

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 2023-2024

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:	232_71MATM10033_01,02		
Hình thức thi: Trắc nghiệm kết hợp Tự luận	Thời gian làm bài:	75	phút
<i>Thí sinh được tham khảo tài liệu:</i>	<input checked="" type="checkbox"/> Có	<input type="checkbox"/> 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/ Upload file bài làm (word, excel, pdf, ...);

2/ Upload hình ảnh bài làm.

Lưu ý: - Sinh viên có 10 phút nộp bài (sau khi hết thời gian làm bài).

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

- Sinh viên tự chuẩn bị bảng tra cứu các phân vị thường dùng.

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 việc phân tích dữ liệu y khoa	Thi cuối kỳ	30%	1-2 TL	3	PI4.2

III. Nội dung câu hỏi thi

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

Cho biết tỉ lệ bệnh B tại cộng đồng là 2%. Khám ngẫu nhiên 5 người. Tính xác suất gặp 1 người mắc bệnh B.

- A. 0,0922
- B. 0,9039
- C. 0,4096
- D. 0,0961

ANSWER: A

Theo Sleep Foundation, giấc ngủ đêm trung bình là 6,8 giờ và độ lệch chuẩn 0,6 giờ. Giả sử phân phối xác suất là phân phối chuẩn. Các bác sĩ đề nghị nên ngủ từ 7 đến 9 giờ mỗi đêm. Hỏi có bao nhiêu phần trăm dân số ngủ ít hơn mức này?

- A. 63,06%
- B. 49,99%
- C. 2,27%
- D. 36,93%

ANSWER: A

Theo Dubbo Osteoporosis Epidemiology Study, dữ liệu về mật độ xương ở cổ xương đùi của 24 người được ghi nhận như sau: (đơn vị tính: g/cm³)

0,37	0,64	0,78	0,75	0,51	0,65	0,96	0,83
0,82	0,69	0,58	0,63	0,83	0,75	0,77	0,71
1,17	1,19	1,14	1	1,1	0,71	0,92	0,86

Tính mật độ xương ở cổ xương đùi trung bình của 24 người trong mẫu trên.

- A. 0,8067 (g/cm³)
- B. 0,7750 (g/cm³)
- C. 0,8200 (g/cm³)
- D. 0,8783 (g/cm³)

ANSWER: A

Theo Dubbo Osteoporosis Epidemiology Study, dữ liệu về mật độ xương ở cổ xương đùi của 24 người được ghi nhận như sau: (đơn vị tính: g/cm³)

0,37	0,64	0,78	0,75	0,51	0,65	0,96	0,83
0,82	0,69	0,58	0,63	0,83	0,75	0,77	0,71
1,17	1,19	1,14	1	1,1	0,71	0,92	0,86

Tìm độ lệch chuẩn (hiệu chỉnh) của mật độ xương ở cổ xương đùi của 24 người trong mẫu trên.

- A. 0,2094 (g/cm³)
- B. 0,0427 (g/cm³)
- C. 0,0438 (g/cm³)
- D. 0,0181 (g/cm³)

ANSWER: A

Giả sử hàm lượng nước X (%) trong một loại cam là biến ngẫu nhiên có phân phối chuẩn: $X \sim N(\mu, \sigma^2)$. Quan sát một mẫu gồm 30 quả cam thuộc loại trên ta thu được số liệu sau:

X (%)	79	80	84,6	86	87,5	89	90
Số quả	2	4	6	8	5	4	1

Hãy ước lượng hàm lượng nước trung bình của loại cam trên với độ tin cậy 95%.

- A. [84,0867; 86,3867] (%)
- B. [83,8695; 86,6038] (%)
- C. [83,7228; 86,7505] (%)
- D. [84,2715; 86,2019] (%)

ANSWER: A

Giả sử hàm lượng nước X (%) trong một loại cam là biến ngẫu nhiên có phân phối chuẩn: $X \sim N(\mu, \sigma^2)$. Quan sát một mẫu gồm 30 quả cam thuộc loại trên ta thu được số liệu sau:

X (%)	79	80	84,6	86	87,5	89	90
Số quả	2	4	6	8	5	4	1

Tìm khoảng tin cậy 95% cho tỉ lệ cam có hàm lượng nước từ 86% trở lên.

- A. [42,47%; 77,53%]
- B. [45,29%; 74,71%]
- C. [39,16%; 80,84%]
- D. [36,92%; 83,08%]

ANSWER: A

Giả sử hàm lượng nước X (%) trong một loại cam là biến ngẫu nhiên có phân phối chuẩn: $X \sim N(\mu, \sigma^2)$. Quan sát một mẫu gồm 30 quả cam thuộc loại trên ta thu được số liệu sau:

X (%)	79	80	84,6	86	87,5	89	90
Số quả	2	4	6	8	5	4	1

Nếu muốn sai số của bài toán ước lượng hàm lượng cam trung bình không vượt quá 1% với độ tin cậy 99% thì cần quan sát thêm tối thiểu bao nhiêu quả cam?

- A. 39
- B. 27
- C. 10
- D. Không cần quan sát thêm vì cỡ mẫu $n = 30$ đã thỏa mãn yêu cầu bài toán

ANSWER: A

Theo dõi số chuột chết khi cho các lô chuột thí nghiệm sử dụng các liều thuốc có độc tăng dần, người ta thu được số liệu sau:

Liều thuốc (mg/kg)	0,015	0,02	0,025	0,03	0,035	0,04
Số chuột mỗi lô	10	10	10	10	10	10
Số chuột chết	0	2	5	8	9	10

Có ý kiến cho rằng tỉ lệ chuột chết (ở tất cả các liều thuốc) bằng 55%. Với mức ý nghĩa 5%, hãy tính giá trị kiểm định và cho kết luận về ý kiến trên.

- A. Giá trị kiểm định: 0,2600. Chấp nhận ý kiến.
- B. Giá trị kiểm định: 0,2600. Bác bỏ ý kiến.
- C. Giá trị kiểm định: 0,1965. Chấp nhận ý kiến.
- D. Giá trị kiểm định: 0,1965. Bác bỏ ý kiến.

ANSWER: A

Theo dõi số chuột chết khi cho các lô chuột thí nghiệm sử dụng các liều thuốc có độc tăng dần, người ta thu được số liệu sau:

Liều thuốc (mg/kg)	0,015	0,02	0,025	0,03	0,035	0,04
Số chuột mỗi lô	10	10	10	10	10	10
Số chuột chết	0	2	5	8	9	10

Có ý kiến cho rằng nên sử dụng liều thuốc độc trung bình tối đa là 0,03 (mg/kg). Yêu cầu tính giá trị kiểm định và cho nhận xét về ý kiến trên với mức ý nghĩa 5%.

- A. Giá trị kiểm định: $-2,2488$. Chấp nhận ý kiến.
- B. Giá trị kiểm định: $-2,2488$. Bác bỏ ý kiến.
- C. Giá trị kiểm định: $2,7800$. Chấp nhận ý kiến.
- D. Giá trị kiểm định: $2,7800$. Bác bỏ ý kiến.

ANSWER: A

Bốn loại thuốc chữa cùng một loại bệnh B, khảo sát trên bốn nhóm bệnh nhân (mỗi nhóm dùng một loại thuốc), thu được kết quả:

	Nhóm 1	Nhóm 2	Nhóm 3	Nhóm 4
Khỏi bệnh	123	95	152	132
Không khỏi bệnh	28	19	63	53

Hỏi hiệu quả của bốn loại thuốc có như nhau không? Yêu cầu tính giá trị kiểm định của phép kiểm định Khi - bình phương và cho kết luận với mức ý nghĩa 5%.

- A. Giá trị kiểm định: $11,0769$. Hiệu quả của bốn loại thuốc khác nhau.
- B. Giá trị kiểm định: $11,0769$. Hiệu quả của bốn loại thuốc như nhau.
- C. Giá trị kiểm định: $0,0113$. Hiệu quả của bốn loại thuốc khác nhau.
- D. Giá trị kiểm định: $0,0113$. Hiệu quả của bốn loại thuốc như nhau.

ANSWER: A

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

Câu 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 Excel (sinh viên tự thực hành và trình bày bảng ANOVA vào bài làm), hãy nêu rõ kết luận. Khuyến khích sinh viên nộp bài làm bằng tập tin Excel.

Đề bài: Một nghiên cứu về mối quan hệ giữa vị trí địa lý và chứng trầm cảm được các nhà xã hội học và bác sĩ tâm lý ở Trung tâm Y tế Goworth - ngoại ô New York thực hiện trên đối tượng người trên 65 tuổi. Một mẫu gồm 60 cá nhân, tất cả đều có sức khỏe khá tốt, được chọn: 20 người là cư dân của Florida, 20 người là cư dân của New York và 20 người là cư dân của Bắc Carolina. Mỗi cá nhân trong mẫu được chọn thực hiện một bài kiểm tra tiêu chuẩn để đo mức độ trầm cảm. Điểm kiểm tra cao hơn cho thấy mức độ trầm cảm cao hơn. Dữ liệu thu thập được trong Medical1.

Phần thứ hai của nghiên cứu xem xét mối quan hệ giữa vị trí địa lý và chứng trầm cảm đối với những người trên 65 tuổi mắc bệnh mãn tính như viêm khớp, tăng huyết áp, bệnh tim. Một mẫu gồm 60 cá nhân, tất cả đều có bệnh mãn tính, được chọn: 20 người là cư dân của

Florida, 20 người là cư dân của New York và 20 người là cư dân của Bắc Carolina. Mỗi cá nhân trong mẫu được chọn thực hiện cùng bài kiểm tra để đo mức độ trầm cảm với nhóm có sức khỏe tốt. Dữ liệu thu thập được trong Medical2.

Data from Medical1			Data from Medical2		
Florida	New York	North Carolina	Florida	New York	North Carolina
3	8	10	13	14	10
7	11	7	12	9	12
7	9	3	17	15	15
3	7	5	17	12	18
8	8	11	20	16	12
8	7	8	21	24	14
8	8	4	16	18	17
5	4	3	14	14	8
5	13	7	13	15	14
2	10	8	17	17	16
6	6	8	12	20	18
2	8	7	9	11	17
6	12	3	12	23	19
6	8	9	15	19	15
9	6	8	16	17	13
7	8	12	15	14	14
5	5	6	13	9	11
4	7	3	10	14	12
7	7	8	11	13	13
3	8	11	17	11	11

Hỏi yếu tố sức khỏe có ảnh hưởng đến chứng trầm cảm không? Kết luận với mức ý nghĩa 5%.

Câu 2: (1,5 điểm)

Yêu cầu: Khuyến khích sinh viên nộp bài làm bằng tập tin Excel.

Khảo sát nhiệt độ của một phản ứng hóa học (Y , đơn vị tính: $^{\circ}\text{C}$) cùng với nồng độ của bốn hóa chất khác nhau ($X_i, i = \overline{1, 4}$, đơn vị tính: mol/lít) ta có số liệu như sau:

x_1	x_2	x_3	x_4	y
7	26	6	60	78,5
1	29	15	52	74,5
11	56	8	20	104,3
11	31	8	47	87,6
7	52	6	33	95,9

11	55	9	22	109,2
3	71	17	6	102,7
1	31	22	44	72,5
2	54	18	22	93,1
21	47	4	26	115,9

a) Tìm phương trình hồi quy ước lượng thể hiện mối liên hệ tuyến tính của y theo $x_i, i = \overline{1, 4}$. Dự đoán nhiệt độ phản ứng hóa học trung bình khi nồng độ của bốn hóa chất lần lượt là $x_1 = 10, x_2 = 30, x_3 = 15, x_4 = 40$ (ước lượng điểm).

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

Câu 3: (2 điểm)

Một xét nghiệm có xác suất sai bằng 0,198. Tỷ lệ bị bệnh tại cộng đồng bằng 0,02. Biết độ chuyên của xét nghiệm bằng 0,8. Dùng xét nghiệm chẩn đoán bệnh. Tìm giá trị tiên đoán dương của xét nghiệm.

HẾT

ĐÁP ÁN PHẦN TỰ LUẬN VÀ THANG ĐIỂM

Câu hỏi 1: (1,5 điểm)

* Kết quả phân tích phương sai từ phần mềm Excel:

(0,5 điểm)

Anova: Two-Factor With Replication

SUMMARY	Florida	New York	North Carolina	Total
<i>Sức khỏe khá tốt</i>				
Count	20	20	20	60
Sum	111	160	141	412
Average	5.55	8	7.05	6.8667
Variance	4.5763	4.8421	8.05	6.6599
<i>Có bệnh mãn tính</i>				
Count	20	20	20	60
Sum	290	305	279	874
Average	14.5	15.25	13.95	14.5667
Variance	10.0526	17.0395	8.6816	11.8090
<i>Total</i>				

Count	40	40	40
Sum	401	465	420
Average	10.025	11.625	10.5
Variance	27.6660	24.1378	20.3590

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	1778.7	1	1778.7	200.4466	7.1E-27	3.9243
Columns	54.01667	2	27.0083	3.0436	0.0516	3.0759
Interaction	24.05	2	12.025	1.3551	0.2620	3.0759
Within	1011.6	114	8.8737			
Total	2868.3667	119				

* Kiểm định sự ảnh hưởng của yếu tố sức khỏe đến chứng trầm cảm:

Giả thuyết H_0 : Yếu tố sức khỏe không ảnh hưởng đến chứng trầm cảm

Đối thuyết H_1 : Yếu tố sức khỏe có ảnh hưởng đến chứng trầm cảm **(0,25 điểm)**

Giá trị của tiêu chuẩn kiểm định: $F - stat = 200,4466$

Phân vị Fisher: $F - crit = 3,9243$ **(0,25 điểm)**

Do $F - stat > F - crit$ nên ta bác bỏ giả thuyết H_0 , chấp nhận đối thuyết H_1 **(0,25 điểm)**

Kết luận: Yếu tố sức khỏe có ảnh hưởng đến chứng trầm cảm **(0,25 điểm)**

Ghi chú 1: Có thể dùng phương pháp P-value: $P\text{-value} = 7 \times 10^{-27} < 5\%$, ...

Ghi chú 2: Sinh viên có thể biến đổi bảng số liệu (tính điểm trầm cảm trung bình cho từng trường hợp: sức khỏe khá tốt, có bệnh mãn tính) để đưa về bài toán phân tích phương sai hai nhân tố không lặp

	Florida	New York	Bắc Carolina
Sức khỏe khá tốt	5,5	8	7,05
Có bệnh mãn tính	14,5	15,25	13,95

Câu hỏi 2: (1,5 điểm)

a) **Kết quả phân tích hồi quy tuyến tính từ phần mềm Excel:** **(0,5 điểm)**

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0.9915
R Square	0.9831
Adjusted R Square	0.9696
Standard Error	2.6108
Observations	10

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	1983.115	495.7787	72.7350	0.0001
Residual	5	34.08117	6.8162		
Total	9	2017.196			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	67.2870	89.3443	0.7531	0.4853	-162.3797	296.9537
X Variable 1	1.4192	0.9449	1.5020	0.1934	-1.0097	3.8480
X Variable 2	0.4993	0.9302	0.5367	0.6145	-1.8919	2.8904
X Variable 3	-0.0981	0.9716	-0.1010	0.9235	-2.5957	2.3995
X Variable 4	-0.1798	0.8971	-0.2004	0.8490	-2.4857	2.1262

Mô hình hồi quy: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$

Phương trình hồi quy tuyến tính ước lượng của y theo x_1, x_2, x_3 và x_4 :

$$\overline{y_{x_1, x_2, x_3, x_4}} = 67,2870 + 1,4192x_1 + 0,4993x_2 - 0,0981x_3 - 0,1798x_4 \quad (0,25 \text{ điểm})$$

Ước lượng giá trị trung bình của y khi $x_1 = 10, x_2 = 30, x_3 = 15, x_4 = 40$:

$$\overline{y_{x_1=10, x_2=30, x_3=15, x_4=40}} = 67,2870 + 1,4192 \times 10 + 0,4993 \times 30 - 0,0981 \times 15 - 0,1798 \times 40$$

$$\overline{y_{x_1=10, x_2=30, x_3=15, x_4=40}} = 87,7945 \text{ (}^\circ\text{C)} \quad (0,25 \text{ điểm})$$

b) Mô hình hồi quy: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$

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 $P\text{-value} = 0,0001 < 5\%$ 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 a có ý nghĩa thống kê. (0,25 điểm)

Câu hỏi 3: (2 điểm)

Tỉ lệ bệnh lưu hành trong dân số là $P(B^+) = 0,02$

Độ chuyên: $P(T^- | B^-) = 0,8$, giả sử độ nhạy: $P(T^+ | B^+) = p \in (0, 1)$ (0,25 điểm)

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

	B^+	B^-	
T^+	$20p$	196	$196 + 20p$
T^-	$20 - 20p$	784	$804 - 20p$

20	980	$N = 1000$
----	-----	------------

Xác suất sai của xét nghiệm là 0,198:

$$P(S) = P(B^+) \cdot P(T^- | B^+) + P(B^-) \cdot P(T^+ | B^-) = 0,198 \quad (0,25 \text{ điểm})$$

$$\Leftrightarrow 0,02 \times (1 - p) + (1 - 0,02) \times (1 - 0,8) = 0,198 \Leftrightarrow 0,018 - 0,02p = 0 \Leftrightarrow p = 0,9 \quad (0,25 \text{ điểm})$$

Giá trị tiên đoán dương của xét nghiệm T :

$$PV^+ = P(B^+ | T^+) = \frac{20p}{196 + 20p} = \frac{20 \times 0,9}{196 + 20 \times 0,9} = \frac{9}{107} \approx 0,0841 \quad (0,5 \text{ điểm})$$

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

HẾT

TP. Hồ Chí Minh, ngày 26 tháng 06 năm 2024

Người duyệt đề



Đinh Tiến Liêm

Giảng viên ra đề



Đỗ Lưu Công Minh