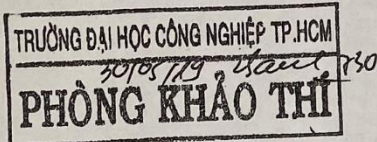


TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP  
THÀNH PHỐ HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN



ĐỀ THI CUỐI KỲ

Môn thi: Thống kê máy tính và Ứng dụng

Lớp/Lớp học phần: ĐHTH13

Ngày thi: 03/06/2019

Thời gian làm bài: 75 phút  
(Không kể thời gian phát đề)

Họ và tên thí sinh

; MSSV:

**Câu 1: (2 điểm)** Giả sử chiều cao của sinh viên năm hai ở một trường Đại Học là đại lượng ngẫu nhiên có phân phối chuẩn với trung bình 172cm, độ lệch chuẩn là 8cm. Hãy tính:

- Tỷ lệ sinh viên có chiều cao lớn hơn 184cm.
- Tỷ lệ sinh viên có chiều cao trong khoảng từ 164cm – 180cm.

**Câu 2: (5 điểm)** Khảo sát lượng sữa bò vắt được trong ngày (lít/ngày) ở một trang trại, người ta thu được bảng số liệu sau:

Lượng sữa (lít/ngày)	7	9	11	13	15
Số con bò	10	32	80	39	8

- Với độ tin cậy 95%, có thể nói lượng sữa trung bình vắt được trong ngày nằm trong khoảng nào?
- Trong câu (a), độ tin cậy được dùng để ước lượng là 95%. Để có kết quả ước lượng chính xác hơn, ta có nên tăng độ tin cậy lên không? Hãy nêu ý kiến của bạn. Nếu độ tin cậy là 100% thì điều gì sẽ xảy ra?
- Với mức ý nghĩa 5%, hãy kiểm định giả thuyết sau: “*Lượng sữa trung bình vắt được trong một ngày của trang trại là 11.3 (lít/ngày)*”.
- Giám đốc trang trại cho rằng: “*Trên 90% số bò của trang trại có năng suất sữa cao*”. Biết rằng bò có năng suất sữa cao là bò cho nhiều hơn 7 lít sữa mỗi ngày. Dựa vào bảng số liệu trên hãy nhận xét ý kiến trên với mức ý nghĩa 5%.

**Câu 3: (3.0 điểm)** Bergans là một công ty Na Uy chuyên sản xuất đồ phục vụ du lịch ngoài trời. Bảng dữ liệu dưới đây thể hiện mức độ chịu nhiệt (Temperature Rating - °F) và giá (Price - \$) của 11 loại túi ngủ do công ty Bergans sản xuất.

Model	Rating	Price
Ranger 3-Seasons	12	319
Ranger Spring	24	289
Ranger Winter	3 ✓	389 ✓
Rondane 3-Seasons	13	239
Rondane Summer	38	149
Rondane Winter	4 ✓	289
Senja Ice	5 ✓	359
Senja Snow	15	259
Senja Zero	25	229
Super Light	45	129
Tight & Light	25	199

- Vẽ đồ thị phân tán (scatter-plot) cho bảng dữ liệu này khi mức độ chịu nhiệt là biến độc lập (đồ thị thể hiện sự thay đổi về giá khi mức độ chịu nhiệt thay đổi).
- Tìm phương trình hồi quy ứng với dữ liệu mẫu đã cho.
- Nếu một túi ngủ có mức độ chịu nhiệt là 20 (°F), hãy ước tính giá của nó.

**Lưu ý:** tính toán các kết quả với độ chính xác là 0.001

Hết

Đề thi được sử dụng tài liệu giấy. Cán bộ coi thi không giải thích gì thêm.



TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP  
THÀNH PHỐ HỒ CHÍ MINH

**ĐÁP ÁN ĐỀ THI CUỐI KỲ**

KHOA CÔNG NGHỆ THÔNG TIN

Môn thi : Thống kê Máy tính và Ứng dụng

Lớp/Lớp học phần: ĐHTH13

Ngày thi: 03/06/2019

Thời gian làm bài: 75 phút

Câu	Nội dung trả lời	Điểm
<b>Câu 1</b>		<b>2.0 điểm</b>
a	<ul style="list-style-type: none"> <li>Gọi X là chiều cao sinh viên năm 2; <math>X \sim N(172; 8^2)</math></li> <li><math>P(X &gt; 180) = 0.0668 \approx 0.067</math></li> </ul>	1
b	<ul style="list-style-type: none"> <li><math>P(164 &lt; X &lt; 184) = 0.6827 \approx 0.683</math></li> </ul>	1
<b>Câu 2</b>		<b>5.0 điểm</b>
	Từ bảng số liệu ta có: $\bar{x} = 11.036$ ; $s = 1.845$ ; $n = 169$	1
a	<ul style="list-style-type: none"> <li>Từ bảng số liệu ta có: <math>\bar{x} = 11.036</math>; <math>s = 1.845</math>; <math>n = 169</math></li> <li><math>\varphi(C) = \frac{Y}{2} = 0.475 \Rightarrow C = t_{0.025}^{99} = 1.984</math></li> <li><math>\varepsilon = C * \frac{s}{\sqrt{n}} = 0.282</math></li> <li><math>\mu \in \{\bar{x} - \varepsilon; \bar{x} + \varepsilon\} \Leftrightarrow \mu \in \{10.754; 11.318\}</math></li> </ul>	1
b	<ul style="list-style-type: none"> <li>Độ tin cậy thể hiện độ chính xác của khoảng ước lượng nên tăng độ tin cậy có nghĩa là tăng độ chính xác</li> <li>Tuy nhiên, khi tăng độ tin cậy thì khoảng ước lượng cũng tăng lên dẫn đến khó tìm giá trị thực của tham số hơn.</li> <li>Khi độ tin cậy là 100% thì khoảng tin cậy sẽ là <math>(-\infty; +\infty)</math></li> </ul>	1
c	<ul style="list-style-type: none"> <li>Xét cặp giả thuyết: <ul style="list-style-type: none"> <li><math>H_0: \mu = 11.3</math></li> <li><math>H_1: \mu \neq 11.3</math></li> </ul> </li> <li>Tính giá trị kiểm định: <math display="block">t_o = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{11.036 - 11.3}{1.845/\sqrt{169}} = -1.86</math> </li> <li><math>t_o = -1.86 \Rightarrow p\text{-value} = 0.0323 &gt; \alpha/2 = 0.025 \Rightarrow</math> Chưa đủ cơ sở để bác bỏ <math>H_0</math>, nghĩa là phát biểu trên đúng với mức ý nghĩa 0.05. (Lưu ý: Nếu sinh viên không sử dụng phương pháp p-value mà vẫn đưa ra kết quả đúng thì vẫn được tính điểm phần này)</li> </ul>	1
d	<ul style="list-style-type: none"> <li>Tỷ lệ bỏ cho năng suất cao của mẫu: <math>f = 159/169 = 0.9408 \approx 0.941</math></li> <li>Xét cặp giả thuyết: <ul style="list-style-type: none"> <li><math>H_0: p \leq 0.9</math></li> <li><math>H_1: p &gt; 0.9</math></li> </ul> </li> <li>Tính giá trị kiểm định: <math display="block">z_o = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.941 - 0.9}{\sqrt{\frac{0.9 * 0.1}{169}}} = 1.78</math> </li> <li><math>z_o = 1.78 \Rightarrow p\text{-value} = 0.0375 &lt; \alpha = 0.05 \Rightarrow</math> Bác bỏ <math>H_0</math>, nghĩa là phát biểu của giám đốc đúng với mức ý nghĩa 0.05.</li> </ul>	1

Mẫu: 2b ĐATL

	(Lưu ý: Nếu sinh viên không sử dụng phương pháp p-value mà vẫn đưa ra kết quả đúng thì vẫn được tính điểm phần này)	
<b>Câu 3</b>		<b>3.0 điểm</b>
a		1
	<ul style="list-style-type: none"> <li>- Scatter plot là 2 trục, trục x tương ứng với dữ liệu cột mức độ chịu nhiệt (Rating), trục y tương ứng với dữ liệu cột giá (Price), dữ liệu xấp xỉ dạng đường thẳng</li> <li>- Ghi chú: vẽ sai trục (Rating: trục y; Price: trục x)</li> </ul>	1 -0,5
b		1
	<ul style="list-style-type: none"> <li>- Phương trình hồi quy có dạng: <math>y = \beta_0 + \beta_1 \cdot x</math></li> <li>- Ta có: <math>\beta_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}</math>; <math>\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}</math></li> <li>Với: <math>\bar{x} = 19</math>; <math>\bar{y} = 259</math>; <math>\overline{xy} = 4003.727</math>; <math>\sigma_x^2 = 13.184</math></li> <li>- Phương trình hồi quy: <math>y = 359.267 - 5.277 \cdot x</math></li> </ul>	0.5 0.5
c		1
	$x_0 = 20 \Rightarrow y_0 = 359.267 - 5.277 \cdot 20 = 253.727 \approx 254$	
	<b>TỔNG ĐIỂM</b>	<b>10.0 điểm</b>



**KHOA CÔNG NGHỆ THÔNG TIN**

**Môn thi :** Thống Kê Máy Tính & Ứng Dụng

**Lớp/Lớp học phần:** DHTH15 – DTDTMT16

**Ngày thi:** 15/12/2021

Thời gian làm bài: 75 phút  
(Không kể thời gian phát đề)

Họ và tên thí sinh ..... ; MSSV: .....

**Lưu ý:**

- Các kết quả tính toán cần làm tròn đến 3 chữ số thập phân
- Mỗi bài làm ghi rõ STT, Họ tên, Mã sinh viên
- Để tránh sót ảnh, sau khi làm xong nên scan bài làm của mình thành file pdf
- Định dạng đặt tên: [stt]\_[hoten]\_[msv].pdf (Ví dụ: 03\_tranminhhieu\_19287754.pdf)

**Câu 1.** (4 điểm)

Nhằm thực hiện việc kiểm định môn TKMT học kì này, người ta rút ngẫu nhiên 40 bài thi và phân loại theo như kết quả bảng sau:

Điểm	0-1	2-3	4-5	6-7	8-9	10
Số bài thi	A	5	6	18	7	B

Trong đó  $B = (STT \% 2) + 3$ ;  $A = 5 - B$ ;

a/ Dựa vào mẫu trên, tìm khoảng ước lượng điểm thi trung bình  $\mu$  của môn TKMT với độ tin cậy 90%.

b/ Bài thi có điểm  $> 5$  được gọi là "Đạt trên mức trung bình". Hãy tìm khoảng ước lượng cho tỉ lệ  $p$  của bài thi "Đạt trên mức trung bình" với độ tin cậy 90%.

**Câu 2:** (4 điểm)

Vì dịch bệnh đã giảm nên nhà trường thực hiện việc học trực tiếp đối với học sinh toàn trường.

a/ Nhà trường dự toán chi phí phòng bệnh mỗi lớp tối đa 450 ngàn/tuần, độ lệch chuẩn 45 ngàn. Nhưng khi xem xét số tiền đã chi cho 40 lớp thì chi phí phòng dịch trung bình mỗi lớp thực tế là 475 ngàn/tuần. Hãy kiểm định dự toán của nhà trường có đúng hay không với độ tin cậy 99%?

b/ Vì học sinh chưa đi học đông đủ nên nhà trường thực hiện việc khảo sát một số phụ huynh đại diện cho phụ huynh có con em đang học tại trường về việc có đồng ý cho con đến trường học trực tiếp không. Kết quả số phụ huynh đồng ý cho con học trực tiếp như sau:

Khối	1	2	3	4	5
Số phụ huynh đồng ý	X	35	60	30	Y

Trong đó X, Y bất kỳ thoả:  $X = (STT \% 2) + 10$ ;  $X + Y = 25$ ;

Nhà trường cho rằng tổng số số phụ huynh các khối 3, 4, 5 đồng ý chiếm 80% phụ huynh của trường đồng ý. Với mức ý nghĩa 2,5% hãy kiểm định ý kiến của nhà trường đúng không?

**Câu 3:** (2 điểm)

Nhằm có kế hoạch khuyến khích sinh viên nghiên cứu khoa học, nhà trường đã khảo sát mức lương của sinh viên sau khi ra trường ứng với kết quả tham gia nghiên cứu khoa học như sau:

Kết quả nghiên cứu khoa học (điểm)	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
Mức lương sau khi ra trường (triệu)	11,9	14,2	16,5	18,8	21,1	23,4	25,7	28+Z

Trong đó  $Z = STT \% 4$ ;

a/ Tính hệ số tương quan giữa kết quả nghiên cứu khoa học của sinh viên và mức lương đi làm?

b/ Tìm hàm hồi qui tuyến tính của mức lương theo kết quả nghiên cứu khoa học của sinh viên?

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP  
THÀNH PHỐ HỒ CHÍ MINH

KHOA CNTT

ĐỀ THI CUỐI KỲ

Môn thi : Thống kê máy tính & ứng dụng

Lớp/Lớp học phần: DHTH16

Ngày thi: 26/05/2022

Thời gian làm bài: 60 phút  
(Không kể thời gian phát đề)

Họ và tên thí sinh .....; MSSV: .....

**Câu 1: (4 điểm) – LO2**

Một khách hàng nhận được lô hàng từ một nhà máy sản xuất bút bi rẻ tiền. Để ước lượng tỉ lệ bút hỏng, khách hàng lấy ngẫu nhiên 300 bút từ lô hàng và thấy có 30 bút hỏng. Hãy cho biết:

- Hãy ước lượng tỷ lệ bút hỏng của nhà máy với độ tin cậy là 95%
- Nếu sử dụng mẫu trên để tính tỉ lệ bút bi hỏng với độ chính xác của ước lượng là 2.5% thì độ tin cậy của ước lượng là bao nhiêu?

**Câu 2: (4 điểm) – LO3**

Trọng lượng trung bình của gà khi xuất chuồng ở một trại chăn nuôi là 3.3kg/con. Năm nay người ta sử dụng một loại thức ăn mới. Cân thử 15 con khi xuất chuồng ta thu được các số liệu như sau:

3.25, 2.50, 4.00, 3.75, 3.80, 3.90, 4.02, 3.60, 3.80, 3.20, 3.82, 3.40, 3.75, 4.00, 3.50

Giả sử trọng lượng gà là đại lượng ngẫu nhiên có phân phối chuẩn. Dựa vào thủ tục kiểm định thống kê bạn hãy:

- Cho biết loại thức ăn mới có làm thay đổi trọng lượng của gà không với mức ý nghĩa là 5%
- Nếu trại chăn nuôi báo cáo trọng lượng trung bình của gà khi xuất chuồng là 3.5kg/con thì có chấp nhận được không với mức ý nghĩa 5%

**Câu 3: (2 điểm)**

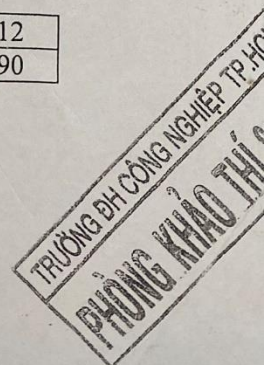
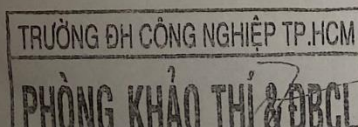
Dựa vào bảng số liệu sau bạn hãy cho biết có thể xây dựng mô hình hồi quy tuyến tính của biến Y theo X hay không? Nếu có bạn hãy tính các hệ số xác định mô hình.

X	6	7	9	9	7	8	6	12
Y	45	80	70	85	60	55	75	90

----- Hết -----

Lưu ý: Đề thi được sử dụng tài liệu giấy.

Cán bộ coi thi không giải thích gì thêm.





**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP  
THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐỀ THI CUỐI KỲ**

**Môn thi :** Thống Kê Máy Tính & Ứng Dụng

**Lớp/Lớp học phần:** DHTH17VL

**Ngày thi:** 24/06/2022

Thời gian làm bài: 60 phút

(Không kể thời gian phát đề)

**Câu 1 (2.0đ):**

Ngày 15/5/2022, phòng đào tạo trường ĐHCN thông báo điểm Toeic của K14 có phân phối chuẩn với giá trị trung bình 415 và độ lệch chuẩn 35. Hỏi:

- Tỷ lệ sinh viên không đủ điều kiện tốt nghiệp là bao nhiêu? (sinh viên cần có điểm Toeic  $\geq 450$  mới được xét tốt nghiệp)
- Nhà trường tìm giải pháp để nâng cao tỷ lệ sinh viên đủ điều kiện xét tốt nghiệp sau 3 tháng huấn luyện cấp tốc đối với sinh đang có điểm Toeic từ 420 đến dưới 450. Hỏi tỷ lệ sinh viên thuộc diện “huấn luyện cấp tốc” là bao nhiêu?
- Sinh viên có điểm Toeic  $\geq 600$  sẽ được miễn thi đầu vào ngoại ngữ ở bậc cao học. Hỏi tỷ lệ sinh viên đủ điều kiện đầu vào ngoại ngữ bậc cao học là bao nhiêu?

**Câu 2 (4.0đ):**

Trong một công ty sản xuất bóng đèn, người quản lý lấy mẫu ngẫu nhiên 13 bóng đèn trong kho và bật bóng đèn cho đến khi nó hỏng để đo tuổi thọ mỗi bóng (đơn vị: giờ). Sau khi thực hiện, người quản lý thu được bảng số liệu dưới đây:

342	426	317	545	264	451	631	512	266	492	562	298	320
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Giả sử rằng tuổi thọ bóng đèn trong công ty là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 120 giờ. Bạn hãy:

- Ước lượng tuổi thọ bóng đèn của công ty với độ tin cậy 90%.
- Bóng đèn có tuổi thọ trên 300 giờ là bóng đèn đạt chuẩn. Tìm khoảng ước lượng tỷ lệ bóng đèn đạt chuẩn với độ tin cậy 95%?

**Câu 3 (1.0đ):**

Xét 2 trường hợp: (1) Cô gái A từ chối hẹn hò với anh B là một người đàn ông tốt (đúng tiêu chuẩn A cần tìm) và (2) Cô gái A chấp nhận hẹn hò với anh C là người đàn ông không tốt (không đúng tiêu chuẩn A cần tìm). Hỏi sai lầm của A trong từng trường hợp thuộc loại nào?

**Câu 4 (3.0đ):**

Dữ liệu sau đây được thu thập về chiều cao-height (inch) và cân nặng (pound) của phụ nữ bơi lội

Height	68	64	62	65	66
Weight	132	108	102	115	128

- Tìm phương trình hồi quy ứng với dữ liệu mẫu đã cho.
- Nếu một vận động viên bơi lội có chiều cao là 63 inch, bạn sẽ ước tính cân nặng của cô ấy là bao nhiêu?

## SỬA BÀI THI CUỐI KỲ 2021

### Câu 1. (4 điểm)

Nhằm thực hiện việc kiểm định môn TKMT học kì này, người ta rút ngẫu nhiên 40 bài thi và phân loại theo như kết quả bảng sau:

Điểm	0-1	2-3	4-5	6-7	8-9	10
Số bài thi	<b>A</b>	5	6	18	7	<b>B</b>

Trong đó **B=(STT % 2) + 3; A= 5 – B;**

a/ Dựa vào mẫu trên, tìm khoảng ước lượng điểm thi trung bình  $\mu$  của môn TKMT với độ tin cậy 90%.

b/ Bài thi có điểm  $> 5$  được gọi là "Đạt trên mức trung bình". Hãy tìm khoảng ước lượng cho tỉ lệ  $p$  của bài thi "Đạt trên mức trung bình" với độ tin cậy 90%.

Điểm	0.5	2.5	4.5	6.5	8.5	10
Số bài	2	5	6	18	7	2

a) Công thức:  $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \rightarrow$  khoảng ước lượng:  $\bar{X} \pm E \rightarrow \bar{x} - E < \mu < \bar{x} + E$ .

Ta có:  $\bar{x} = \frac{0.5 \cdot 2 + 2.5 \cdot 5 + 4.5 \cdot 6 + 6.5 \cdot 18 + 8.5 \cdot 7 + 10 \cdot 2}{40} = \dots$

$$s = \sqrt{\sum \frac{(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{2(0.5 - \bar{x})^2 + 5(2.5 - \bar{x})^2 + 6(4.5 - \bar{x})^2 + 18(6.5 - \bar{x})^2 + 7(8.5 - \bar{x})^2 + 2(10 - \bar{x})^2}{39}}$$

$$z_{\alpha/2} = 1.65 \rightarrow E =$$

b) Số bài  $> 5$  là:  $18 + 7 + 2 = 27 \rightarrow p = 27/40$  và  $q = 1 - 27/40 = 13/40$ .

$$\hat{p} - E < p < \hat{p} + E$$

Khoảng tin cậy 90%  $\rightarrow z_{\alpha/2} = 1.63 \rightarrow E = 1.63 \sqrt{\frac{(27/40)(13/40)}{40}} = 0.121$  nên

Ta có:  $p = 27/40 = 0.675$  nên

$$0.675 - 0.121 < p < 0.675 + 0.121.$$

**Câu 2: (4 điểm)**

Vì dịch bệnh đã giảm nên nhà trường thực hiện việc học trực tiếp đối với học sinh toàn trường.

a/ Nhà trường dự toán chi phí phòng bệnh mỗi lớp tối đa 450 ngàn/tuần, độ lệch chuẩn 45 ngàn. Nhưng khi xem xét số tiền đã chi cho 40 lớp thì chi phí phòng dịch trung bình mỗi lớp thực tế là 475 ngàn/tuần. Hãy kiểm định dự toán của nhà trường có đúng hay không với độ tin cậy 99%?

b/ Vì học sinh chưa đi học đông đủ nên nhà trường thực hiện việc khảo sát một số phụ huynh đại diện cho phụ huynh có con em đang học tại trường về việc có đồng ý cho con đến trường học trực tiếp không. Kết quả số phụ huynh đồng ý cho con học trực tiếp như sau:

Khối	1	2	3	4	5
Số phụ huynh đồng ý	<b>X</b>	35	60	30	<b>Y</b>

**Trong đó X, Y bất kỳ thoả:  $X = (\text{STT} \% 2) + 10$ ;  $X + Y = 25$ ;**

Nhà trường cho rằng tổng số số phụ huynh các khối 3, 4, 5 đồng ý chiếm 80% phụ huynh của trường đồng ý. Với mức ý nghĩa 2,5% hãy kiểm định ý kiến của nhà trường đúng hay không?

a) Kiểm định giả thuyết:  $H_0: \mu = 450$  và  $H_1: \mu > 450$ .

Ta có  $n = 40$ ,  $\sigma = 45$ , độ tin cậy  $1 - \alpha = 99\% \rightarrow \alpha / 2 = 0.005$ .

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{475 - 450}{45 / \sqrt{40}} \sim 3.51.$$

$\alpha / 2 = 0.005 \rightarrow p\text{-value} = 2.57 < z$  nên bác bỏ  $H_0$ .

Kết luận: chi phí dự toán của nhà trường vượt mức 450 ngàn / tuần với độ tin cậy 99%.

b) Ta có bảng:

Khối	1	2	3	4	5
Số PH	10	35	60	30	15

Tổng số PH = 150 = n, tỷ lệ **p mẫu =  $105 / 150 = 0.7$**

Ta có:  $np \geq 5$  và  $n(1-p) \geq 5$  nên áp dụng được việc xấp xỉ bằng pp chuẩn.

Kiểm định giả thuyết:  $H_0$  là  $p = 0.8$  và  $H_1: p \neq 0.8$  mức ý nghĩa  $\alpha / 2 = 0.025$ .

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Công thức:  $z = \frac{0.7 - 0.8}{\sqrt{0.8 * 0.2}} \sqrt{150} \sim -3.06.$

Ta có:  $p\text{-value} = 1.96 * 2 = 3.92 > |z| = 3.06$  nên không bác bỏ  $H_0$ .



**Câu 3:** (2 điểm)

Nhằm có kế hoạch khuyến khích sinh viên nghiên cứu khoa học, nhà trường đã khảo sát mức lương của sinh viên sau khi ra trường ứng với kết quả tham gia nghiên cứu khoa học như sau:

Kết quả nghiên cứu khoa học (điểm)	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
Mức lương sau khi ra trường (triệu)	11,9	14,2	16,5	18,8	21,1	23,4	25,7	<b>28+Z</b>

**Trong đó Z = STT % 4;**

a/ Tính hệ số tương quan giữa kết quả nghiên cứu khoa học của sinh viên và mức lương đi làm?

b/ Tìm hàm hồi qui tuyến tính của mức lương theo kết quả nghiên cứu khoa học của sinh viên?

$$\begin{aligned} \text{Cor}(Y, X) &= \frac{\text{Cov}(Y, X)}{s_y s_x} \\ &= \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}} \end{aligned}$$

Tính trước  $\bar{x}, \bar{y} \rightarrow$  tổng các tích, và tổng bình phương.

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

SV tự áp dụng các công thức trên và tính lại giá trị cụ thể cho câu a), b).

**Đề cuối kỳ TKMT 2022.**

**Câu 1:** (2 điểm) Một người đánh cờ với xác suất thắng là 0.65 đã đăng ký tham gia một giải cờ tướng online đánh tổng cộng 200 ván cờ trong vòng 1 tuần. Nếu người này thắng được từ 150 ván trở lên thì sẽ giành được giải thưởng lớn từ ban tổ chức. Tính xác suất để điều đó xảy ra.

**Câu 2:** (3 điểm) Thống kê lượng calo có trong một loại bánh, người ta thu được mẫu số liệu:

Lượng calo	160	170	180	190	200	210	220
Số bánh	8	12	16	20	28	6	6

- Tính trung bình và độ lệch chuẩn của mẫu bánh này.
- Hãy ước lượng khoảng cho lượng calo trung bình có trong mỗi chiếc bánh, với độ tin cậy là 90%.
- Giả sử bánh đạt chuẩn là bánh có hàm lượng calo lớn hơn 180 calo. Hãy ước lượng tỉ lệ bánh đạt chuẩn với độ tin cậy là 95%.

**Câu 3:** (2 điểm) Để điều tra tình hình sản xuất của một nhà máy, người ta kiểm tra ngẫu nhiên một số sản phẩm và thu được bảng số liệu như sau:

$x$ (cm)	4	4.5	5	5.5	6	6.5	7
$n$ (số sản phẩm)	7	11	20	8	25	9	20

- Những sản phẩm có chiều dài lớn hơn 5 cm được gọi là *sản phẩm loại A*. Với độ tin cậy 99%, hãy ước lượng khoảng cho tỉ lệ sản phẩm loại A của nhà máy.
- Theo một báo cáo, tỉ lệ sản phẩm loại A của nhà máy là 65%. Với mức ý nghĩa 5%, hỏi báo cáo này có đáng tin hay không?

**Câu 4:** (3 điểm) Cho bảng số liệu khảo sát về lượng hút thuốc  $X$  (điều/ngày) và tuổi thọ  $Y$  (tuổi):

Số điều/ngày	0	6	20	2	5	6	20	0	4	3
Tuổi thọ	90	68	55	70	60	65	50	82	75	63

- Hãy tính giá trị hiệp phương sai và hệ số tương quan của hai đại lượng trên.
- Lập hàm hồi quy tuyến tính mẫu của tuổi thọ ( $Y$ ) theo số điều thuốc hút/ngày ( $X$ ). Từ đó dự báo tuổi thọ trung bình khi hút thuốc 15 điều/ngày.

a) Ta có:  $n = 10$ ,  $\bar{x} = 6,6$ ,  $s_x = 7,382$ ,  $\bar{y} = 67,8$ ,  $s_y = 12,127$ .

$$\text{cov}(X, Y) = \frac{\sum (\bar{x} - x_i)(\bar{y} - y_i)}{n - 1} = -71,978.$$

Hàm hồi quy:  $y = b + ax$  với

$$a = \frac{\text{cov}(X, Y)}{s_x^2} = \frac{-71,978}{7,382^2} = -1,321 \text{ và } b = \bar{y} - a\bar{x} = 67,8 - (-1,321) \cdot 6,6 = 76,518.$$

Đáp số:  $y = 76,518 - 1,321x$ .

b) Thay  $x = 15$  thì được  $y = 56,703$  tuổi.