0
4	30	22.6	2.6
5	57	25.4	4.5
6	25	23.1	3.0
7	28	22.7	2.9
8	36	24.9	3.8
9	22	19.8	2.1
10	43	25.3	3.8
11	57	23.2	4.1
12	33	21.8	3.0
13	22	20.9	2.5
14	63	26.7	4.6
15	40	26.4	3.2
16	48	21.2	4.2
17	28	21.2	2.3
18	49	22.8	4.0

- Vẽ biểu đồ tán xạ (scatter plot) giữa độ tuổi và cholesterol. Thêm đường hồi quy giữa 2 biến
- Xác định hệ số tương quan (r) giữa hai biến số độ tuổi (x) và cholesterol (y)? Cho biết mức độ tương quan giữa 2 biến độ tuổi và cholesterol như thế nào?
- Viết công thức mô hình hồi quy tuyến tính giữa biến độ tuổi và cholesterol: $y = ax + b$. Tìm giá trị a và b

Hướng dẫn:

```
age <- c(46,20,52,30,57,25,28,36,22,43,57,33,22,63,40,48,28,49)
```

```
bmi <-c(25.4,20.6,26.2,22.6,25.4,23.1,22.7,24.9,19.8,25.3,23.2,21.8,20.9,26.7,26.4,21.2,21.2,22.8)
```

```
chol <- c(3.5,1.9,4.0,2.6,4.5,3.0,2.9,3.8,2.1,3.8,4.1,3.0, 2.5,4.6,3.2, 4.2,2.3,4.0)
```

```
data <- data.frame(age, bmi, chol)
```

```
# Vẽ biểu đồ tán xạ (scatter plot) giữa độ tuổi và cholesterol
```

```
plot(chol ~ age, pch=16)
```

```
# Vẽ đường hồi quy giữa 2 biến
```

```
abline(lm(chol~age),col="red")
```

```
# Xác định hệ số tương quan Pearson
```

```
cor(age, chol)
```

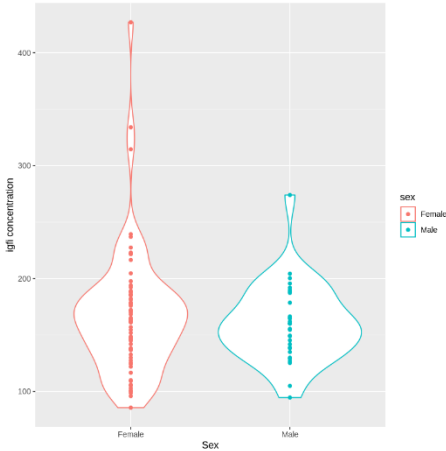
```
# Kiểm định giả thiết hệ số tương quan bằng 0 (tức hai biến x và y không có liên hệ)
```

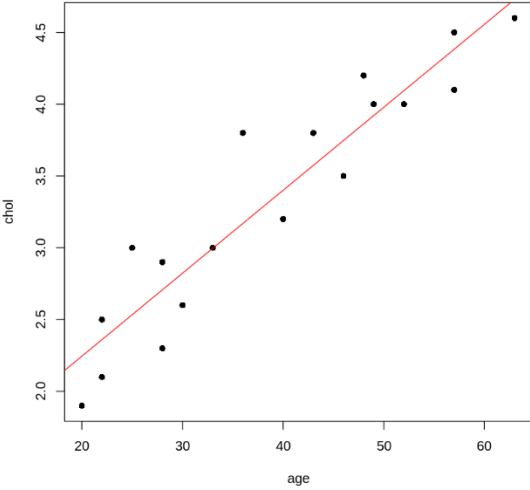
```
cor.test(age, chol)
```

```
# Mô hình của hồi qui tuyến tính
```

```
model <- lm(chol ~ age)
summary(model)
```

ĐÁP ÁP VÀ THANG ĐIỂM

Phần câu hỏi	Nội dung đáp án	Thang điểm	Ghi chú
I. Tự luận			
Câu 1	<ul style="list-style-type: none"> - Có 100 bệnh nhân và 12 biến - Biến nhị phân (sex), biến phân loại (ethnicity), biến rời rạc (age), biến liên tục (height, weight, igfi, igfbp3, als, pinp, ictp, p3np) - Trung bình, trung vị độ tuổi: 38.83, 38 - Trung bình độ tuổi của nam (19.59) và nữ (18.23) - Nồng độ trung bình của hormone igfi ở nam (167.97) và nữ (160.29) 	2.5	
Câu 2	<ul style="list-style-type: none"> - Biểu đồ về phân bố nồng độ hormone igfi ở nam và nữ.  <ul style="list-style-type: none"> - Kết quả kiểm định t-test cho thấy không có sự khác biệt nào về hormone igfi giữa nam và nữ <pre>Welch Two Sample t-test data: igfi by sex t = 0.84117, df = 88.329, p-value = 0.4025 alternative hypothesis: true difference in means between group Female and group Male is not equal to 0 95 percent confidence interval: -10.46855 25.83627 sample estimates: mean in group Female mean in group Male 167.9741 160.2903</pre> <ul style="list-style-type: none"> - Xác định số mẫu cần cho nghiên cứu cho dựa trên trung bình chỉ số igfi giữa nam và nữ <pre>Two-sample t test power calculation n = 801.8676 delta = 7 sd = 50 sig.level = 0.05 power = 0.8 alternative = two.sided NOTE: n is number in *each* group</pre>	2.5	

	<p>Kết luận: cần thu thập 801 bệnh nhân ở mỗi nhóm nam và nữ để xác định mức độ khác biệt về chỉ số igfi với mức ảnh hưởng 0.8 và mức ý nghĩa <0.05</p>		
Câu 3	<p>- Kết quả kiểm định Student t-test cho biến theo cặp</p> <pre>Paired t-test data: before and after t = 2.7924, df = 9, p-value = 0.02097 alternative hypothesis: true mean difference is not equal to 0 95 percent confidence interval: 1.993901 19.006099 sample estimates: mean difference 10.5</pre> <p>- Kết quả trên cho thấy sau khi điều trị áp suất máu giảm 10.5 mmHg, và khoảng tin cậy 95% là từ 2.0 mmHg đến 19 mmHg, với trị số p = 0.0209. Như vậy, chúng ta có bằng chứng để phát biểu rằng mức độ giảm huyết áp có ý nghĩa thống kê.</p>	2.5	
Câu 4	<p># Vẽ biểu đồ tán xạ (scatter plot) giữa độ tuổi và cholesterol</p> <pre>plot(chol ~ age, pch=16)</pre> <p># Vẽ đường hồi quy giữa 2 biến</p> <pre>abline(lm(chol~age),col="red")</pre>  <p># Xác định hệ số tương quan Pearson</p> <pre>cor(age, chol)</pre> <p>0.93</p> <p>- Có sự tương quan mạnh giữa độ tuổi và cholesterol</p> <p>- Kết quả kiểm định giả thiết hệ số tương quan bằng 0 (tức hai biến x và y không có liên hệ) cho thấy sự tương quan giữa độ tuổi và cholesterol có ý nghĩa về mặt thống kê với P= 1.058e-08 và ngưỡng tương quan từ 0.83-0.97.</p>	2.5	

	<pre># Xây dựng mô hình hồi quy tuyến tính giữ độ tuổi và cholesterol Call: lm(formula = chol ~ age) Residuals: Min 1Q Median 3Q Max -0.40729 - 0.24133 -0.04522 0.17939 0.63040 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 1.089218 0.221466 4.918 0.000154 *** age 0.057788 0.005399 10.704 1.06e-08 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.3027 on 16 degrees of freedom Multiple R- squared: 0.8775, Adjusted R-squared: 0.8698 F- statistic: 114.6 on 1 and 16 DF, p-value: 1.058e-08 - Mô hình hồi quy tuyến tính giữa độ tuổi và cholesterol là: $y = 0.05x + 1.08$ ($a=0.05$, $b=1.08$)</pre>		
	Điểm tổng	10.0	

TP. Hồ Chí Minh, ngày tháng năm 2024

Người duyệt đề

Giảng viên ra đề



TS. Vũ Thị Quyền

TS. Bùi Hoàng Phúc