

TRƯỜNG ĐẠI HỌC VĂN LANG

Đơn vị: Khoa Khoa học cơ bản

ĐỀ THI VÀ ĐÁP ÁN
THI KẾT THÚC HỌC PHẦN
Học kỳ 2, năm học 2024-2025

I. Thông tin chung

Tên học phần:	Xác suất thống kê Y học		
Mã học phần:	71MATM10033	Số tin chỉ:	3
Mã nhóm lớp học phần:	242_71MATM10033_01,02,03		
Hình thức thi: Trắc nghiệm kết hợp Tự luận	Thời gian làm bài:	75	phút
<i>Thí sinh được tham khảo tài liệu:</i>	☒ Có <i>(Chỉ được tham khảo tài liệu giấy)</i>		☐ Không

Cách thức nộp bài phần tự luận:

Nộp bài làm riêng cho **từng câu hỏi** dưới 1 trong 2 dạng sau đây (Không cần chép lại đề bài)

1/ Gõ trực tiếp vào chỗ dành sẵn trên trang thi CTE.

2/ Upload hình ảnh bài làm.

Lưu ý: - Các đáp án làm tròn 4 chữ số thập phân sau dấu phẩy.

- Sinh viên chỉ được tham khảo tài liệu **GIẤY**.

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

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	Áp dụng kiến thức cơ bản về xác suất trong việc chẩn đoán và quyết định lâm sàng	Thi cuối kỳ	30%	1-2 TN, 3TL	3	PI1.1
CLO2	Vận dụng kiến thức về thống kê trong quá trình thu thập và phân tích dữ liệu y khoa	Thi cuối kỳ	40%	3-10 TN	4	PI1.1
CLO3	Sử dụng phần mềm thống kê hỗ trợ cho	Thi cuối kỳ	30%	1-2 TL	3	PI4.2

	việc phân tích dữ liệu y khoa					
--	-------------------------------	--	--	--	--	--

III. Nội dung câu hỏi thi

PHẦN TRẮC NGHIỆM (10 câu hỏi, mỗi câu 0,5 điểm, tổng số điểm: 5 điểm)

Tỉ lệ mắc bệnh B trong cộng đồng là 20%. Hỏi phải quan sát tối thiểu bao nhiêu người để xác suất gặp được ít nhất một người mắc bệnh B không nhỏ hơn 99%?

A. 21

B. 23

C. 25

D. 19

ANSWER: A

Do nồng độ Na^+ trong một mẫu huyết thanh bằng quang kế ngọn lửa, lặp đi lặp lại nhiều lần phép đo, ta ghi nhận có 20% kết quả trên 143 mEq/lít và 30% kết quả dưới 141 mEq/lít. Tính nồng độ Na^+ trung bình (μ) và độ lệch tiêu chuẩn (σ) của phép đo trên. Giả sử nồng độ Na^+ tuân theo luật phân phối chuẩn.

A. $\mu = 141,76$; $\sigma = 1,47$

B. $\mu = 142,16$; $\sigma = 1,25$

C. $\mu = 142,56$; $\sigma = 1,57$

D. $\mu = 141,16$; $\sigma = 1,15$

ANSWER: A

Theo dõi nhịp tim (đơn vị tính: lần/phút) của người trưởng thành, dữ liệu về nhịp tim của 75 nam được ghi nhận như sau:

68	68	69	69	70	71	71	71	71	71
71	72	72	72	72	73	73	73	74	74
74	74	75	75	75	75	75	76	76	76
76	76	78	78	78	78	79	79	79	79
80	80	80	81	81	82	82	82	83	83
83	83	85	85	86	87	87	87	88	88
89	90	92	93	93	94	96	98	99	100
102	104	104	106	111					

Tính nhịp tim trung bình của 75 nam trong mẫu trên.

- A. 81,3333
- B. 79
- C. 71
- D. 83,3333

ANSWER: A

Khảo sát trọng lượng (đơn vị tính: kg) của 20 trẻ sơ sinh, ta có số liệu cho bởi bảng sau:

2,5	2,3	3,2	3	1,6	1,8	2,6	2,5	3,6	1,9
2,8	1,9	3,1	2,9	1,6	1,9	2,8	2,9	3,1	3,6

Tính phương sai (hiệu chỉnh) về trọng lượng của 20 trẻ sơ sinh được khảo sát trong mẫu trên.

- A. 0,3943
- B. 0,6279
- C. 0,3746
- D. 0,6120

ANSWER: A

Khảo sát hàm lượng Acid Uric (đơn vị tính: mg/dL) của một số người, ta có số liệu cho bởi bảng sau:

5,8	6,1	4,9	7,3	6,5	5,2	6,9	5,7
6,3	7	4,8	6,6	5,9	6,2	5	

Tìm khoảng tin cậy 95% cho hàm lượng Acid Uric trung bình trong dân số.

- A. [5,5774; 6,4493]
- B. [5,5801; 6,4466]
- C. [5,5072; 6,5195]
- D. [5,6149; 6,4117]

ANSWER: A

Điều tra một số gia đình 8 con, người ta thu được số liệu như sau:

<i>Số con trai</i>	0	1	2	3	4	5	6	7	8
<i>Số gia đình</i>	215	1485	5331	10649	14959	11929	6678	2092	342

Tìm khoảng tin cậy 99% cho tỉ lệ hộ gia đình có số con trai nhiều hơn số con gái trong gia đình 8 con trong dân số.

- A. [0,3867; 0,3974]
- B. [0,3879; 0,3961]
- C. [0,6654; 0,6758]
- D. [0,6666; 0,6746]

ANSWER: A

Điều tra một số gia đình 8 con, người ta thu được số liệu như sau:

<i>Số con trai</i>	0	1	2	3	4	5	6	7	8
<i>Số gia đình</i>	215	1485	5331	10649	14959	11929	6678	2092	342

Nếu muốn ước lượng số con trai trung bình trong gia đình 8 con trong dân số với độ chính xác không vượt quá 0,02 thì cần điều tra thêm tối thiểu bao nhiêu gia đình 8 con nữa? Chọn độ tin cậy 95%.

- A. Không cần điều tra thêm
- B. 19857
- C. 53680
- D. 33823

ANSWER: A

Điều tra năm đầu thấy tỉ lệ trẻ sâu răng bằng 48,53%. Sau 9 năm điều trị và súc họng bằng thuốc, điều tra lại 1250 trẻ thấy tỉ lệ trẻ sâu răng bằng 14,48%. Hỏi điều trị và súc họng bằng thuốc có hiệu quả không? Yêu cầu tính giá trị kiểm định và cho kết luận với mức ý nghĩa 1%.

- A. Giá trị kiểm định: -24,0874. Điều trị và súc họng bằng thuốc có hiệu quả.
- B. Giá trị kiểm định: -24,0874. Điều trị và súc họng bằng thuốc không có hiệu quả.
- C. Giá trị kiểm định: -34,2101. Điều trị và súc họng bằng thuốc có hiệu quả.
- D. Giá trị kiểm định: -34,2101. Điều trị và súc họng bằng thuốc không có hiệu quả.

ANSWER: A

Khoảng tin cậy 95% của chiều cao trung bình (đơn vị tính: cm) của 2790 thanh niên nam 17 tuổi là [162,5389; 162,9211]. Khoảng tin cậy 95% của chiều cao trung bình (đơn vị tính: cm) của 2310 thanh niên nam 18 tuổi là [163,2563; 163,6437].

Có ý kiến cho rằng chiều cao trung bình của thanh niên nam 17 tuổi nhỏ hơn chiều cao trung bình của thanh niên nam 18 tuổi. Với mức ý nghĩa 1%, hãy tính P-giá trị và cho kết luận về ý kiến trên.

- A. P-giá trị: 10^{-7} . Chấp nhận ý kiến.

- B. P-giá trị: 10^{-7} . Bác bỏ ý kiến.
 C. P-giá trị: 10^{-17} . Chấp nhận ý kiến.
 D. P-giá trị: 10^{-17} . Bác bỏ ý kiến.

ANSWER: A

Theo dõi cân thấp và lần sinh, người ta thu được số liệu như sau:

	Lần sinh		
	Ít lần	Nhiều lần	Rất nhiều lần
Số trẻ cân thấp	411	42	21
Tổng số trẻ	2205	264	67

Có ý kiến cho rằng cân thấp phụ thuộc vào lần sinh. Hãy tính giá trị kiểm định của phép kiểm định Khi - bình phương và cho nhận xét về ý kiến trên với mức ý nghĩa 5%.

- A. Giá trị kiểm định: 8,4056. Chấp nhận ý kiến.
 B. Giá trị kiểm định: 8,4056. Bác bỏ ý kiến.
 C. Giá trị kiểm định: 5,3146. Chấp nhận ý kiến.
 D. Giá trị kiểm định: 5,3146. Bác bỏ ý kiến.

ANSWER: A

PHẦN TỰ LUẬN (3 câu hỏi, tổng số điểm: 5 điểm)

Yêu cầu: Khuyến khích sinh viên nộp bài trực tiếp tại chỗ dành sẵn trên trang thi CTE.

Câu hỏi 1: (1,5 điểm)

Yêu cầu: Đối với bài toán kiểm định so sánh nhiều giá trị trung bình, yêu cầu phát biểu rõ giả thuyết, đối thuyết và dựa vào kết quả phân tích phương sai từ phần mềm Jasp (bảng ANOVA cho sẵn trong phần đề bài), hãy nêu rõ kết luận. Khuyến khích sinh viên nộp bài trực tiếp tại chỗ dành sẵn trên trang thi CTE.

Đề bài: Một nghiên cứu nhằm đánh giá xem việc điều trị bằng 4 loại thuốc khác nhau có ảnh hưởng đến mức đường huyết (đơn vị tính: mg/dL) của bệnh nhân hay không. Nhà nghiên cứu đã chọn ngẫu nhiên các bệnh nhân và chia thành 4 nhóm, mỗi nhóm điều trị bằng một loại thuốc A, B, C, D khác nhau. Sau một tháng, nhà nghiên cứu thu thập dữ liệu về mức đường huyết của từng bệnh nhân của mỗi nhóm như trong bảng sau đây:

A	85	90	80	92	89	87	91	86
B	80	74	85	84	81	79	82	83
C	95	96	94	99	97	93	98	100
D	88	87	86	89	87	84	83	90

Biết kết quả phân tích phương sai từ phần mềm Jasp như sau:

ANOVA

ANOVA - Muc duong huyet

Cases	Sum of Squares	df	Mean Square	F	p
Loai thuoc	984.375	3	328.125	33.840	< .001
Residuals	271.500	28	9.696		

Note. Type III Sum of Squares

Post Hoc Tests

Standard (HSD)

Post Hoc Comparisons - Loai thuoc

		Mean Difference	SE	df	t	p _{tukey}
A	B	6.500	1.557	28	4.175	0.001
	C	-9.000	1.557	28	-5.781	< .001
	D	0.750	1.557	28	0.482	0.963
B	C	-15.500	1.557	28	-9.955	< .001
	D	-5.750	1.557	28	-3.693	0.005
C	D	9.750	1.557	28	6.262	< .001

Note. P-value adjusted for comparing a family of 4 estimates.

Với mức ý nghĩa 5%:

a) Hỏi yếu tố thuốc có ảnh hưởng đến mức đường huyết của bệnh nhân không?

b) Nếu kết luận ở câu a là khẳng định thì hãy so sánh từng cặp nhằm chỉ ra các loại thuốc có mức đường huyết trung bình thấp nhất, cao nhất trong 4 loại thuốc A, B, C, D.

Câu hỏi 2: (1,5 điểm)

Yêu cầu: Khuyến khích sinh viên nộp bài trực tiếp tại chỗ dành sẵn trên trang thi CTE.

Nghiên cứu mối liên hệ giữa hàm lượng Cholesterol trong cơ thể, độ tuổi và chỉ số khối cơ thể (BMI) trên một số người, ta có số liệu như sau:

Mã Id	1	2	3	4	5	6	7	8	9
Độ tuổi	46	20	52	30	57	25	28	36	22
BMI	25,4	20,6	26,2	22,6	25,4	23,1	22,7	24,9	19,8
Cholesterol (mmol/L)	3,5	1,9	4,0	2,6	4,5	3,0	2,9	3,8	2,1

Mã Id (tt)	10	11	12	13	14	15	16	17	18
Độ tuổi	43	57	33	22	63	40	48	28	49
BMI	25,3	23,2	21,8	20,9	26,7	26,4	21,2	21,2	22,8
Cholesterol (mmol/L)	3,8	4,1	3,0	2,5	4,6	3,2	4,2	2,3	4,0

Biết kết quả phân tích hồi quy tuyến tính từ phần mềm Excel như sau:

SUMMARY OUTPUT

<i>Regression Statistics</i>						
Multiple R	0.9389					
R Square	0.8815					
Adjusted R Square	0.8657					
Standard Error	0.3074					
Observations	18					
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	2	10.5423	5.2711	55.7702	1E-07	
Residual	15	1.4177	0.0945			
Total	17	11.96				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0.4555	0.9182	0.4960	0.6271	-1.5017	2.4126
Age	0.0541	0.0076	7.1204	4E-06	0.0379	0.0702
Bmi	0.0334	0.0469	0.7119	0.4875	-0.0665	0.1333

a) Tìm hệ số tương quan tuyến tính toàn phần giữa hàm lượng cholesterol, độ tuổi và chỉ số khối cơ thể. Tìm hệ số xác định bội và nêu ý nghĩa của hệ số xác định bội vừa tìm được.

b) Tìm phương trình hồi quy ước lượng thể hiện mối liên hệ tuyến tính của hàm lượng cholesterol theo độ tuổi và chỉ số khối cơ thể. Dự đoán hàm lượng cholesterol trung bình của một người 60 tuổi và có chỉ số khối cơ thể là 30 (ước lượng điểm).

c) Hãy kiểm định ý nghĩa của mô hình hồi quy vừa tìm được ở câu b với mức ý nghĩa 5%.

Câu hỏi 3: (2 điểm)

Một xét nghiệm dùng chẩn đoán bệnh. Xét nghiệm có độ nhạy bằng 0,9, độ chuyên bằng 0,7 và xác suất dương tính bằng 0,5625. Chẩn đoán bệnh cho 500 người. Tìm tỉ lệ bệnh lưu hành trong dân số.

HẾT

ĐÁP ÁN PHẦN TỰ LUẬN VÀ THANG ĐIỂM

Đáp án câu hỏi 1: (1,5 điểm)

a) Kiểm định sự ảnh hưởng của yếu tố thuốc đến mức đường huyết của bệnh nhân:

Giả thuyết H_0 : Yếu tố thuốc không ảnh hưởng đến mức đường huyết của bệnh nhân

Đối thuyết H_1 : Yếu tố thuốc có ảnh hưởng đến mức đường huyết của bệnh nhân **(0,25 điểm)**

Giá trị P-value: $P\text{-value} < 0,001 < 5\%$ **(0,25 điểm)**

Do đó ta bác bỏ giả thuyết H_0 , chấp nhận đối thuyết H_1 . **(0,25 điểm)**

Kết luận: Yếu tố thuốc có ảnh hưởng đến mức đường huyết của bệnh nhân. **(0,25 điểm)**

b) Dựa vào kết quả phân tích sâu ANOVA để so sánh từng cặp:

Ta thấy chỉ có p_{tukey} của cặp A-D là $0,963 > 5\%$, còn các cặp khác đều có $p_{\text{tukey}} < 5\%$.

Do đó: A khác B, C có ý nghĩa thống kê, B khác C, D và C khác D có ý nghĩa thống kê.

(0,25 điểm)

Dựa vào giá trị “Mean Difference” ta có thể kết luận: $\mu_B < \mu_A \approx \mu_D < \mu_C$

Kết luận: Loại thuốc có mức đường huyết trung bình thấp nhất là B, loại thuốc có mức đường huyết trung bình cao nhất là D.

(0,25 điểm)

Đáp án câu hỏi 2: (1,5 điểm)

a) Hệ số tương quan tuyến tính toàn phần giữa hàm lượng cholesterol, độ tuổi và chỉ số khối cơ thể: $R = 0,9389$

(0,25 điểm)

Hệ số xác định bội: $R^2 = 0,8815 = 88,15\%$

Ý nghĩa của hệ số xác định bội: 88,15% giá trị biến thiên của hàm lượng cholesterol được xác định bởi tác động tuyến tính của độ tuổi và chỉ số khối cơ thể

(0,25 điểm)

b) Mô hình hồi quy: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Trong đó: Y là hàm lượng cholesterol; X_1 là độ tuổi, X_2 là chỉ số khối cơ thể

Phương trình hồi quy tuyến tính ước lượng của Y theo X_1, X_2 :

$$\overline{y_{x_1, x_2}} = 0,4555 + 0,0541x_1 + 0,0334x_2$$

(0,25 điểm)

Ước lượng giá trị trung bình của Y khi $X_1 = 60$, $X_2 = 30$:

$$\overline{y}_{x_1=60, x_2=30} = 0,4555 + 0,0541 \times 60 + 0,0334 \times 30 = 4,7035 \text{ (mmol/L)} \quad (0,25 \text{ điểm})$$

c) Mô hình hồi quy: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Phương trình hồi quy tuyến tính ước lượng của Y theo X_1, X_2 :

$$\overline{y}_{x_1, x_2} = 0,4555 + 0,0541x_1 + 0,0334x_2$$

Giả thuyết H_0 : Mô hình hồi quy không có ý nghĩa thống kê (0,25 điểm)

Đối thuyết H_1 : Mô hình hồi quy có ý nghĩa thống kê

Do $P\text{-value} = 10^{-7} < 5\%$ nên ta bác bỏ giả thuyết H_0 và chấp nhận đối thuyết H_1 .

Nghĩa là mô hình hồi quy vừa tìm được ở câu b có ý nghĩa thống kê. (0,25 điểm)

Đáp án câu hỏi 3: (2,0 điểm)

Giả sử tỉ lệ bệnh lưu hành trong dân số là: $P(B^+) = p \in (0, 1)$

Độ nhạy: $P(T^+ | B^+) = 0,9$, độ chuyên: $P(T^- | B^-) = 0,7$ (0,25 điểm)

Xét bảng 2×2 đối với xét nghiệm T : (0,75 điểm)

	B^+	B^-	
T^+	$450p$	$150 - 150p$	$150 + 300p$
T^-	$50p$	$350 - 350p$	$350 - 300p$
	$500p$	$500 - 500p$	$N = 500$

Xác suất dương tính bằng 0,5625:

$$P(T^+) = P(B^+) \cdot P(T^+ | B^+) + P(B^-) \cdot P(T^+ | B^-) = 0,5625 \quad (0,5 \text{ điểm})$$

$$\Leftrightarrow p \times 0,9 + (1 - p) \times (1 - 0,7) = 0,5625 \Leftrightarrow 0,9p + (1 - p) \times 0,3 = 0,5625$$

$$\Leftrightarrow 0,6p + 0,3 = 0,5625 \Leftrightarrow 0,6p = 0,2625 \Leftrightarrow p = 0,4375 \quad (0,25 \text{ điểm})$$

Kết luận: Tỉ lệ bệnh lưu hành trong dân số là 43,75% (0,25 điểm)

Ghi chú: Sinh viên có thể dùng phương pháp công thức Bayes hoặc tỉ số khả năng.

----- Hết -----

TP. Hồ Chí Minh, ngày 16 tháng 04 năm 2025

Người duyệt đề

Giảng viên ra đề

Đinh Tiến Liêm


Đỗ Lưu Công Minh