



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-DBCL ghi)
CK18191-
MTH10404

Tên học phần: Lý Thuyết Thống Kê Mã HP: MTH10404
Thời gian làm bài: 90 phút Ngày thi: 09/01/2019
Ghi chú: Sinh viên [☒ được phép / ☐ không được phép] sử dụng tài liệu khi làm bài.

Họ tên sinh viên: MSSV: STT:

Câu 1. (4 điểm)

Một công ty sản xuất xe ô-tô chế tạo ra một dòng xe ô-tô Z, công ty kiểm tra xe bằng cách cho 30 xe chạy trên cùng 1 đoạn đường 100 km và đo lượng xăng sử dụng X (Đv: lít), kết quả cho ở bảng sau

X (lít)	4	4,1	4,2	4,3	4,4	4,5
Số xe	5	7	6	6	3	3

Giả sử lượng xăng tiêu hao là đại lượng ngẫu nhiên có phân phối chuẩn.

- (a) Ước lượng lượng xăng tiêu hao trung bình của dòng xe với độ tin cậy 95%.
- (b) Những xe có lượng xăng tiêu hao từ 4,3 lít trở lên được xếp vào loại không đạt tiêu chuẩn về nhiên liệu, hãy ước lượng tỷ lệ xe không đạt do nhà máy sản xuất với độ tin cậy 95%.
- (c) Lời khẳng định: "Lượng xăng tiêu hao trung bình của dòng xe này là 4,2 lít (trên 100 km)" có được chấp nhận hay không? (Mức ý nghĩa $\alpha = 5\%$)
- (d) Khảo sát lượng xăng tiêu hao Y (Đv: lít/100 km) của một dòng xe W cùng hãng, tính được $\bar{Y} = 4,35$ (lít) và $s_Y = 0,2$. So sánh mức xăng tiêu hao của hai dòng xe Z và W. (Mức ý nghĩa $\alpha = 5\%$)

Câu 2. (2 điểm) Theo dõi sự phụ thuộc giữa màu mắt và màu tóc ở 124 phụ nữ ở một nước Châu Âu ta có kết quả sau:

Màu mắt \ Màu tóc	Màu tóc			
	Vàng nâu	Nâu	Đen	Vàng hoe
Xanh	25	9	3	7
Xám	13	17	10	7
Nâu mực	7	13	8	5

Với mức ý nghĩa 5%, hãy kiểm tra giả thuyết cho rằng màu của tóc và màu của mắt độc lập với nhau.



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

ĐỀ THI KẾT THÚC HỌC PHẦN

Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ

(do phòng KT-ĐBCL ghi)

Câu 3. (2 điểm) Xét mẫu ngẫu nhiên $X_1, \dots, X_4$ kích thước 4 được lấy từ phân phối Poisson với trung bình θ .

(a) Chứng tỏ rằng miền bác bỏ C xác định bởi $\sum_{i=1}^4 X_i \geq 3$ là tối ưu để kiểm định $H_0 : \theta = 0,5$ so với $H_1 : \theta = 0,8$.

(b) Với kiểm định trong câu (a), xác định mức ý nghĩa của kiểm định α .

Câu 4. (2 điểm) Gọi X_1, X_2 là một mẫu ngẫu nhiên kích thước 2 được lấy từ tổng thể có phân phối đều $\mathcal{U}(0, \theta)$. Ta định nghĩa thống kê sau:

$$\hat{\theta} = \max(X_1, X_2)$$

(a) Tìm hàm phân phối xác suất của $\hat{\theta}$. Từ đó suy ra hàm mật độ xác suất của $\hat{\theta}$.

(b) Tính kỳ vọng của $\hat{\theta}$. Từ đó suy ra một ước lượng không chệch của θ và ký hiệu là $\tilde{\theta}$.

HẾT.



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-ĐBCL ghi)

CK18191-MTH10401

Tên học phần: Độ đo xác suất Mã HP: MTH10401
Thời gian làm bài: 120 phút Ngày thi: 03/01/2019
Ghi chú: Sinh viên [☐ được phép / ☒ không được phép] sử dụng tài liệu khi làm bài.

Câu 1 (1đ). Sinh viên chọn một trong 2 câu sau: **a)** Hàm số $f(x) = \frac{e^{-x}}{x^2}$ có khả tích

Lebesgue trên $(0, \infty)$ không? **b)** Cho $G(\lambda) = \int_0^1 e^{-\lambda t} g(t) dt$ với $g: [0, 1] \rightarrow \mathbb{R}$ là hàm khả tích. Hỏi G có liên tục trên $(0, \infty)$ không?

Câu 2 (3đ). Cho $f(x) = Cx(2-x)I_{(0,2)}(x)$

- a)** (1 đ). Tìm C để f là hàm mật độ (MĐ) xác suất,
b) (1đ). Gọi X là biến ngẫu nhiên có hàm MĐ $f(x)$, tính EX , $\text{var } X$,
c) (1đ). Cho $Y = 2X + 3$, tìm hàm MĐXS của Y .

Câu 3 (1đ). Một kỳ thi trắc nghiệm môn Thống kê có 50 câu. Một sinh viên loại Khá có thể trả lời đúng một câu trắc nghiệm với xác suất 0.65. Tính xác suất để sinh viên này không rớt (nghĩa là làm đúng tối thiểu 25 câu).

Câu 4 (3đ) Cho biến ngẫu nhiên Poisson $X \sim P(\lambda_1)$, độc lập với $Y \sim P(\lambda_2)$

- (a)** (1đ). Tìm hàm sinh moment của X và Y , đặt là $M_X(t)$ và $M_Y(t)$.
(b) (1đ) Chứng minh rằng $M_{X+Y}(t) = M_X(t)M_Y(t)$
(c) (1đ). Từ đó suy ra rằng $X + Y$ cũng có phân phối Poisson và $X + Y \sim P(\lambda_1 + \lambda_2)$.

Câu 5 (1đ). Các kỹ sư cho rằng trọng lượng W (tính bằng tấn) mà một nhịp cầu có thể chịu đựng và không làm hỏng cấu trúc cầu thỏa phân phối chuẩn với trung bình 400 và độ lệch chuẩn 40. Giả sử trọng lượng một chiếc xe (cũng tính bằng tấn) là biến ngẫu nhiên với trung bình 3 và độ lệch chuẩn 0.3. Giả thiết các biến ngẫu nhiên này độc lập. Ước lượng số xe tối đa trên một nhịp cầu sao cho xác suất làm hỏng cấu trúc cầu không quá 0.05.

Câu 6 (1 đ) Viết code R

- a) mô phỏng một mẫu ngẫu nhiên (2000 phần tử) có phân phối đều $\mathcal{U}([160, 180])$
b) xuất ra biểu đồ histogram cho mẫu vừa tạo được.

Hết

Lưu ý: Thí sinh không được dùng tài liệu.

Tên học phần: Xác suất nâng cao Mã HP: MTH10423
 Thời gian làm bài: 120 phút Ngày thi: 03/01/2019
 Ghi chú: Sinh viên [☒ được phép / ☐ không được phép] sử dụng tài liệu giấy khi làm bài.

Câu 1 (2.5 điểm).

- a) Phát biểu bổ đề Borel-Cantelli.
 b) Cho $X_1, X_2, \dots$ là các biến ngẫu nhiên độc lập cùng phân phối có hàm mật độ như sau

$$f(x) = \begin{cases} \theta x^{-(\theta+1)}, & x \in (1, \infty) \\ 0, & x \notin (1, \infty) \end{cases}$$

Với giá trị nào của θ thì $\frac{X_n}{n} \xrightarrow{a.s} 0$ (hội tụ hầu chắc chắn).

Câu 2 (2.5 điểm).

Cho $X_1, X_2, \dots$ là các biến ngẫu nhiên độc lập cùng phân phối có hàm mật độ như sau

$$f(x) = \begin{cases} e^{-x+\theta}, & x \geq \theta \\ 0, & x < \theta \end{cases}$$

Kí hiệu $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$

- a) Chứng minh rằng $\bar{X}_n \xrightarrow{P} 1 + \theta$.
 b) Chứng minh rằng $\min\{X_1, X_2, \dots, X_n\} \xrightarrow{P} \theta$.

Câu 3 (2.5 điểm).

- a) Phát biểu định lý Slutsky.
 b) Cho $X_1, X_2, \dots$ là các biến ngẫu nhiên độc lập cùng phân phối có kì vọng bằng 0 và phương sai bằng 1. Chứng minh rằng

$$\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{j=1}^n X_j^2}} \xrightarrow{d} N(0,1) \quad (\text{hội tụ theo phân phối}).$$

Câu 4 (2.5 điểm).

Cho $\{X_n, n = 1, 2, \dots\}$ là dãy các biến ngẫu nhiên có phân phối Bernoulli($0.5 + 0.5/n$). Chứng minh rằng X_n hội tụ đến X theo phân phối nhưng X_n không hội tụ đến X theo xác suất với X là biến ngẫu nhiên có phân phối Bernoulli(0.5).



Tên học phần: Thống Kê Kinh Tế

Mã HP: MTH10512

Thời gian làm bài: 75 phút

Ngày thi: 04/01/2019

Ghi chú: Sinh viên [☐ được phép / ☒ không được phép] sử dụng tài liệu khi làm bài.

Họ tên sinh viên: MSSV: STT: ... đề 1

Điểm số	Điểm chữ	Chữ ký CB chấm thi	CBCT1:
			CBCT2:

SINH VIÊN LÀM BÀI TRỰC TIẾP TRÊN ĐỀ VÀ NỘP LẠI ĐỀ THI

- Với mỗi câu hỏi, chỉ có 1 đáp án đúng nhất. Sử dụng bút chì tô kín đáp án được chọn.

Bảng trả lời:

1. (A) (B) (C) (D)	6. (A) (B) (C) (D)	11. (A) (B) (C) (D)	16. (A) (B) (C) (D)
2. (A) (B) (C) (D)	7. (A) (B) (C) (D)	12. (A) (B) (C) (D)	17. (A) (B) (C) (D)
3. (A) (B) (C) (D)	8. (A) (B) (C) (D)	13. (A) (B) (C) (D)	18. (A) (B) (C) (D)
4. (A) (B) (C) (D)	9. (A) (B) (C) (D)	14. (A) (B) (C) (D)	19. (A) (B) (C) (D)
5. (A) (B) (C) (D)	10. (A) (B) (C) (D)	15. (A) (B) (C) (D)	20. (A) (B) (C) (D)

Để trả lời các câu hỏi từ 1 đến 4 xét mô hình ước lượng sau (theo OLS), với *return* là tổng lợi nhuận của việc nắm giữ cổ phiếu của một công ty trong một năm, *dkr* là tỷ lệ nợ trên vốn của công ty, *eps* là thu nhập trên mỗi cổ phiếu, *netinc* là thu nhập ròng và *salary* là tổng thù lao, triệu đô, cho tổng giám đốc điều hành (CEO) (các sai số chuẩn ước lượng của các tham số trong các dấu ngoặc đơn bên dưới các ước lượng). Mô hình được ước lượng bằng cách sử dụng dữ liệu trên $n = 142$ công ty.

$$\widehat{return}_i = -12.3 + 0.32 dkr_i + 0.043 eps_i - 0.005 netinc_i + 0.0035 salary_i$$

(6.89) (0.150) (0.078) (0.0047) (0.0022)

$$n = 142 \quad R^2 = 0.0395$$

- Bạn có thể nói gì về hệ số ước lượng của biến *salary*? (xét đối thuyết một phía cho kiểm định mức ý nghĩa của các tham số và sử dụng xấp xỉ chuẩn)

☐ Nó có ý nghĩa thống kê ở mức 5% nhưng không có ý nghĩa ở mức 1%.

☐ Với mỗi triệu đô thù lao tăng thêm của CEO, lợi nhuận được dự đoán tăng 0.0035, về trung bình, các yếu tố khác giữ nguyên không đổi. Và nó có ý nghĩa thống kê ở mức 1%.



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(đơn vị KT-DBCL ghi)

- ☐ Với mỗi triệu đô thù lao tăng thêm của CEO, lợi nhuận được dự đoán giảm 0.0035, về trung bình, các yếu tố khác giữ nguyên không đổi. Và nó có ý nghĩa thống kê ở mức 5%.
- ☐ Với mỗi triệu đô thù lao tăng thêm của CEO, lợi nhuận được dự đoán tăng 0.0035, về trung bình, các yếu tố khác giữ nguyên không đổi. Nhưng nó không có ý nghĩa thống kê ở mức 5%.
2. Mô hình được ước lượng mà không có các biến dkr và eps , và nhận được $R^2 = 0.0387$. Giá trị của thống kê F khi kiểm định giả thuyết không “các hệ số trên dkr và eps đều bằng không” là?
- ☐ 0.0808
- ☐ Ta không có đủ thông tin để trả lời câu hỏi này, ta sẽ cần phải thu thập nhiều thông tin hơn từ mô hình bị hạn chế này.
- ☐ 0.0570
- ☐ 32.821
3. Hệ số tương quan giữa các giá trị phù hợp (the fitted values) và biến phụ thuộc là
- ☐ 0.1987 ☐ 0.00156025 ☐ 0.0395 ☐ $\pi/12$
4. Bạn có thể nói gì về hệ số của dkr (xét đối thuyết một phía khi kiểm định mức ý nghĩa của các tham số và sử dụng xấp xỉ chuẩn)
- ☐ Nó có ý nghĩa thống kê ở mức 1%
- ☐ Nó có ý nghĩa thống kê ở mức 1% nhưng không có ý nghĩa ở mức 5%
- ☐ Nó có ý nghĩa thống kê ở mức 5% nhưng không có ý nghĩa ở mức 1%
- ☐ Nó có ý nghĩa thống kê ở mức 5% và cũng có ý nghĩa ở mức 1%
5. Nghĩa của thuật ngữ “heteroscedasticity” là?
- ☐ Các sai số không có độc lập tuyến tính với nhau
- ☐ Phương sai của biến phụ thuộc thay đổi
- ☐ Phương sai của sai số thay đổi
- ☐ Các sai số có trung bình khác không
6. Nếu bạn muốn kiểm định, mức ý nghĩa 5%, một hệ số độ dốc cụ thể có bằng 1 hay không, thì bạn nên
- ☐ cộng và trừ 1.96 từ hệ số độ dốc và kiểm tra xem khoảng này có chứa 1 hay không.
- ☐ kiểm tra xem R^2 hiệu chỉnh có gần 1 hay không.
- ☐ xem hệ số độ dốc có nằm giữa 0.95 và 1.05 hay không.
- ☐ trừ 1 từ hệ số được ước lượng, chia hiệu cho sai số chuẩn, và kiểm tra xem tỷ số kết quả có lớn hơn 1.96 hay không.
7. Giả sử rằng hồi quy sau được ước lượng bằng cách sử dụng 27 quan trắc:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

Giá trị tới hạn phù hợp cho kiểm định hai phía 5% của $H_0 : \beta_3 = 1$?

Cho biết: $z_{0.975} = 1.96$, $z_{0.95} = 1.64$, $t_{0.95}^{24} = 1.71$, $t_{0.975}^{24} = 2.06$.

☐ 1.64

☐ 1.71

☐ 2.06

☐ 1.96

(Đề thi gồm 4 trang)

[Trang 2/4]

8. Các giả thuyết tuyến tính sau có thể được kiểm định bằng cách dùng F-test ngoại trừ

☐ $\beta_2 = 1$ và $\beta_3 = \beta_4/\beta_5$

☐ $\beta_2 = 0$

☐ $\beta_0 = \beta_1$ và $\beta_1 = 0$

☐ $\beta_1 + \beta_2 = 1$ và $\beta_3 = -2\beta_4$

9. Trong kiểm định việc loại bỏ đa biến trong mô hình hồi quy bội dưới các giả sử cổ điển, ta có nhiều khả năng bác bỏ giả thuyết không rằng một số hệ số là bằng không nếu:

☐ Tổng bình phương các phần dư của mô hình bị hạn chế là lớn tương đối so với của mô hình không hạn chế

☐ Tổng bình phương các phần dư của mô hình bị hạn chế là nhỏ tương đối so với của mô hình không bị hạn chế

☐ Tổng bình phương toàn phần, SST, là lớn

☐ Hệ số chặn là lớn hơn mức ý nghĩa

10. Nếu một ước lượng được gọi là có phương sai bé nhất, phát biểu nào sau đây KHÔNG ĐÚNG?

☐ Ước lượng là hiệu quả.

☐ Một ước lượng như vậy luôn luôn là không chệch.

☐ Xác suất ước lượng đó cách xa khỏi giá trị đúng đạt cực tiểu.

☐ Ước lượng như vậy được gọi là “tốt nhất”.

11. Lấy một khoảng tin cậy 95% ước lượng cho một tham số của hồi quy tuyến tính bội. Nếu bạn tăng độ tin cậy lên 99% thì:

☐ Chiều dài của khoảng tin cậy giảm

☐ Chiều dài của khoảng tin cậy giữ nguyên không đổi

☐ Chiều dài của khoảng tin cậy tăng

☐ Không đáp án nào đúng

Đối với các câu hỏi 12 đến 16, xét phương trình hồi quy sau được ước lượng bằng cách dùng 1,000 quan trắc:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad (1)$$

12. Mô hình nào sau đây là mô hình bị hạn chế để kiểm định giả thuyết không $H_0 : \beta_2 + \beta_3 = 1$?

☐ Mô hình bị hạn chế sẽ là mô hình như trong phương trình (1) ở trên

☐ $y_t = \beta_1 + \beta_3(x_{3t} - x_{2t}) + \beta_4 x_{4t} + u_t$

☐ $(y_t - x_{2t}) = \beta_1 + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$

☐ $(y_t - x_{2t}) = \beta_1 + \beta_3(x_{3t} - x_{2t}) + \beta_4 x_{4t} + u_t$

13. Giả thuyết không nào sau đây có thể được kiểm định bằng cách sử dụng F-test?

i) $\beta_2 = 1$

ii) $\beta_3^2 = 1$

iii) $\beta_4 = -\beta_2$

iv) $\beta_3\beta_4 = 0$

☐ chỉ (i), (ii), và (iii)

☐ (i), (ii), (iii), và (iv)

☐ chỉ (ii) và (iv)

☐ chỉ (i) và (iii)

(Đề thi gồm 4 trang)

... [Trang 3/4]

.....



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-ĐBCL ghi)

14. Giả sử rằng kiểm định trong câu hỏi (12) được thực hiện, giá trị tới hạn nào được dùng để so sánh với thống kê kiểm định?
☐ A 254 ☐ B 3.00 ☐ C 253 ☐ D 3.84
15. Giả sử rằng kiểm định trong câu hỏi (12) được thực hiện, và hai tổng bình phương phần dư cần có là 30.2 và 28.1, thống kê kiểm định F là?
☐ A -74.4 ☐ B 37.2 ☐ C 74.4 ☐ D -37.2
16. Giả thuyết không cho kiểm định F cho phương trình (1) ở trên là?
☐ A $\beta_1 = 0$ hoặc $\beta_2 = 0$ hoặc $\beta_3 = 0$ hoặc $\beta_4 = 0$ ☐ B $\beta_1 = 0$ và $\beta_2 = 0$ và $\beta_3 = 0$ và $\beta_4 = 0$
☐ C $\beta_2 = 0$ hoặc $\beta_3 = 0$ hoặc $\beta_4 = 0$ ☐ D $\beta_2 = 0$ và $\beta_3 = 0$ và $\beta_4 = 0$
17. Dưới các giả thiết bình phương bé nhất cho bài toán hồi quy bội (MLR. 1-4), các ước lượng OLS cho hệ số độ dốc và hệ số chặn
☐ A có phân phối chuẩn trong mẫu nhỏ chừng nào các sai số có phương sai không đổi.
☐ B là không chệch và vững.
☐ C có phân phối chuẩn chính xác với $n > 25$.
☐ D là BLUE.
18. Khi bạn có một vấn đề biến bị bỏ sót, giả sử rằng $E(u_i|X_i) = 0$ bị vi phạm. Điều này suy ra
☐ A tổng các phần dư nhân với bất kỳ biến giải thích nào không còn bằng không.
☐ B có một ước lượng khác được gọi là bình phương bé nhất có trọng số (weighted least squares), là BLUE.
☐ C ước lượng OLS không còn vững.
☐ D tổng các phần dư không còn bằng không.
19. Điều nào sau đây có thể là các hậu quả của một hoặc vài giả thiết MLR bị vi phạm?
i) Các ước lượng OLS không còn tối ưu
ii) Các ước lượng sai số chuẩn không còn tối ưu
iii) Các phân phối được giả sử cho các thống kê kiểm định không còn phù hợp
iv) Các kết luận về độ mạnh yếu của mối quan hệ giữa các biến độc lập và phụ thuộc có thể không hợp lệ.
☐ A (i) và (iii) ☐ B (i), (ii), (iii), và (iv) ☐ C (ii) và (iv) ☐ D (i), (ii), và (iii)
20. Nếu một ước lượng là vững, thì
☐ A các ước lượng sẽ hội tụ về giá trị thật sự khi cỡ mẫu tăng lên.
☐ B các ước lượng OLS sẽ gần với các giá trị đúng khi mẫu nhỏ và lớn.
☐ C về trung bình, các ước lượng OLS sẽ bằng với các giá trị thật sự.
☐ D ước lượng OLS là không chệch và không có ước lượng không chệch nào khác có phương sai nhỏ hơn.



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-DBCL ghi)

CK18191.MTH10512

Tên học phần: Thông Kê Kinh Tế

Mã HP: MTH10512

Thời gian làm bài: 75 phút

Ngày thi: 04/01/2019

Ghi chú: Sinh viên [☐ được phép / ☒ không được phép] sử dụng tài liệu khi làm bài.

Họ tên sinh viên: MSSV: STT: ... Đề 2

Điểm số	Điểm chữ	Chữ ký CB chấm thi	CBCT1:
			CBCT2:

SINH VIÊN LÀM BÀI TRỰC TIẾP TRÊN ĐỀ VÀ NỘP LẠI ĐỀ THI

- Với mỗi câu hỏi, chỉ có 1 đáp án đúng nhất. Sử dụng bút chì tô kín đáp án được chọn.

Bảng trả lời:

1. (A) (B) (C) (D)	6. (A) (B) (C) (D)	11. (A) (B) (C) (D)	16. (A) (B) (C) (D)
2. (A) (B) (C) (D)	7. (A) (B) (C) (D)	12. (A) (B) (C) (D)	17. (A) (B) (C) (D)
3. (A) (B) (C) (D)	8. (A) (B) (C) (D)	13. (A) (B) (C) (D)	18. (A) (B) (C) (D)
4. (A) (B) (C) (D)	9. (A) (B) (C) (D)	14. (A) (B) (C) (D)	19. (A) (B) (C) (D)
5. (A) (B) (C) (D)	10. (A) (B) (C) (D)	15. (A) (B) (C) (D)	20. (A) (B) (C) (D)

- Dưới các giả thiết bình phương bé nhất cho bài toán hồi quy bội (MLR. 1-4), các ước lượng OLS cho hệ số độ dốc và hệ số chặn
(A) là BLUE.
(B) có phân phối chuẩn chính xác với $n > 25$.
(C) có phân phối chuẩn trong mẫu nhỏ chừng nào các sai số có phương sai không đổi.
(D) là không chệch và vững.
- Khi bạn có một vấn đề biến bị bỏ sót, giả sử rằng $E(u_i|X_i) = 0$ bị vi phạm. Điều này suy ra
(A) tổng các phần dư nhân với bất kỳ biến giải thích nào không còn bằng không.
(B) tổng các phần dư không còn bằng không.
(C) có một ước lượng khác được gọi là bình phương bé nhất có trọng số (weighted least squares), là BLUE.
(D) ước lượng OLS không còn vững.

Đề thi gồm 4 trang)

[Trang 1/4]



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-ĐBCL ghi)

3. Điều nào sau đây có thể là các hậu quả của một hoặc vài giả thiết MLR bị vi phạm?

- i) Các ước lượng OLS không còn tối ưu
- ii) Các ước lượng sai số chuẩn không còn tối ưu
- iii) Các phân phối được giả sử cho các thống kê kiểm định không còn phù hợp
- iv) Các kết luận về độ mạnh yếu của mối quan hệ giữa các biến độc lập và phụ thuộc có thể không hợp lệ.

☐ A (i) và (iii) ☐ B (ii) và (iv) ☐ C (i), (ii), và (iii) ☐ D (i), (ii), (iii), và (iv)

4. Nếu một ước lượng là vững, thì

- ☐ A các ước lượng OLS sẽ gần với các giá trị đúng khi mẫu nhỏ và lớn.
- ☐ B ước lượng OLS là không chệch và không có ước lượng không chệch nào khác có phương sai nhỏ hơn.
- ☐ C về trung bình, các ước lượng OLS sẽ bằng với các giá trị thật sự.
- ☐ D các ước lượng sẽ hội tụ về giá trị thật sự khi cỡ mẫu tăng lên.

Để trả lời các câu hỏi từ 5 đến 8 xét mô hình ước lượng sau (theo OLS), với *return* là tổng lợi nhuận của việc nắm giữ cổ phiếu của một công ty trong một năm, *dkr* là tỷ lệ nợ trên vốn của công ty, *eps* là thu nhập trên mỗi cổ phiếu, *netinc* là thu nhập ròng và *salary* là tổng thù lao, triệu đô, cho tổng giám đốc điều hành (CEO) (các sai số chuẩn ước lượng của các tham số trong các dấu ngoặc đơn bên dưới các ước lượng). Mô hình được ước lượng bằng cách sử dụng dữ liệu trên $n = 142$ công ty.

$$\widehat{return}_i = -12.3 + 0.32 dkr_i + 0.043 eps_i - 0.005 netinc_i + 0.0035 salary_i$$

$n = 142 \quad R^2 = 0.0395$

(6.89) (0.150) (0.078) (0.0047) (0.0022)

5. Bạn có thể nói gì về hệ số ước lượng của biến *salary*? (xét đối thuyết một phía cho kiểm định mức ý nghĩa của các tham số và sử dụng xấp xỉ chuẩn)

- ☐ A Nó có ý nghĩa thống kê ở mức 5% nhưng không có ý nghĩa ở mức 1%.
- ☐ B Với mỗi triệu đô thù lao tăng thêm của CEO, lợi nhuận được dự đoán tăng 0.0035, về trung bình, các yếu tố khác giữ nguyên không đổi. Và nó có ý nghĩa thống kê ở mức 1%.
- ☐ C Với mỗi triệu đô thù lao tăng thêm của CEO, lợi nhuận được dự đoán giảm 0.0035, về trung bình, các yếu tố khác giữ nguyên không đổi. Và nó có ý nghĩa thống kê ở mức 5%.
- ☐ D Với mỗi triệu đô thù lao tăng thêm của CEO, lợi nhuận được dự đoán tăng 0.0035, về trung bình, các yếu tố khác giữ nguyên không đổi. Nhưng nó không có ý nghĩa thống kê ở mức 5%.

6. Mô hình được ước lượng mà không có các biến *dkr* và *eps*, và nhận được $R^2 = 0.0387$. Giá trị của thống kê F khi kiểm định giả thuyết không “các hệ số trên *dkr* và *eps* đều bằng không” là?

- ☐ A 0.0570
- ☐ B 0.0808
- ☐ C Ta không có đủ thông tin để trả lời câu hỏi này, ta sẽ cần phải thu thập nhiều thông tin hơn từ mô hình bị hạn chế này.
- ☐ D 32.821

7. Hệ số tương quan giữa các giá trị phù hợp (the fitted values) và biến phụ thuộc là
☐ A $\pi/12$ ☐ B 0.00156025 ☐ C 0.0395 ☐ D 0.1987
8. Bạn có thể nói gì về hệ số của dkr (xét đối thuyết một phía khi kiểm định mức ý nghĩa của các tham số và sử dụng xấp xỉ chuẩn)
☐ A Nó có ý nghĩa thống kê ở mức 1%
☐ B Nó có ý nghĩa thống kê ở mức 5% nhưng không có ý nghĩa ở mức 1%
☐ C Nó có ý nghĩa thống kê ở mức 1% nhưng không có ý nghĩa ở mức 5%
☐ D Nó có ý nghĩa thống kê ở mức 5% và cũng có ý nghĩa ở mức 1%
9. Nghĩa của thuật ngữ “heteroscedasticity” là?
☐ A Phương sai của biến phụ thuộc thay đổi
☐ B Các sai số có trung bình khác không
☐ C Các sai số không có độc lập tuyến tính với nhau
☐ D Phương sai của sai số thay đổi
10. Nếu bạn muốn kiểm định, mức ý nghĩa 5%, một hệ số độ dốc cụ thể có bằng 1 hay không, thì bạn nên
☐ A cộng và trừ 1.96 từ hệ số độ dốc và kiểm tra xem khoảng này có chứa 1 hay không.
☐ B trừ 1 từ hệ số được ước lượng, chia hiệu cho sai số chuẩn, và kiểm tra xem tỷ số kết quả có lớn hơn 1.96 hay không.
☐ C kiểm tra xem R^2 hiệu chỉnh có gần 1 hay không.
☐ D xem hệ số độ dốc có nằm giữa 0.95 và 1.05 hay không.
11. Giả sử rằng hồi quy sau được ước lượng bằng cách sử dụng 27 quan trắc:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

 Giá trị tối hạn phù hợp cho kiểm định hai phía 5% của $H_0 : \beta_3 = 1$?
 Cho biết: $z_{0.975} = 1.96$, $z_{0.95} = 1.64$, $t_{0.95}^{24} = 1.71$, $t_{0.975}^{24} = 2.06$.
☐ A 2.06 ☐ B 1.71 ☐ C 1.96 ☐ D 1.64
12. Các giả thuyết tuyến tính sau có thể được kiểm định bằng cách dùng F-test ngoại trừ
☐ A $\beta_1 + \beta_2 = 1$ và $\beta_3 = -2\beta_4$ ☐ B $\beta_0 = \beta_1$ và $\beta_1 = 0$
☐ C $\beta_2 = 0$ ☐ D $\beta_2 = 1$ và $\beta_3 = \beta_4/\beta_5$
13. Trong kiểm định việc loại bỏ đa biến trong mô hình hồi quy bội dưới các giả sử cổ điển, ta có nhiều khả năng bác bỏ giả thuyết không rằng một số hệ số là bằng không nếu:
☐ A Tổng bình phương các phần dư của mô hình bị hạn chế là lớn tương đối so với của mô hình không hạn chế
☐ B Tổng bình phương các phần dư của mô hình bị hạn chế là nhỏ tương đối so với của mô hình không bị hạn chế
☐ C Tổng bình phương toàn phần, SST, là lớn
☐ D Hệ số chặn là lớn hơn mức ý nghĩa



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-ĐBCL ghi)

14. Nếu một ước lượng được gọi là có phương sai bé nhất, phát biểu nào sau đây KHÔNG ĐÚNG?
- ☐ A Ước lượng là hiệu quả.
 - ☐ B Xác suất ước lượng đó cách xa khỏi giá trị đúng đạt cực tiểu.
 - ☐ C Một ước lượng như vậy luôn luôn là không chệch.
 - ☐ D Ước lượng như vậy được gọi là “tốt nhất”.

15. Lấy một khoảng tin cậy 95% ước lượng cho một tham số của hồi quy tuyến tính bội. Nếu bạn tăng độ tin cậy lên 99% thì:
- ☐ A Chiều dài của khoảng tin cậy giảm
 - ☐ B Chiều dài của khoảng tin cậy tăng
 - ☐ C Chiều dài của khoảng tin cậy giữ nguyên không đổi
 - ☐ D Không đáp án nào đúng

Đối với các câu hỏi 16 đến 20, xét phương trình hồi quy sau được ước lượng bằng cách dùng 1,000 quan trắc:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \quad (1)$$

16. Mô hình nào sau đây là mô hình bị hạn chế để kiểm định giả thuyết không $H_0 : \beta_2 + \beta_3 = 1$?

- ☐ A Mô hình bị hạn chế sẽ là mô hình như trong phương trình (1) ở trên
- ☐ B $(y_t - x_{2t}) = \beta_1 + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$
- ☐ C $y_t = \beta_1 + \beta_3(x_{3t} - x_{2t}) + \beta_4 x_{4t} + u_t$
- ☐ D $(y_t - x_{2t}) = \beta_1 + \beta_3(x_{3t} - x_{2t}) + \beta_4 x_{4t} + u_t$

17. Giả thuyết không nào sau đây có thể được kiểm định bằng cách sử dụng F-test?

- i) $\beta_2 = 1$
- ii) $\beta_3^2 = 1$
- iii) $\beta_4 = -\beta_2$
- iv) $\beta_3 \beta_4 = 0$

- ☐ A chỉ (i) và (iii)
- ☐ B chỉ (ii) và (iv)
- ☐ C (i), (ii), (iii), và (iv)
- ☐ D chỉ (i), (ii), và (iii)

18. Giả sử rằng kiểm định trong câu hỏi (16) được thực hiện, giá trị tới hạn nào được dùng để so sánh với thống kê kiểm định?

- ☐ A 254
- ☐ B 253
- ☐ C 3.84
- ☐ D 3.00

19. Giả sử rằng kiểm định trong câu hỏi (16) được thực hiện, và hai tổng bình phương phần dư cần có là 30.2 và 28.1, thống kê kiểm định F là?

- ☐ A -74.4
- ☐ B 37.2
- ☐ C -37.2
- ☐ D 74.4

20. Giả thuyết không cho kiểm định F cho phương trình (1) ở trên là?

- ☐ A $\beta_1 = 0$ hoặc $\beta_2 = 0$ hoặc $\beta_3 = 0$ hoặc $\beta_4 = 0$
- ☐ B $\beta_2 = 0$ hoặc $\beta_3 = 0$ hoặc $\beta_4 = 0$
- ☐ C $\beta_2 = 0$ và $\beta_3 = 0$ và $\beta_4 = 0$
- ☐ D $\beta_1 = 0$ và $\beta_2 = 0$ và $\beta_3 = 0$ và $\beta_4 = 0$



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
(do phòng KT-DBCL ghi)

CK18191.MTH00040

Tên học phần: **Xác Suất Thống Kê**

Mã HP: **MTH00040**

Thời gian làm bài: 90 phút

Ngày thi: **28/12/2018, 7g45**

Ghi chú: Sinh viên [☐ được phép / ☒ không được phép] sử dụng tài liệu khi làm bài.

Họ tên sinh viên: MSSV: STT: ...

1. (7 điểm)

Gọi X (giờ) là thời gian tự học hàng ngày của sinh viên, khảo sát 120 sinh viên trường Đại học KHXHNV. Kết quả cho bởi bảng sau

Thời gian tự học (giờ)	1	2	3	4	5	6	7	8
Số sinh viên	13	18	14	23	15	16	17	4

Giả sử thời gian tự học của sinh viên có phân phối chuẩn.

(a) Ước lượng thời gian tự học trung bình của sinh viên trường KHXHNV với độ tin cậy 98%.
(1,5 đ)

(b) Với độ tin cậy 95%, ước lượng tỷ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày. (1,5 đ)

(c) Khảo sát thời gian tự học của 90 sinh viên trường Đại học Kinh tế

Thời gian tự học (giờ)	1	2	3	4	5	6
Số sinh viên	7	8	17	24	20	14

Có ý kiến cho rằng thời gian tự học của sinh viên trường KHXHNV lớn hơn sinh viên trường Kinh Tế. Với mức ý nghĩa 5%, hãy kiểm định ý kiến trên. (2 đ)

(d) So sánh tỷ lệ những sinh viên có thời gian tự học trên 5 giờ mỗi ngày giữa hai trường KHXHNV và Kinh Tế. ($\alpha = 1\%$) (2 đ)

2. (3 điểm)

Trong cấu tạo một loại dây thừng, người ta quan tâm đến hàm lượng nylon x (đv: %) ảnh hưởng như thế nào đến lực căng y (đv: psi) (lực kéo tối đa trước khi sợi dây bị đứt). Số liệu bên dưới cho kết quả đo tương ứng (x, y) của 8 sợi dây

Hàm lượng nylon	0	10	20	20	30	40	50	50
Lực căng	160	240	320	340	395	450	510	520

(a) Tìm phương trình hồi quy tuyến tính biểu diễn mối liên hệ của x và y dưới dạng:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

(2 đ)

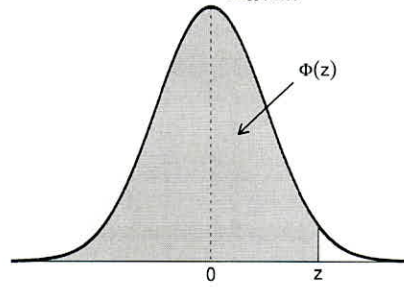
(b) Giải thích ý nghĩa của hệ số $\hat{\beta}_1$ nhận được và dự đoán lực căng của một sợi dây có hàm lượng nylon bằng 45. (1 đ)

HẾT.

thi gồm 3 trang)

[Trang 1/3]

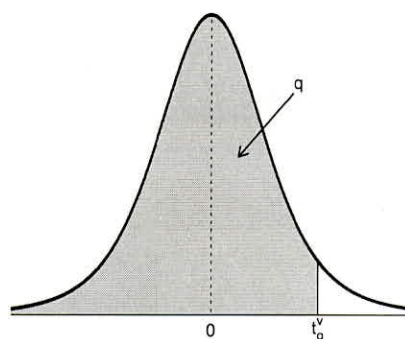
$$\Phi(z) = P(Z < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$



Bảng A.3: Phân phối chuẩn tắc (tt)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

*Với $z \geq 3.50$, xác suất sẽ lớn hơn hoặc bằng 0.9998.



Bảng A.4: Phân vị t_q^v của phân phối Student

$v \backslash q$	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.3249	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	0.2887	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	0.2767	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240
4	0.2707	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103
5	0.2672	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	0.2648	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	0.2632	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	0.2619	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	0.2610	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	0.2579	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495
21	0.2566	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314	3.8193
22	0.2564	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921
23	0.2563	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676
24	0.2562	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454
25	0.2561	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251
26	0.2560	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066
27	0.2559	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896
28	0.2558	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739
29	0.2557	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594
30	0.2556	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460
40	0.2550	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045	3.5510
60	0.2545	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603	3.4602
120	0.2539	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174	3.3735



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
 Học kỳ I – Năm học 2018-2019

MÃ LƯU TRỮ
 (do phòng KT-DBCL ghi)

CK18191-MTH00040

Tên học phần: **Xác Suất Thống Kê**

Mã HP: **MTH00040**

Thời gian làm bài: **90 phút**

Ngày thi: **28/12/2018, 9g55**

Ghi chú: Sinh viên [☐ được phép / ☒ không được phép] sử dụng tài liệu khi làm bài.

Họ tên sinh viên: MSSV: STT: ...

1. (7 điểm)

Thực hiện một khảo sát xã hội về số tiền chi trả cho các hoạt động vui chơi giải trí trong 1 tháng của 400 thanh niên tại TP. HCM người ta thu được bảng sau:

Số tiền (USD)	50 - 80	80 - 120	120 - 160	160 - 200	200 - 220	220 - 250
Số người	50	80	100	80	60	30

Giả thiết số tiền phải bỏ ra cho các hoạt động vui chơi giải trí trong 1 tháng của 1 thanh niên tại TP. HCM là một đại lượng ngẫu nhiên phân phối theo qui luật chuẩn.

- Ước lượng số tiền trung bình một thanh niên phải bỏ ra với độ tin cậy 95%. **(1,5 đ)**
- Những thanh niên bỏ ra trên 200 USD/tháng cho các hoạt động vui chơi là những thanh niên khá giả. Hãy ước lượng tỉ lệ những thanh niên khá giả với độ tin cậy 97%. Nếu muốn sai số ≤ 0.1 thì phải khảo sát thêm bao nhiêu thanh niên? **(2,5 đ)**
- Một nhà nghiên cứu xã hội cho rằng cứ 100 thanh niên ở TP. HCM thì có 30 người thuộc diện khá giả, trong khi nhà thống kê lại tỏ ra nghi ngờ và họ cho rằng con số này thực sự phải nhỏ hơn con số thống kê do nhà nghiên cứu này đưa ra. Vậy theo các bạn, ý kiến nào là đúng đắn với mức ý nghĩa $\alpha = 5\%$. **(1,5 đ)**
- Lời khẳng định: "Tỷ lệ thanh niên có thu nhập hạn chế là 50%" có được chấp nhận hay không, mức ý nghĩa 1%. Biết rằng một thanh niên được gọi là có thu nhập hạn chế nếu số tiền bỏ ra cho hoạt động vui chơi dưới 120 USD/tháng. **(1,5 đ)**

2. (3 điểm)

Bảng số liệu bên dưới mô tả về chỉ số khối cơ thể x (Body Mass Index - BMI) và huyết áp y của 8 người được chọn ngẫu nhiên. BMI được tính bởi công thức (cân nặng (kg)/chiều cao (m))².

BMI	20.3	22.0	26.4	28.2	31	32.6	17.6	19.4
Huyết áp	116	110	131	136	144	138	122	115

- Tìm phương trình hồi quy tuyến tính biểu diễn mối liên hệ của x và y dưới dạng:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

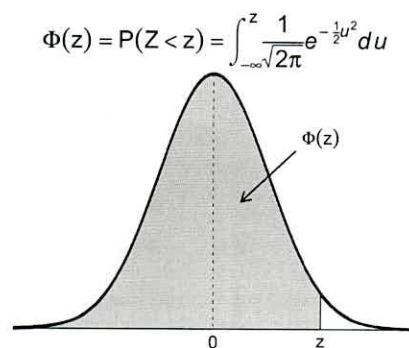
(2 đ)

- Giải thích ý nghĩa của hệ số $\hat{\beta}_1$ nhận được và dự đoán huyết áp của một người có BMI bằng 30. **(1 đ)**

HẾT.

(Đề thi gồm 3 trang)

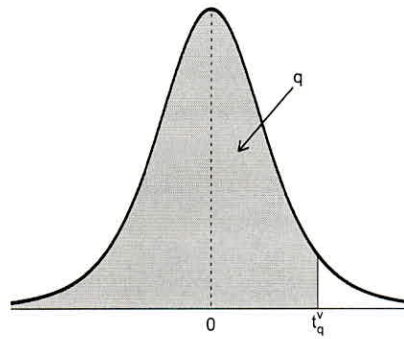
[Trang 1/3]



Bảng A.3: Phân phối chuẩn tắc (tt)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

*Với $z \geq 3.50$, xác suất sẽ lớn hơn hoặc bằng 0.9998.



Bảng A.4: Phân vị t_q^v của phân phối Student

$v \backslash q$	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.3249	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	0.2887	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	0.2767	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240
4	0.2707	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103
5	0.2672	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	0.2648	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	0.2632	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	0.2619	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	0.2610	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	0.2579	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495
21	0.2566	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314	3.8193
22	0.2564	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921
23	0.2563	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676
24	0.2562	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454
25	0.2561	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251
26	0.2560	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066
27	0.2559	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896
28	0.2558	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739
29	0.2557	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594
30	0.2556	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460
40	0.2550	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045	3.5510
60	0.2545	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603	3.4602
120	0.2539	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174	3.3735



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

ĐỀ THI KẾT THÚC HỌC PHẦN

Học kỳ I – Năm học: 2018-2019

MÃ LƯU TRỮ
(do Phòng KT-ĐBCL ghi)

CK18191-MTH00085

Tên học phần: THỰC HÀNH XÁC SUẤT THÔNG KÊ

Mã HP: MTH00085

Thời gian làm bài: 60 phút

Ngày thi: 05/01/2019

Ghi chú: Sinh viên không được sử dụng tài liệu. Đề thi gồm 8 trang.

Họ tên sinh viên: _____ MSSV: _____ STT: _____

Câu 1. Câu lệnh R nào dùng để mô phỏng việc tung đồng xu cân đối đồng chất 5 lần.

- A. `sample(c('H','T'), 5)`.
 B. `sample(c(0,1), 5, prob=c(1/2,1/2))`.
 C. `sample(c(0,1), 5, replace= T)`.
 D. `sample(5, c('H','T'), replace= T)`.

Cho đoạn code sau, trả lời các câu hỏi 2-3.

```
x <- c(rep(1, 3), rep(2, 4), rep(3,5), rep(4,4), rep(5,3))
a = mean(x)
b = length(x)
c = median(x)
```

Câu 2. Kết quả của $a * b$ bằng và $a * c$ bằngCâu 3. Giá trị của $a * b - c$ bằng

- A. 54. B. 24. C. 10. D. 60.

Câu 4. Trong các đoạn code R sau, đoạn nào trả ra kết quả là TRUE .

- A. `(2 != 2) && (3 <= 3)` . B. `(3 != 4) || (4 >= 3)` .
 C. `(2 <= 5) && (4 <= 3)` . D. `isTRUE(4 == 5)` .

Cho đoạn code sau, trả lời các câu hỏi : 5 - 6

```
perm <- function(n){
  if(n==1){
    return(matrix(1))
  } else {
    sp <- perm(n-1)
    p <- nrow(sp)
    A <- matrix(nrow=n*p, ncol=n)
    for(i in 1:n){
      A[(i-1)*p+1:p,] <- cbind(i, sp+(sp>=i))
    }
    return(A)
  }
}
Z=perm(3);
```

Câu 5. Cho biết số dòng của ma trận Z

- A. 3. B. 9. C. 8. D. 6.

Câu 6. Tìm số tự nhiên n sao cho ma trận $perm(n)$ có 24 dòng.

- A. 3. B. 4. C. 12. D. 24.

Cho đoạn code sau, trả lời các câu hỏi 7-8.

```
df <- data.frame(
  STT = c(seq(1,6,1)),
  Name= c("Pearson", "Neymann", "Fisher", "Gosset", "Bayes", "Poisson"),
  Birthday = c(1857, 1894, 1890, 1876, 1702, 1781),
  Died=c(1936, 1981, 1962, 1937, 1761, 1840)
)
```

Câu 7. Dựa vào dataframe df, hãy cho biết năm sinh và năm mất của cha đẻ kiểm định t (Gosset, William Sealy Gosset).

- A. 1890-1962. B. 1936-1937. C. 1876-1937. D. 1890-1962.

Câu 8. Lệnh subset(Name, Birthday%%25>13) cho ra tên của hai nhà thống kê đã có quan điểm đối lập nhau trong vấn đề kiểm định:

- A. [1] "Bayes" "Poisson". B. [1] "Pearson" "Neymann".
C. [1] "Gosset" "Poisson". D. [1] "Neymann" "Fisher".

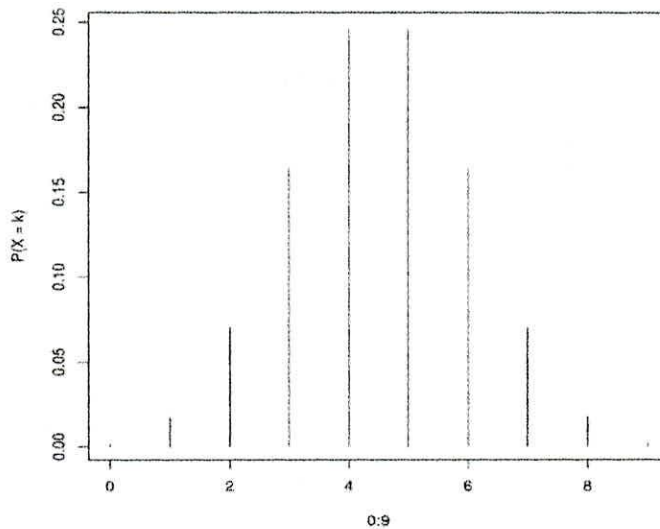
Câu 9. Câu lệnh R: 3*length(sample(1:6, 5))+10 cho ra kết quả nào sau đây?

- A. 25. B. 28. C. 43. D. 13.

Câu 10. Giá trị của ppois(x0, lambda) bằng với

- A. Giá trị của hàm phân phối (tích lũy) của biến ngẫu nhiên phân phối Poisson $\mathcal{P}(\lambda)$ tại x_0 .
B. $\sum_{k \in \mathbb{Z}: 0 \leq k \leq x_0} \frac{e^{-\lambda} \lambda^k}{k!}$.
C. $\mathbb{P}(X \leq x_0)$ trong đó $X \sim \mathcal{P}(\lambda)$.
D. Tất cả các giá trị liệt kê ở trên.

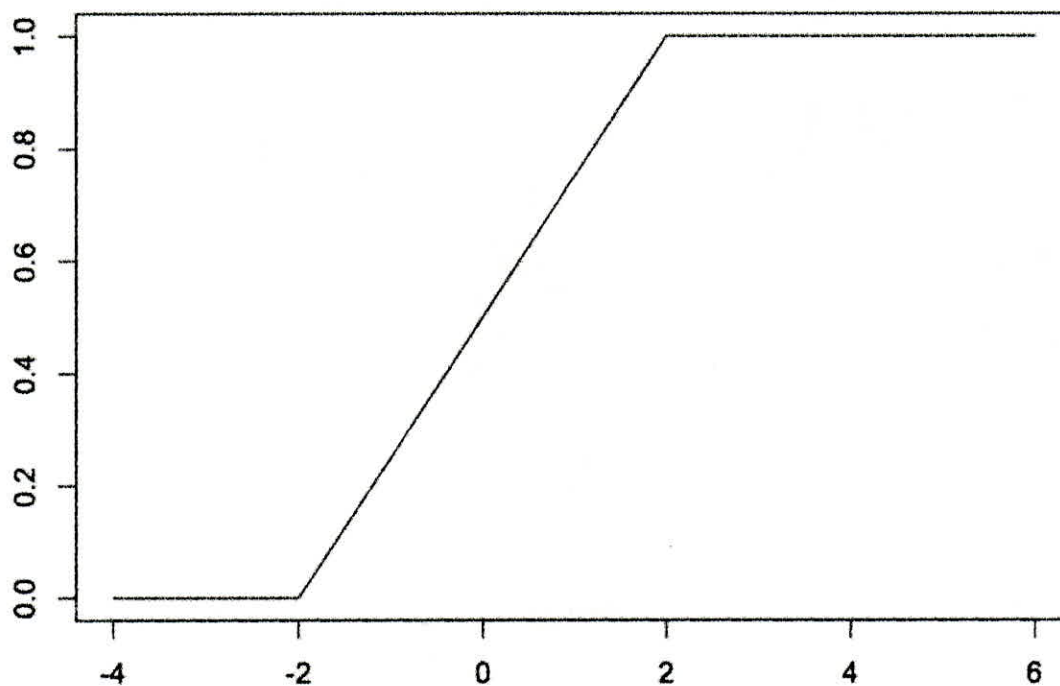
Câu 11. Trong các lệnh sau, lệnh nào có thể vẽ được hình bên dưới.



Hình 1: Biểu đồ cột

- A. curve(dbinom(x, 9, 0.5), from = 0, to = 10).
B. curve(dnorm(x, 2, 1), from = -1, to = 5).
C. hist(c(0:9), dbinom(0:9, 9, 0.5)).
D. plot(0:9, dbinom(0:9, 9, 0.5), type='h', ylab = "P(X = k)").

Câu 12. Cho biết đồ thị hàm phân phối (tích lũy), như Hình 2, là của phân phối nào



Hình 2: Hàm phân phối tích lũy

A. PP chuẩn $\mathcal{N}(1, 1)$.
C. PP mũ $\text{Exp}(2)$.

B. PP đều $U([-2, 2])$.
D. PP Student(10).

Câu 13. Để phát sinh hàm mật độ của biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu = 2, \sigma^2 = 1)$ và mẫu 1000 phần tử là biến ngẫu nhiên có phân phối Poisson $\mathcal{P}(\lambda = 2)$ lần lượt bằng các lệnh và

Đề thi cuối kỳ môn THỰC HÀNH XÁC SUẤT THỐNG KÊ dạng trắc nghiệm có 50 câu hỏi, mỗi câu 5 đáp án. Sinh viên A không học bài, khi đi thi thì xác suất trả lời đúng của mỗi câu là như nhau. Sinh viên B học khá trong lớp, cảm thấy xác suất để mình chọn đúng mỗi câu là 0.6. Đặt

$$X = \sum_{i=1}^{50} X_i \text{ và } Y = \sum_{i=1}^{50} Y_i$$

trong đó, X_i, Y_i lần lượt là các biến ngẫu nhiên phản ánh kết quả chọn câu thứ i là đúng của sinh viên A và sinh viên B (với $i = 1, \dots, 50$), trả lời các câu hỏi 14 và 15

Câu 14. X và Y có phân phối gì?

- A. Nhị thức $B(50, 0.2)$ và $B(50, 0.4)$. B. Nhị thức $B(50, 0.2)$ và $B(50, 0.6)$.
- C. Nhị thức $B(50, 0.8)$ và $B(50, 0.6)$. D. Nhị thức $B(50, 0.8)$ và $B(50, 0.4)$.

Câu 15. Sinh viên sẽ qua môn nếu trả lời đúng ít nhất 25 câu hỏi. Xác suất sinh viên A không qua môn; được xấp xỉ bằng câu lệnh nào sau đây,

- A. `pnorm(15/sqrt(8))` . B. `pnorm(15/8)` .
- C. `1-pnorm(15/sqrt(8))` . D. `dnorm(0.5)` .

Câu 16. Một nhóm sinh viên đo nhiệt độ ở những độ cao khác nhau và thu được bảng số liệu sau.

Elevation(ft)	600	1000	1250	1600	1800	2100	2500	2900
Temperature(F)	56	54	56	50	47	49	47	45

Sử dụng các câu lệnh R để vẽ đồ thị phân tán và đường hồi quy nhiệt độ theo độ cao cùng hệ trục tọa độ.

Câu 17. Các cuộc gọi điện đến tổng đài tuân theo phân phối Poisson với mức λ trên mỗi phút. Từ kinh nghiệm có được trong quá khứ, ta biết rằng xác suất nhận được chính xác một cuộc gọi trong một phút bằng ba lần xác suất không nhận được cuộc gọi nào trong cùng thời gian. Ta xét khoảng 100 khoảng thời gian một phút liên tiếp và gọi U là số khoảng thời gian một phút không nhận được cuộc gọi nào. Viết câu lệnh tính $\mathbb{P}(U \leq 1)$

Câu 18. Cho

$$z_{1-\alpha/2} \triangleq \text{qnorm}(1 - \alpha/2) \text{ và } t_{1-\alpha/2}^{n-1} \triangleq \text{qt}(1 - \alpha/2, \text{df} = n-1)$$

hoàn thành các chỗ trống trong đoạn code sau

```
path = 'D://Works'
setwd(path)
dtf = read.csv('data01.csv', header=TRUE)
Age = dtf$Age
KTC_mean <- function(data, alpha, sig = 'None'){
  n = length(data)
  m = mean(data)
  sd = sd(data)
  zalp = qnorm(1 - ...../2)
  talp = qt(1 - alpha/2, ..... )
  if(sig != 'None')
    eps = sig*...../sqrt(n)
  else if(sig == 'None')
    if( n < 30)
      eps = sd*talp/sqrt(n)
    else if (n >= ..... )
      eps = sd*...../sqrt(n)
  return(c(m - eps, m + eps))
}
KTC_mean(Age, 0.05)
```

Câu 19. Hàm KTC_mean cho biết

- A. Input các tham số dữ liệu mẫu (data), độ tin cậy (alpha) và giả thiết về sigma .
- B. Output là khoảng tin cậy của trung bình trong các trường hợp biết phương sai, không biết phương sai và cỡ mẫu.
- C. A, B đều sai .
- D. A, B đều đúng.

Câu 20. path = 'D://Works'

```
setwd(path)
dtf = read.csv('data01.csv', header=TRUE)
Age = dtf$Age
U70 = Age[Age > 70]
KTC_prop <- function(data.p, data, alpha){
  phat = length(data.p)/ length(data) ## ty le mau
  eps = qnorm(1 - alpha/2)*sqrt(phat*(1-phat)/n)
  print('KTC cho ty le la')
  return(c(phat - eps, phat + eps))
}
KTC_prop(U70, Age, 0.05)
```

Hàm KTC_prop cho biết

- A. Input các tham số dữ liệu mẫu (data), dữ liệu thỏa tính chất nào đó để truy xuất tỷ lệ mẫu (datap) và mức ý nghĩa (alpha) .

- B. Output là khoảng tin cậy cho tỷ lệ p với độ tin cậy α .
- C. A, B đều đúng.
- D. A, B đều sai.

Xem đoạn code và kết quả sau

```
path = 'D://Works'
setwd(path)
data = read.csv('rocket.motor.csv', header=TRUE)
SK = data$streng; mu_0 = 2000
test = t.test(SK, alternative = "two.sided", mu = mu_0, conf.level = 0.95)
```

One Sample t-test

```
data: SK
t = 1.9799, df = 19, p-value = 0.06238
alternative hypothesis: true mean is not equal to 2000
95 percent confidence interval:
 1992.438 2272.377
sample estimates:
mean of x
 2132.407
```

Câu 21. Hàm `t.test(Age, alternative = "two.sided", mu = mu_0, conf.level = 0.95)` dùng để:

- A. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu \neq 2000$ và độ tin cậy $1 - \alpha = 95\%$.
- B. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu < 2000$ và độ tin cậy $1 - \alpha = 95\%$.
- C. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu \neq 2000$ và mức ý nghĩa $1 - \alpha = 0.05$.
- D. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu > 2000$ và độ tin cậy $1 - \alpha = 0.95$.

Câu 22. Kết quả của lệnh `test$statistic` bằng

- A. 0.9218. B. 1.979949. C. 19. D. 0.06238.

Câu 23. Để kiểm định trung bình của hai mẫu X, Y độc lập với đối thuyết $\mu_X > \mu_Y$ và độ tin cậy $1 - \alpha = 0.95$, hãy viết một đoạn code thực hiện điều đó

Xét dữ liệu trong file `house.price.csv` với các tên biến như đoạn lệnh bên dưới, hãy trả lời các câu hỏi 23 và 24. 24 & 25

```
dtf = read.csv('house.price.csv', header=TRUE)
Tax = dtf$taxes
Sales = dtf$sale.price
Tax2 = Tax[Tax > 8]
Sale2 = Sales[Sales > 35]
```

Câu 24. Cho kết quả của kiểm định sau

test1

Welch Two Sample t-test

data: Tax and sales

t = -22.2571, df = 26.179, p-value = 1

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-30.36865 Inf

sample estimates:

mean of x mean of y

6.404917 34.612500

Hãy cho biết kết quả trên nói về kiểm định của
....., trong đó đối thuyết của kiểm định là, cùng với
mức ý nghĩa, ta có thể kết luận rằng

Câu 25. Cho kết quả của kiểm định sau

test2

2-sample test for equality of proportions with continuity correcti

data: y out of n

X-squared = 2.2222, df = 1, p-value = 0.06802

alternative hypothesis: less

95 percent confidence interval:

-1.00000000 0.01374729

sample estimates:

prop 1 prop 2

0.25 0.50

Hãy cho biết kết quả trên nói về kiểm định của
....., trong đó đối thuyết của kiểm định là, cùng với
 p -giá trị bằng, ta có thể kết luận rằng

Câu 26. Cho mẫu X , dùng những hàm có sẵn hãy viết các đoạn lệnh thực hiện *trung bình mẫu* :, *phương sai mẫu* :, *trung vị mẫu* và *độ lệch chuẩn (mẫu)*

Câu 27. Để tính p -giá trị của $Z_0 = 18$ với $Z_0 \sim \mathcal{B}(50, 0.5)$ với đối thuyết $H_1 : p \neq p_0$, hãy hoàn thành đoạn code sau

Z_0 =

p.value = 2*min(pnorm(...), 1 -(Z_0))

Xem đoạn lệnh sau và kết quả của nó để trả lời các câu hỏi ~~25~~ và 26. 28 và 29

```
df = read.csv('chloride.csv', header=TRUE)
y = df$y ## Nồng độ clorua
x = df$x ## diện tích ở đầu nguồn
lm(y ~ x)
```

```
Call:
lm(formula = y ~ x)
```

```
Coefficients:
(Intercept)          x
      0.4705      20.5673
```

Câu 28. Đoạn lệnh trên cho biết giá trị các hệ số ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ lần lượt bằng và

Câu 29. Đoạn lệnh trên cho biết

- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho y- nồng độ clorua(đv: mg/l) theo x- diện tích ở đầu nguồn x(đv:%) với các hệ số hồi quy $\hat{\beta}_0 = 20.5673$ và $\hat{\beta}_1 = 0.4705$.
- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho y- nồng độ clorua(đv: mg/l) theo x- diện tích ở đầu nguồn x(đv:%) với các hệ số hồi quy $\hat{\beta}_0 = 0.4705$ và $\hat{\beta}_1 = 20.5673$.
- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho x- diện tích ở đầu nguồn x(đv:%) theo y- nồng độ clorua(đv: mg/l) với các hệ số hồi quy $\hat{\beta}_0 = 20.5673$ và $\hat{\beta}_1 = 0.4705$.
- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho x- diện tích ở đầu nguồn x(đv:%) theo y- nồng độ clorua(đv: mg/l) với các hệ số hồi quy $\hat{\beta}_0 = 0.4705$ và $\hat{\beta}_1 = 20.5673$.

Câu 30. Để tìm tính khoảng tin cậy 99% cho β_0 , hãy hoàn chỉnh vào đoạn lệnh sau , biết rằng $\left[\hat{\beta}_0 - t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}}{Sxx} \right)}; \hat{\beta}_0 + t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}}{Sxx} \right)} \right]$ trong đó $\hat{\beta}_0$ là hệ số góc trong mô hình hồi quy và MSE, SSE, Sxx thoả các công thức sau

$$MSE = \frac{SSE}{n-2}, \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad Sxx = \sum_{i=1}^n (x_i - \bar{x})^2.$$

```
beta0_CI <- function(x, y, .....){
  n = length(x); x.mean = mean(x)
  result = lm(y~x)
  res = resid(result)
  beta0.hat = (coef(result))[[1]]
  MSE = .....
  Sxx = sum((x-x.mean)^...)
  eps = qt(1 - alpha/2, df=n-2)*sqrt(.....*(1/n + x.mean**2/.....))
  print('KTC cho beta 0')
  return(c(beta0.hat - eps, beta0.hat + eps))
}
beta0_CI(x, y, 0.01)
```

HẾT.



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM

ĐỀ THI KẾT THÚC HỌC PHẦN

Học kỳ I – Năm học: 2018-2019

MÃ LƯU TRỮ
(do Phòng KT-ĐBCL ghi)

CK18191.MTH00085

Tên học phần: THỰC HÀNH 'XÁC SUẤT THỐNG KÊ

Mã HP: MTH00085

Thời gian làm bài: 60 phút

Ngày thi: 05/01/2019

Ghi chú: Sinh viên không được sử dụng tài liệu. Đề thi gồm 8 trang.

Họ tên sinh viên: _____ MSSV: _____ STT: _____

Câu 1. Trong các đoạn code R sau, đoạn nào trả ra kết quả là TRUE .

- A. $(2 != 2) \&\& (3 <= 3)$. B. $(3 != 4) \&\& (4 >= 3)$.
 C. $(2 <= 5) \&\& (4 <= 3)$. D. $isTRUE(4 == 5)$.

Cho đoạn code sau, trả lời các câu hỏi 2-3.

```
perm <- function(n){
  if(n==1){
    return(matrix(1))
  } else {
    sp <- perm(n-1)
    p <- nrow(sp)
    A <- matrix(nrow=n*p,ncol=n)
    for(i in 1:n){
      A[(i-1)*p+1:p,] <- cbind(i,sp+(sp>=i))
    }
    return(A)
  }
}
Z=perm(3);
```

Câu 2. Cho biết số dòng của ma trận Z

- A. 3. B. 9. C. 8. D. 6.

Câu 3. Tìm số tự nhiên n sao cho ma trận $perm(n)$ có 24 dòng.

- A. 3. B. 4. C. 12. D. 24.

Câu 4. Câu lệnh R nào dùng để mô phỏng việc tung đồng xu cân đối đồng chất 5 lần.

- A. `sample(c('H','T'),5)`.
 B. `sample(c(0,1), 5, prob=c(1/2,1/2))`.
 C. `sample(c(0,1), 5, replace= T)`.
 D. `sample(5, c('H','T'), replace= T)`.

Cho đoạn code sau, trả lời các câu hỏi 5-6.

```
x <- c(rep(1, 3), rep(2, 4), rep(3,5), rep(4,4) ,rep(5,3))
a = mean(x)
b = length(x)
c = median(x)
```

Câu 5. Kết quả của $a * b$ bằng và $a * c$ bằngCâu 6. Giá trị của $a * b - c$ bằng

- A. 54. B. 24. C. 10. D. 60.

Câu 7. Câu lệnh R: `3*length(sample(1:6, 5))+10` cho ra kết quả nào sau đây?
 A. 25. B. 28. C. 43. D. 13.

Cho đoạn code sau, trả lời các câu hỏi 8-9.

```
df <- data.frame(
  STT = c(seq(1,6,1)),
  Name= c("Pearson", "Neymann", "Fisher", "Gosset", "Bayes", "Poisson"),
  Birthday = c(1857, 1894, 1890, 1876, 1702, 1781),
  Died=c(1936, 1981, 1962, 1937, 1761, 1840)
)
```

Câu 8. Dựa vào dataframe df, hãy cho biết năm sinh và năm mất của cha để kiểm định t (Gosset, William Sealy Gosset).

- A. 1890-1962. B. 1936-1937. C. 1876-1937. D. 1890-1962.

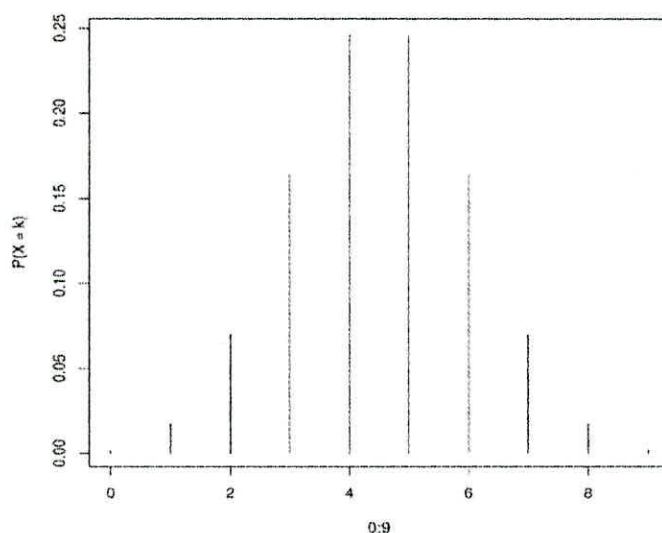
Câu 9. Lệnh `subset(Name, Birthday%%25>13)` cho ra tên của hai nhà thống kê đã có quan điểm đối lập nhau trong vấn đề kiểm định:

- A. [1] "Bayes" "Poisson". B. [1] "Pearson" "Neymann".
 C. [1] "Gosset" "Poisson". D. [1] "Neymann" "Fisher".

Câu 10. Giá trị của `ppois(x0, lambda)` bằng với

- A. Giá trị của hàm phân phối (tích lũy) của biến ngẫu nhiên phân phối Poisson $\mathcal{P}(\lambda)$ tại x_0 .
 B. $\sum_{k \in \mathbb{Z}: 0 \leq k \leq x_0} \frac{e^{-\lambda} \lambda^k}{k!}$.
 C. $\mathbb{P}(X \leq x_0)$ trong đó $X \sim \mathcal{P}(\lambda)$.
 D. Tất cả các giá trị liệt kê ở trên.

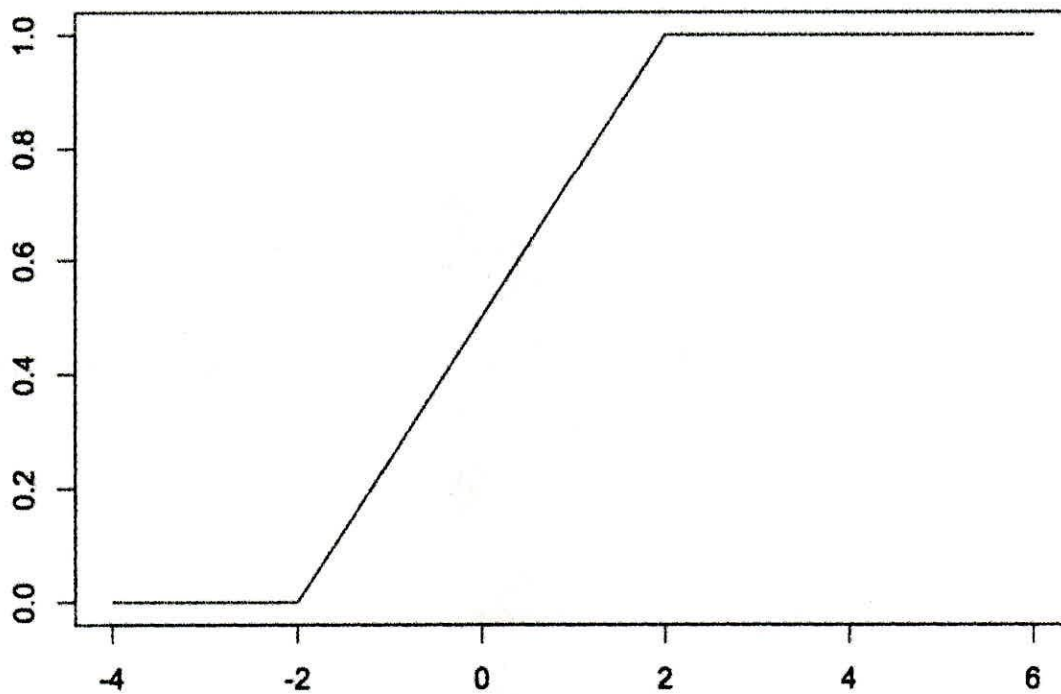
Câu 11. Trong các lệnh sau, lệnh nào có thể vẽ được hình bên dưới.



Hình 1: Biểu đồ cột

- A. `curve(dbinom(x, 9, 0.5), from = 0, to = 10)`.
 B. `curve(dnorm(x, 2, 1), from = -1, to = 5)`.
 C. `hist(c(0:9), dbinom(0:9, 9, 0.5))`.
 D. `plot(0:9, dbinom(0:9, 9, 0.5), type='h', ylab = "P(X = k)")`.

Câu 12. Cho biết đồ thị hàm phân phối (tích lũy), như Hình 2, là của phân phối nào



Hình 2: Hàm phân phối tích lũy

- A. PP chuẩn $\mathcal{N}(1, 1)$.
B. PP đều $U([-2, 2])$.
C. PP mũ $\text{Exp}(2)$.
D. PP Student(10).

Câu 13. Để phát sinh hàm mật độ của biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu = 2, \sigma^2 = 1)$ và mẫu 1000 phần tử là biến ngẫu nhiên có phân phối Poisson $\mathcal{P}(\lambda = 2)$ lần lượt bằng các lệnh và

Đề thi cuối kỳ môn THỰC HÀNH XÁC SUẤT THỐNG KÊ dạng trắc nghiệm có 50 câu hỏi, mỗi câu 5 đáp án. Sinh viên A không học bài, khi đi thi thì xác suất trả lời đúng của mỗi câu là như nhau. Sinh viên B học khá trong lớp, cảm thấy xác suất để mình chọn đúng mỗi câu là 0.6. Đặt

$$X = \sum_{i=1}^{50} X_i \text{ và } Y = \sum_{i=1}^{50} Y_i$$

trong đó, X_i, Y_i lần lượt là các biến ngẫu nhiên phản ánh kết quả chọn câu thứ i là đúng của sinh viên A và sinh viên B (với $i = 1, \dots, 50$), trả lời các câu hỏi 13 và 14: ~~va~~ và 15

Câu 14. X và Y có phân phối gì?

- A. Nhị thức $B(50, 0.2)$ và $B(50, 0.4)$. B. Nhị thức $B(50, 0.2)$ và $B(50, 0.6)$.
- C. Nhị thức $B(50, 0.8)$ và $B(50, 0.6)$. D. Nhị thức $B(50, 0.8)$ và $B(50, 0.4)$.

Câu 15. Sinh viên sẽ qua môn nếu trả lời đúng ít nhất 25 câu hỏi. Xác suất sinh viên A không qua môn; được xấp xỉ bằng câu lệnh nào sau đây,

- A. `pnorm(15/sqrt(8))` . B. `pnorm(15/8)` .
- C. `1-pnorm(15/sqrt(8))` . D. `dnorm(0.5)` .

Câu 16. Một nhóm sinh viên đo nhiệt độ ở những độ cao khác nhau và thu được bảng số liệu sau.

Elevation(ft)	600	1000	1250	1600	1800	2100	2500	2900
Temperature(F)	56	54	56	50	47	49	47	45

Sử dụng các câu lệnh R để vẽ đồ thị phân tán và đường hồi quy nhiệt độ theo độ cao cùng hệ trục tọa độ.

Câu 17. Các cuộc gọi điện đến tổng đài tuân theo phân phối Poisson với mức λ trên mỗi phút. Từ kinh nghiệm có được trong quá khứ, ta biết rằng xác suất nhận được chính xác một cuộc gọi trong một phút bằng ba lần xác suất không nhận được cuộc gọi nào trong cùng thời gian. Ta xét khoảng 100 khoảng thời gian một phút liên tiếp và gọi U là số khoảng thời gian một phút không nhận được cuộc gọi nào. Viết câu lệnh tính $\mathbb{P}(U \leq 1)$

```

Câu 18. path = 'D://Works'
setwd(path)
dtf = read.csv('data01.csv', header=TRUE)
Age = dtf$Age
U70 = Age[Age > 70]
KTC_prop <- function(data.p, data, alpha){
  phat = length(data.p)/ length(data) ## ty le mau
  eps = qnorm(1 - alpha/2)*sqrt(phat*(1-phat)/n)
  print('KTC cho ty le la')
  return(c(phat - eps, phat + eps))
}
KTC_prop(U70, Age, 0.05)

```

Hàm KTC_prop cho biết

- A. Input các tham số dữ liệu mẫu (data), dữ liệu thỏa tính chất nào đó để truy xuất tỷ lệ mẫu (datap) và mức ý nghĩa (alpha) .
- B. Output là khoảng tin cậy cho tỷ lệ p với độ tin cậy α ..
- C. A, B đều đúng .
- D. A, B đều sai.

Câu 19. Cho

$$z_{1-\alpha/2} \triangleq \text{qnorm}(1 - \alpha/2) \text{ và } t_{1-\alpha/2}^{n-1} \triangleq \text{qt}(1 - \alpha/2, \text{df} = n-1)$$

hoàn thành các chỗ trống trong đoạn code sau

```

path = 'D://Works'
setwd(path)
dtf = read.csv('data01.csv', header=TRUE)
Age = dtf$Age
KTC_mean <- function(data, alpha, sig = 'None'){
  n = length(data)
  m = mean(data)
  sd = sd(data)
  zalp = qnorm(1 - ...../2)
  talp = qt(1 - alpha/2, ..... )
  if(sig != 'None')
    eps = sig*...../sqrt(n)
  else if(sig == 'None')
    if( n < 30)
      eps = sd*talp/sqrt(n)
    else if (n >= ..... )
      eps = sd*...../sqrt(n)
  return(c(m - eps, m + eps))
}
KTC_mean(Age, 0.05)

```

Câu 20. Hàm KTC_mean cho biết

- A. Input các tham số dữ liệu mẫu (data), độ tin cậy (alpha) và giả thiết về sigma .
- B. Output là khoảng tin cậy của trung bình trong các trường hợp biết phương sai, không biết phương sai và cỡ mẫu.

- C. A, B đều sai .
D. A, B đều đúng.

Câu 21. Để kiểm định trung bình của hai mẫu X,Y độc lập với đối thuyết $\mu_X > \mu_Y$ và độ tin cậy $1-\alpha = 0.95$, hãy viết một đoạn code thực hiện điều đó

Xem đoạn code và kết quả sau

```
path = 'D://Works'
setwd(path)
data = read.csv('rocket.motor.csv', header=TRUE)
SK = data$streng; mu_0 = 2000
test = t.test(SK, alternative = "two.sided", mu = mu_0, conf.level = 0.95)
```

One Sample t-test

```
data: SK
t = 1.9799, df = 19, p-value = 0.06238
alternative hypothesis: true mean is not equal to 2000
95 percent confidence interval:
 1992.438 2272.377
sample estimates:
mean of x
 2132.407
```

Câu 22. Hàm `t.test(Age, alternative = "two.sided", mu = mu_0, conf.level = 0.95)` dùng để :

- A. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu \neq 2000$ và độ tin cậy $1-\alpha = 95\%$.
B. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu < 2000$ và độ tin cậy $1-\alpha = 95\%$.
C. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu \neq 2000$ và mức ý nghĩa $1-\alpha = 0.05$.
D. Khoảng tin cậy cho trung bình của mẫu với đối thuyết $\mu > 2000$ và độ tin cậy $1-\alpha = 0.95$.

Câu 23. Kết quả của lệnh `test$statistic` bằng

- A. 0.9218 . B. 1.979949. C. 19 . D. 0.06238 .

Xét dữ liệu trong file `house.price.csv` với các tên biến như đoạn lệnh bên dưới, hãy trả lời các câu hỏi 21 và 22. 24 và 25 .

```
dtf = read.csv('house.price.csv', header=TRUE)
Tax = dtf$taxes
Sales = dtf$sale.price
Tax2 = Tax[Tax > 8]
Sale2 = Sales[Sales > 35]
```

Câu 24. Cho kết quả của kiểm định sau

test1

Welch Two Sample t-test

```
data: Tax and sales
t = -22.2571, df = 26.179, p-value = 1
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -30.36865      Inf
sample estimates:
mean of x mean of y
 6.404917 34.612500
```

Hãy cho biết kết quả trên nói về kiểm định của
....., trong đó đối thuyết của kiểm định là, cùng với
mức ý nghĩa, ta có thể kết luận rằng

Câu 25. Cho kết quả của kiểm định sau

```
test2
      2-sample test for equality of proportions with continuity correcti
```

```
data: y out of n
X-squared = 2.2222, df = 1, p-value = 0.06802
alternative hypothesis: less
95 percent confidence interval:
 -1.000000000  0.01374729
sample estimates:
prop 1 prop 2
 0.25   0.50
```

Hãy cho biết kết quả trên nói về kiểm định của
....., trong đó đối thuyết của kiểm định là, cùng với
 p -giá trị bằng, ta có thể kết luận rằng

Câu 26. Cho mẫu X , dùng những hàm có sẵn hãy viết các đoạn lệnh thực hiện *trung bình mẫu* :, *phương sai mẫu* :, *trung vị mẫu* và *độ lệch chuẩn (mẫu)*

Câu 27. Để tính p -giá trị của $Z_0 = 18$ với $Z_0 \sim \mathcal{B}(50, 0.5)$ với đối thuyết $H_1 : p \neq p_0$, hãy hoàn thành đoạn code sau

```
Z_0 = .....
p.value = 2*min(pnorm(...), 1 - .....(Z_0))
```

Xem đoạn lệnh sau và kết quả của nó để trả lời các câu hỏi ~~25~~ và ~~26~~ 28 và 29

```
df = read.csv('chloride.csv', header=TRUE)
y = df$y ## Nồng độ cloride
x = df$x ## tỷ lệ phân tram
lm(y ~ x)
```

Call:

```
lm(formula = y ~ x)
```

Coefficients:

```
(Intercept)          x
      0.4705      20.5673
```

Câu 28. Đoạn lệnh trên cho biết giá trị các hệ số ước lượng $\hat{\beta}_0$ và $\hat{\beta}_1$ lần lượt bằng và

Câu 29. Đoạn lệnh trên cho biết

- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho y- nồng độ clorua(đv: mg/l) theo x- diện tích ở đầu nguồn x(đv:%) với các hệ số hồi quy $\hat{\beta}_0 = 20.5673$ và $\hat{\beta}_1 = 0.4705$.
- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho y- nồng độ clorua(đv: mg/l) theo x- diện tích ở đầu nguồn x(đv:%) với các hệ số hồi quy $\hat{\beta}_0 = 0.4705$ và $\hat{\beta}_1 = 20.5673$.
- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho x- diện tích ở đầu nguồn x(đv:%) theo y- nồng độ clorua(đv: mg/l) với các hệ số hồi quy $\hat{\beta}_0 = 20.5673$ và $\hat{\beta}_1 = 0.4705$.
- Kết quả mô hình hồi quy $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, cho x- diện tích ở đầu nguồn x(đv:%) theo y- nồng độ clorua(đv: mg/l) với các hệ số hồi quy $\hat{\beta}_0 = 0.4705$ và $\hat{\beta}_1 = 20.5673$.

Câu 30. Để tìm tính khoảng tin cậy 99% cho β_0 , hãy hoàn chỉnh vào đoạn lệnh sau , biết rằng $\left[\hat{\beta}_0 - t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}}{Sxx} \right)}; \hat{\beta}_0 + t_{1-\alpha/2}^{n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{\bar{x}}{Sxx} \right)} \right]$ trong đó $\hat{\beta}_0$ là hệ số góc trong mô hình hồi quy và MSE, SSE, Sxx thoả các công thức sau

$$MSE = \frac{SSE}{n-2}, \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad Sxx = \sum_{i=1}^n (x_i - \bar{x})^2.$$

```
beta0_CI <- function(x, y, .....){
  n = length(x); x.mean = mean(x)
  result = lm(y~x)
  res = resid(result)
  beta0.hat = (coef(result))[[1]]
  MSE = .....
  Sxx = sum((x-x.mean)^...)
  eps = qt(1 - alpha/2, df=n-2)*sqrt(.....*(1/n + x.mean**2/.....))
  print('KTC cho beta 0')
  return(c(beta0.hat - eps, beta0.hat + eps))
}
beta0_CI(x, y, 0.01)
```

HẾT.

Tên học phần: Giải tích hàm trong thông kê Mã HP: MTH10518
 Thời gian làm bài: 90 phút Ngày thi: 09/01/2019
 Ghi chú: Sinh viên [☐ được phép / ☒ không được phép] sử dụng tài liệu khi làm bài.

Câu 1 (.2...điểm). Cho không gian xác suất $\Omega = [0,1]$ với độ đo Lebesgue thông thường. Cho biến ngẫu nhiên $X(\omega) = 1$ với $\omega \in [0,1/3]$ và $X(\omega) = 0$ khi $\omega \in (1/3,1]$. Cho $Y : \Omega \rightarrow \mathbb{R}$ là biến ngẫu nhiên với $Y(\omega) = \omega^2$. Tìm $E(Y | X)$.

Câu 2 (.2...điểm). Cho $(X_n)_{n \geq 1}$ độc lập với $E(|X_n|) < \infty$ và $E(X_n) = 0 \forall n$. Với $n \geq 1$ đặt $F_0 = \{\emptyset, \Omega\}$ và $F_n = \sigma\{X_k; k \leq n\}$, $S_n = \sum_{k=1}^n X_k$. CM (S_n) là một martingale.

Câu 3 (.2...điểm). Cho biến ngẫu nhiên độc lập $X_1, X_2, \dots, X_n$ i.i.d. có hàm mật độ $f(x)$. Đặt

$$f_n(x) = \frac{1}{n} \sum_{k=1}^n K\left(\frac{x - X_k}{h}\right) \text{ với } n \in \mathbb{N}, h > 0 \text{ và } K : \mathbb{R} \rightarrow \mathbb{R}, K \geq 0 \text{ đo được. Tính}$$

$E f_n, \text{Var} f_n$ theo f, n, h . Từ đó suy ra biểu thức của $E |f_n - f|^2$. Nếu K liên tục và

$$\int K(t) dt = 1 \text{ thì } f_n \xrightarrow{L^2} f?$$

Câu 4 (2...điểm). Cho $\mu_n \rightarrow \mu, \sigma_n \rightarrow \sigma$ khi $n \rightarrow \infty$. Giả sử X_n có phân phối $N(\mu_n, \sigma_n)$. Chứng minh $X_n \rightarrow N(\mu, \sigma)$.

Câu 5 (.2...điểm). Cho biến ngẫu nhiên X có hàm mật độ $f_X(x) = \frac{1}{2} e^{-|x|}$. Tìm $\varphi_X(u)$.

Hết