I. XÁC SUẤT - BIẾN SỐ NGẪU NHIÊN

Bài 1:

Tỷ lệ mắc bệnh X ở lô chuột thứ nhất (lô I) là 0,10 và ở lô chuột thứ hai (lô II) là 0,07.

1) Lấy ngẫu nhiên 3 chuột ở lô I. Tính xác suất có ít nhất một chuột mắc bệnh X. Phải lấy ít nhất bao nhiêu chuột ở lô I để xác suất có ít nhất một chuột mắc bệnh X lớn hơn 0,90?

2) Lấy ngẫu nhiên ra mỗi lô một con chuột. Tính xác suất để có một chuột mắc bệnh X và một chuột không mắc bệnh X?

3) Chọn ngẫu nhiên một lô, rồi từ đó lấy ngẫu nhiên ra hai chuột. Tính xác suất để có một chuột mắc bệnh X và một chuột không mắc bệnh X?

Bài giải

[1] Gọi K là số chuột bị bệnh X. Ta có:

$$P(K \geq 1) = 1 - P(K = 0) = 1 - (0,90)^3 = 0,271$$

Gọi n là số chuột lấy ra, ta có:

$$P(K > 1) = 1 - P(K = 0) = 1 - (0,90)^n > 0,90$$

$$(0,90)^n < 0,10 \Leftrightarrow n \ln(0,90) < \ln(0,10)$$

$$n > \frac{\ln(0,10)}{\ln(0,90)} = 21,85$$

Vậy phải lấy ra ít nhất 22 chuột.

[2] Gọi B = biến cố "lấy ra chuột mắc bệnh X" ; và C là biến cố "có một chuột mắc bệnh X và một chuột không mắc bệnh X", ta có:

$$P(C) = P(B|I).P(\overline{B}|II) + P(\overline{B}|I).P(B|II)$$

$$P(C) = (0,10)(0,93) + (0,90)(0,70) = 0,156$$

[3] Nếu chọn ngẫu nhiên một trong hai lô, ta có:

$$P(I) = P(II) = \frac{1}{2}$$

$$P(C) = P(C|I).P(I) + P(C|II).P(II)$$

$$P(C) = 2(0,10)(0,90) \times \frac{1}{2} + 2(0,07)(0,93) \times \frac{1}{2} = 0,155$$

Bài 2:

Số bệnh nhân đến khám thường bị một trong ba bệnh B_1, B_2, B_3 với tỷ lệ 0,5; 0,3; 0,2. Để chẩn đoán, bác sĩ dùng ba xét nghiệm T_1, T_2, T_3, kết quả mỗi xét nghiệm sẽ là + hay – (ký hiệu là 1 hay 0)

Cho biết:

$$P(111)|B_3) = 1,0; P(011)|B_2) = 0,8; P(000)|B_2) = 0,2$$

$$P(011|B_1) = 0,5; P(001|B_1) = 0,4; P(100|B_1) = 0,1$$

1) Một bệnh nhân đến khám có T_1 âm, T_2 dương thì khả năng người này bị B_1, B_2, B_3 là bao nhiêu?

2) Chứng tỏ rằng có 2 trong 3 xét nghiệm đó đủ để chẩn đoán một bệnh nhân đến khám là bị B_1 hoặc B_2 hoặc B_3.

Bài giải

[1]

$T_1T_2T_3$	000	001	010	100	011	101	110	111
B_1		0,4		0,1	0,5			
B_2	0,2		0,8					
B_3								1

$$P(01) = P(01|B_1)P(B_1) + P(01|B_2)P(B_2)$$

$$P(01) = (0,50)(0.50) + (0,80)(0,30) = 0,49$$

$$P(B_1|01) = \frac{P(01|B_1)P(B_1)}{P(01)} = \frac{(0,5)(0,5)}{0,49} = 0,51$$

$$P(B_2|01) = \frac{P(01|B_2)P(B_2)}{P(01)} = \frac{(0,8)(0,3)}{0,49} = 0,49$$

$$P(B_3|01) = \frac{P(01|B_3)P(B_3)}{P(01)} = \frac{0(0,2)}{0,49} = 0$$

[2] • Nếu dùng hai xét nghiệm T_1, T_2 thì:

$(T_1 = 0, T_2 = 0)$, kết quả là B_1 hay B_2

$(T_1 = 0, T_2 = 1)$, kết quả là B_1 hay B_2

$(T_1 = 1, T_2 = 0)$, kết quả là B_1

$(T_1 = 1, T_2 = 1)$, kết quả là B_3

• Nếu dùng hai xét nghiệm T_2, T_3 thì:

$(T_2 = 0, T_3 = 0)$, kết quả là B_1 hay B_2

$(T_2 = 0, T_3 = 1)$, kết quả là B_1

$(T_2 = 1, T_3 = 0)$, kết quả là B_2

$(T_2 = 1, T_3 = 1)$, kết quả là B_3

- Nếu dùng hai xét nghiệm T_1, T_3 thì:

$(T_1 = 0, T_3 = 0)$, kết quả là B_2

$(T_1 = 0, T_3 = 1)$, kết quả là B_1

$(T_1 = 0, T_3 = 0)$, kết quả là B_2

$(T_1 = 1, T_3 = 1)$, kết quả là B_3

Vậy chỉ cần hai xét nghiệm T_1, T_3 đủ để chẩn đoán B_1, B_2, B_3

Bài 3:

Hai xét nghiệm (XN) T_1 và T_2 được dùng để chẩn đoán bệnh B. Bệnh này không lây lan và diễn tiến đến tử vong nếu không được chẩn đoán và điều trị, mà việc điều trị cũng ít tốn kém. T_1 cho âm giả 2%, dương giả 25%; còn T_2 cho dương giả 2%, âm giả 25%.

1) Hỏi độ nhạy và độ chuyên của T_1 và T_2.

2) Nếu chỉ được phép dùng một trong hai xét nghiệm để chẩn đoán một bệnh nhân nghi ngờ bị bệnh B, Anh (Chị) chọn xét nghiệm nào ? Giải thích.

Bài giải

[1] Tính các độ nhạy nhạy và độ chuyên:

* Độ nhạy T_1: $P(T_1^+ | B) = 1 - P(T_1^- | B) = 0,98$

* Độ chuyên T_1: $P(T_1^- | \overline{B}) = 1 - P(T_1^+ | \overline{B}) = 0,75$

* Độ nhạy T_2: $P(T_2^+ | B) = 1 - P(T_2^- | B) = 0,75$

* Độ chuyên T_2: $P(T_2^- | \overline{B}) = 1 - P(T_2^+ | \overline{B}) = 0,98$

[2] Ta nên chọn xét nghiệm T_1 vì T_1 có độ nhạy cao hơn T_2, nên xác suất bỏ sót bệnh ít hơn.

Bài 4:

Một bệnh nhân uống nhầm một trong hai loại thuốc A hoặc B. Các lọ thuốc bề ngoài trông thật giống nhau, lại để chung trong một ngăn kéo. Cả hai loại đều có hại đối với bệnh nhân này.

1) Có 2 lọ loại A và 3 lọ loại B để trong một ngăn kéo. Bệnh nhân vô tình lấy một lọ ra dùng. Dùng phải A hay B đều có khả năng bị hạ huyết áp nghiêm trọng. Khả năng đó là 75% nếu dùng A, 20% nếu dùng B. Tính khả năng người này bị hạ huyết áp.

2) Quả thật người này bị hạ huyết áp nghiêm trọng sau khi dùng thuốc. Tùy theo bệnh nhân uống nhầm A hay B mà cách xử trí là hoàn toàn khác nhau và không tương thích (incompatibles). Nếu không xử trí thích hợp thì khả năng bị di chứng là 10% nếu dùng A, 20% nếu dùng B.

Phải xử trí bệnh nhân này theo hướng nào ? (Hướng nhầm A hay nhầm B ?)

Bài giải

[1] Gọi K là biến cố "bị hạ huyết áp nghiêm trọng".

$$P(K) = P(K|A)P(A) + P(K|B)P(B)$$

$$P(K) = (0,75)\frac{2}{5} + (0.20)\frac{3}{5} = 0,42$$

[2] $$P(A') = P(A|K) = \frac{P(K|A)P(A)}{P(K)} = \frac{(0,75)\frac{2}{5}}{0,42} = 0,714$$

$$P(B') = P(B|K) = \frac{P(K|B)P(B)}{P(K)} = \frac{(0,20)\frac{3}{5}}{0,42} = 0,286$$

Gọi D là biến cố "bệnh nhân bị di chứng", ta có:

$$P(A'D) = P(D|A')P(A') = (0,10)(0,714) = 0,0714$$

$$P(B'D) = P(D|B')P(B') = (0,20)(0,286) = 0,0572$$

Vậy phải xử trí theo hướng nhầm A.

Bài 5:

Một người đang khỏe mạnh, nếp sống không đến nổi buông thả mà cứ sợ mình "bị SIDA" nên xin làm xét nghiệm để kiểm tra. Có hai xét nghiệm, cách thực hiện và chi phí như nhau: T_1 có độ nhạy 90%, độ chuyên 80%. còn T_2 có độ nhạy 80% độ chuyên 90%.

1) Anh (Chị) chọn xét nghiệm nào để kiểm tra người này? Giải thích.

2) Cho biết trong dân số tỷ lệ bệnh này là 1/1000, và xét nghiệm vừa chọn lại trả lời dương tính. Tính khả năng người này bị bệnh trên.

Bài giải

[1] Tránh sai lầm "không bệnh mà cho là có", vậy chọn xét nghiệm nào có tỷ lệ dương giả thấp nghĩa là tỷ lệ âm thật cao. Do đó chọn xét nghiệm T_2.

[2] $P(T_2^+) = P(T_2^+|B)P(B) + P(T_2^+|\overline{B})P(\overline{B})$

$$P(T_2^+) = (0,80)(0,001) + (1 - 0,90)(0,999) = 0,1007$$

$$P(B|T_2^+) = \frac{P(T_2^+|B)P(B)}{P(T_2^+)}$$

$$P(B|T_2^+) = \frac{(0,80)(0,001)}{0,1007} = 0,0079 = 0,008$$

Bài 6:

Một hồi cứu về bệnh ung thư vú đã phẩu thuật cho biết: Tỷ lệ sống quá 5 năm là 60%, tỷ lệ có (hạch) di căn là 30%. Trong hồi cứu này số ca vừa sống qua 5 năm và có di căn chỉ bằng nữa số ca không có di căn và không sống qua 5 năm.

Một người bị ung thư vú và không có hạch di căn, tính khả năng người này sống quá 5 năm sau khi phẩu thuật.

Bài giải

	Sống > 5 năm	Sống ≤ 5 năm	Cộng
Có hạch	a	b	0,30
Không hạch	c	d	0,70
Cộng	0,60	0,40	1

Ta có hệ phương trình: $\begin{cases} a + b = 0.30 \\ a + c = 0.60 \\ a \quad = \dfrac{1}{2}d \\ a + b + c + d = 1 \end{cases} \Rightarrow \begin{cases} a = 0,10 \\ b = 0,20 \\ c = 0,50 \\ d = 0,20 \end{cases}$

P (sống quá 5 năm không hạch di căn) = $\dfrac{c}{c+d}$

$= \dfrac{0,50}{0,70} = 0,71$

Bài 7:

Một người nghi ngờ bị bệnh B, với B(B) = 0,3; cho làm xét nghiệm T.

Xét nghiệm T sẽ trả về hoặc dương tính T(+) hoặc âm tính T(–). Trong số những người T(+) chỉ có 80% là bị bệnh B, còn trong số những người T(–) có 90% không bị bệnh này.

1) Tính độ nhạy và độ chuyên của xét nghiệm T.

2) Khả năng người này xét nghiệm trả về T(+) là bao nhiêu?

Bài giải

[1] Gọi: $x = P(T^+ | B^+)$ là độ nhạy của xét nghiệm T.

$y = P(T^- | B^-)$ là độ chuyên của xét nghiệm T.

Ta có: $P(T^+) = P(T^+ | B^+)P(B^+) + P(T^+ | B^-)P(B^-)$

$P(T^+) = (0,3)x + (0,7)(1 - y)$

$P(T^-) = P(T - | B^+)P(B^+) + P(T^- | B^-)P(B^-)$

$$P(T^-) = (0,3)(1-x) + (0,7)y$$

$$P(B^+|T^+) = \frac{P(T^+|B^+)P(B^+)}{P(T^+)}$$

$$P(B^+|T^-) = \frac{(0,30)x}{(0,3)x + (0,7)(1-y)} = 0,80$$

$$P(B^-|T^-) = \frac{P(T^-|B^-)(B^-)}{P(T^-)}$$

$$P(B^-|T^-) = \frac{(0,70)y}{(0,3)(1-y) + (0,7)y} = 0.90$$

Ta có hệ phương trình: $\begin{cases} \dfrac{(0,30)x}{(0,3)x + (0,7)(1-y)} = 0,80 & (1) \\ \dfrac{(0,70)y}{(0,3)(1-y) + (0,7)y} = 0,90 & (2) \end{cases}$

(1): $(0,30)x = (0,24)x + (0,56)(1-y) \Rightarrow (0,06)x + (0,56)y = 0,56$

(2): $(0,70)y = (0,27)(1-y) + (0,63)y \Rightarrow (0,27)x + (0,07)y = 0,27$

Giải hệ phương trình này ta được kết quả: x = 0,7619, y = 0,9183

[2] $P(T^+) = P(T^+|B^+)P(B^+) + P(T^+|B^-)P(B^-)$

$$P(T^+) = (0,76)(0,30) + (1-0,92)(0,70) = 0,28$$

Bài 8:

Một nhà Hộ Sinh trong điều kiện hiện tại trung bình giải quyết 6 ca/ngày, và khả năng tối đa 10 ca/ngày. Anh (Chị) dự đoán xem trong một năm có bao nhiêu ngày Nhà Hộ Sinh đó chịu quá mức tối đa? Cho biết số ca sinh trong ngày X ~ Poisson(μ).

Bài giải

Gọi X = số ca cần giải quyết mỗi ngày, X ~ Poisson(6).

$$f(x) = \begin{cases} e^{-6} \times \dfrac{6^x}{x!}; x = 0,1,2,... \\ 0 \qquad\qquad ; \text{nơi khác} \end{cases}$$

$$P(X \le 10) = \sum_{x=0}^{10} e^{-6} \times \frac{6^x}{x!} = 0,975$$

$$P(X > 10) = 1 - 0.957 = 0,043$$

Số ngày quá tải trong năm: n = 365 × 0,043 = 16 ngày

Bài 9:

Một hồi cứu về ung thư vú đã phẩu thuật cho biết: Tỷ lệ sống quá 5 năm là 90%, tỷ lệ không có hạch di căn trong số những người sống quá 5 năm là 80%, tỷ lệ có hạch di căn trong số những người không sống quá 5 năm là 70%.

Một người bị ung thư vú mà có hạch di căn, tính khả năng người này sống quá 5 năm sau khi phẩu thuật.

Bài giải

Gọi A là biến cố "sống quá 5 năm", $P(A) = 0.90$

B là biến cố "có hạch di căn", $P(\overline{B}|A) = 0,80, P(B|\overline{A}) = 0,70$

Ta có: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

$$P(B) = (1 - 0,80)(0,90) + (0,70)(1 - 0,90) = 0,25$$

$$P(A|B) = \frac{(1 - 0,80)(0,90)}{0,25} = 0,72$$

Bài 10:

Cho biết tỷ lệ K–vú đối với bà trên 50 tuổi là 0,3%, những bà này có mẹ hay chị em ruột bị K–vú thì khả năng bị bệnh là 1%, nếu một vú có khối u thì khả năng bị bệnh là 38%. MMG (Mammography) có sens = 73,3%, spec = 98,4%; FNA (Fine needle aspiration) có sens = 93%, spec = 92%. Một cô giáo 54 tuổi lo ngại về ung thư vú, đến phòng khám Anh (Chị).

1) Khi khỏi bệnh, được biết có chị ruột đã chết vì K–vú. Khả năng bà này bị K–vú là bao nhiêu?

2) Lúc khám bệnh, một vú có khối u sờ được (palpable mass). Anh (Chị) cần một xét nghiệm giúp chẩn đoán bệnh này.

Sai lầm nghiêm trọng ở đây là bỏ sót bệnh. Vì để khối u càng lớn thì khả năng di căn càng cao, dự đoán sống quá 5 năm càng thấp.

Giữa MMG và FNA, Anh (Chị) chọn xét nghiệm nào? Lý do? Nếu xét nghiệm đó trả về là + thì khả năng người này bị bệnh K là bao nhiêu?

Bài giải

[1] $P(B^+) = 0,01$

[2] Vì sai lầm nghiêm trọng là bỏ sót bệnh, vậy ta chọn xét nghiệm nào có xác suất âm giả thấp, nghĩa là dương thật cao. Do đó chọn FNA có sens = 0,93

$$P(T^+) = P(T^+|B^+)P(B^+) + P(T^+|B^-)P(B^-)$$

$$P(T^+) = (0,93)(0,38) + (0,08)(0,62) = 0,403$$

$$P(B^+|T^+ = \frac{P(T+|B+)P(B+)}{P(T+)}$$

$$= \frac{0,93 \times 0,38}{0,403} = 0,88$$

Bài 11:

Tại một địa phương trong dân số, tỷ lệ bệnh sốt rét là 20%, tỷ lệ lách to là 30%. Trong số người bị sốt rét thì tỷ lệ lách to chiếm 80%.

1) Chọn ngẫu nhiên một người trong địa phương đó, thấy người này có lách to, tính khả năng người này không bị sốt rét.

2) Khám 10 người tại địa phương đó, tính xác suất có ít nhất:

a) 1 người bị sốt rét.

b) 1 người lách to.

c) 1 người sốt rét có lách to.

3) Khám ít nhất bao nhiêu người để có ít nhất 1 người có lách to và không sốt rét với xác suất ít nhất 90%?

Bài giải

Ta có: $P(SR) = 0,20$; $P(LT) = 0,30$; $P(LT|SR) = 0,80$

Do đó: $P(LT \& SR) = P(LT|SR)P(SR) = (0,80)(0,20) = 0,16$

	SR	$\overline{SR}$	
LT	0,16	0,14	0,30
$\overline{LT}$	0,04	0,66	0,70
	0,20	0,8	1.00

[1] $P(SR|LT) = \dfrac{P(LT|SR)P(SR)}{P(LT)} = \dfrac{(0,80)(0,20)}{0,30} = 0,53$

$P(\overline{SR}|LT) = 1 - P(SR|LT) = 0,47$

[2] a) Gọi A = biến cố "có ít nhất một người bị sốt rét".

$P(A) = 1 - P(\overline{A}) = 1 - (0.80)^{10} = 0,89$

b) Gọi B = biến cố "có ít nhất một người lách to".

$P(B) = 1 - P(\overline{B}) = 1 - (0,70)^{10} = 0,97$

c) Gọi C = biến cố "có ít nhất một người bị sốt rét có lách to".

$P(C) = 1 - P(\overline{C}) = 1 - (0,84)^{10} = 0,83$

[3] $P(LT \& \overline{SR}) = 0,14$

Gọi D = biến cố "có ít nhất một người lách to và không sốt rét".

$$P(D) = 1 - P(\overline{D}) = 1 - (0,86)^n \geq 0,90$$

$$1 - 0,90 \geq (0,86)^n \Rightarrow (0,86)^n \leq 0,10$$

$$n \ln 0,86 \leq \ln 0,10 \Rightarrow n \geq \frac{\ln 0,10}{\ln 0,86} = 15,26$$

Do đó khám ít nhất có 16 người.

Bài 12:

Tại một nệnh viện số bệnh nhân bị bệnh tim chiếm tỷ lệ 35%. Trong số đó, xác suất chọn một bệnh nhân có hút thuốc lá là 86%.

Chọn ngẫu nhiên một người trong bệnh viện này. Tính khả năng người này bị bệnh tim và không hút thuốc lá.

Bài giải

$$P(T) = 0,35\ ;\ P(H|T) = 0,86$$

$$P(\overline{H}T) = P(\overline{H}|T)P(T) = (1 - P(H|T)P(T)$$

Bài 13:

Một hồi cứu (Ducan 1986) về ung thư vú đã phẩu thuật cho biết: Tỷ lệ sống quá 5 năm là 90%, tỷ lệ có hạch di căn trong số những người sống quá 5 năm là 17%, tỷ lệ có hạch di căn trong số những người không số quá 5 năm là 72%.

Một người bị ung thư vú và không có hạch di căn, tính khả năng người này sống quá 5 năm sau khi phẩu thuật.

Bài giải

B = biến cố: "*có hạch di căn*".

Ta có: $P(A) = 0,90$

$$P(B|A) = 0,17 \Rightarrow P(\overline{B}|A) = 1 - 0,17 = 0,83$$

$$P(B|A = 0,72$$

$$P(A|\overline{B}) = \frac{P(\overline{B}|A)P(A)}{P(\overline{B})}$$

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

$$P(B) = (0,17)(0.90) + (0,72)(0,10) = 0,2250$$

$$P(\overline{B}) = 1 - P(B) = 0,7750$$

$$P(A|\overline{B}) = \frac{(0,83)(0,90)}{0,7750} = 0,96$$

Bài 14:

Giả sử biết tỷ lệ viên thuốc bị sứt mẻ của máy dập A là $p_A = 0,10$

1) Lấy ngẫu nhiên 5 viên từ máy dập đó. Tính xác suất có ít nhất 1 viên bị sứt mẻ.

2) Quan sát tối thiểu mấy viên để xác suất có ít nhất 1 viên bị sứt mẻ $\geq 0,95$.

Bài giải

[1] Gọi X là số viên bị sứt mẻ ở mẫu, ta có:

$$P(X \geq 1) = 1 - P(X) = 0) = 1 - (0,90)^5 = 0,41$$

[2] Gọi n là số lần quan sát, ta có:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,90)^n \geq 0,95$$

$$1 - 0,95 \geq (0,90)^n \Rightarrow (0,90)^n \leq 0,05$$

$$n.\ln 0,90 \leq \ln 0,05 \Rightarrow n \geq \frac{\ln 0,05}{\ln 0,90} = 28,43$$

$n \geq 29$ viên

Bài 15:

Trong một dân số tỷ lệ những người có dấu hiệu lách to là 20%, những người bị sốt rét là 23%, những người vừa sốt rét vừa lách to là 18%. Một người đến ngẫu nhiên từ dân số đó, người này không có dấu hiệu lách to. Tính khả năng người này bị sốt rét.

Bài giải

	SR	$\overline{SR}$	Cộng
LT	0,18	0,02	0,20
$\overline{LT}$	0,05	0,75	0,80
Cộng	0,23	0,77	1,00

$$P(SR|\overline{LT}) = \frac{P(SR, \overline{LT})}{P(\overline{LT})} = \frac{0,05}{0,80} = 0,0625$$

Bài 16:

Anh (Chị) đem giao một lô hàng, và biết rõ lô hàng của mình có tỷ lệ xấu là 10%. Người nhận hàng đề nghị lấy ngẫu nhiên 6 lọ để kiểm tra, nếu số lọ xấu quá k lọ thì từ chối. Anh (Chị) đề nghị k là bao nhiêu để vừa thuyết phục được người nhận, vừa hy vọng khả năng lô hàng không bị từ chối ít nhất là 95%.

Bài giải

$X \sim B(n = 6, p = 0,10)$

Hàm mật độ xác suất của X là:

$$f(x) = \begin{cases} C_6^x(0,10)^x(0,90)^{6-x} & ; (x = 0,1,\ldots,6) \\ 0 & ; \text{nơi khác} \end{cases}$$

$$P(X = 0) = (0,90)^6 = 0,5314$$

$$P(X = 1) = 6(0,10)(0,90)^5 = 0,3543$$

$$P(X = 2) = 15(0,10)^2(0,90)^4 = 0,0984$$

Ta có:

$$P(X \le 0) = P(X = 0) = 0,5314$$

$$P(X \le 1) = P(X = 0) + P(X = 1) = 0,8867$$

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0,9842 > 0,95$$

Vậy $k = 2$

Bài 17:

1) Chứng tỏ rằng: nếu 2 biến cố A, B độc lập thì $\overline{A}, \overline{B}$ cũng độc lập.

2) Có hai xét xét nghiệm T_1 và T_2: T_1 có độ nhạy 93% và độ chuyên 95%, T_2 dương giả 7% và âm giả 5%. T_1 dùng sàng lọc người có nguy cơ bị bệnh B; T_2 dùng chẩn đoán bệnh này trên những người mà T_1 cho kết quả dương tính.

Một người từ dân số có tỷ lệ bệnh B là 0,001, cho người này làm xét nghiệm T_1, kết quả dương tính, cho làm tiếp xét nghiệm T_2 cũng thấy dương tính. Tính khả năng người này mắc bệnh B.

Bài giải

[1] Vì A, B độc lập nên: $P(AB) = P(A)P(B)$

Ta có: $P(\overline{AB}) = 1 - P(A \cup B)$

$$P(\overline{AB}) = 1 - P(A) - P(B) + P(AB)$$
$$P(\overline{AB}) = 1 - P(A) - P(B) + P(A)P(B)$$
$$P(\overline{AB}) = (1 - P(A))(1 - P(B) = P(\overline{A})P(\overline{B})$$

Do đó $\overline{A}\ \overline{B}$ độc lập.

[2] $$P(T_1^+) = P(T_1^+ | \overline{B}^+)P(B^+) + P(T_1^+ | B^-)P(B^-)$$

$$P(T_1^+) = (0,93)(0,001) + (1 - 0,95)(1 - 0,001) = 0,05088$$

$$p = P(B^+ | T_1^+) = \frac{P(T_1^+ | B^+)P(B^+)}{P(T_1^+)}$$

$$p = P(B^+ | T_1^+) = \frac{(093)(0,001)}{0,05088} = 0,0183$$

Cho làm xét nghiệm T_2, kết quả T_2^+, vậy:

$$P(T_2^+) = P(T_2^+ | \overline{B}^+)P(B^+) + P(T_2^+ | B^-)P(B^-)$$

$$P(T_2^+) = 1 - 0,05)(0.0183) + (0,07)(1 - 0,0183) = 0,0861$$

$$P(B^+ | T_2^+) = \frac{P(T_2^+ | B^+)P(B^+)}{P(T_2^+)}$$

$$P(B^+ | T_2^+) = \frac{(1 - 0,05)(0,0183)}{0,0681} = 0,2017$$

Bài 18:

Người ta tổng kết về các phương pháp chẩn đoán bệnh dạ dày tá tràng. Trên lâm sàng chẩn đoán đúng 60%, X quang 70%, nội soi 80%. Kết hợp cả ba phương pháp thì khả năng chẩn đoán đúng là bao nhiêu?

Bài giải

Gọi: A = biến cố: "lâm sàng chẩn đoán đúng".

B = Biến cố: "X quang chẩn đoán đúng".

C = Biến cố: "nội soi chẩn đoán đúng".

$$\begin{aligned} P(A \cup B \cup C) &= 1 - P(\overline{A}\ \overline{B}\ \overline{C}) \\ &= 1 - (0,40)(0,30)(0,20) = 0976 \end{aligned}$$

Bài 19:

Một hộp thuốc có 5 lọ, gồm 4 lọ tốt và 1 lọ hỏng.

1) Lấy ngẫu nhiên 3 lọ. Tính xác suất được 3 lọ tốt? được 2 lọ tốt và 1 lọ hỏng?

2) Kiểm tra từng lọ (không hoàn lại) cho đến khi phát hiện lọ hỏng thì dừng. Gọi X là số lần kiểm tra. Tìm hàm mật độ của X. Tính $\mu(X)$ và $\sigma^2(X)$.

Bài giải

[1] $P(3T) = \dfrac{C_4^3}{C_5^3} = \dfrac{4}{10} = 0,40$

$$P(2T, 1H) = \frac{C_4^2 C_1^1}{C_5^3} = \frac{6 \times 1}{10} = 0,60$$

[2] Gọi X = số lần kiểm tra, (X = 1, 2, 3, 4, 5)

$$P(X=1) = P(H) \qquad\qquad = \frac{1}{5}$$

$$P(X=2) = P(TH) = \frac{4}{5}\cdot\frac{1}{4} = \frac{1}{5}$$

$$P(X=3) = P(TTH) = \frac{4}{5}\cdot\frac{3}{4}\cdot\frac{1}{3} = \frac{1}{5}$$

$$P(X=4) = P(TTTH) = \frac{4}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} = \frac{1}{5}$$

$$P(X=5) = P(TH) = \frac{4}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2}\cdot\frac{1}{1} = \frac{1}{5}$$

X	1	2	3	4	5
P	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$$\mu = \sum x \cdot f(x)$$

$$\mu = 1(\frac{1}{5}) + 2(\frac{1}{5}) + 3(\frac{1}{5}) + 4(\frac{1}{5}) + 5(\frac{1}{5}) = 3$$

$$\sigma^2 = \sum x^2 f(x) - \mu^2$$

$$\sigma^2 = 1^2(\frac{1}{5}) + 2^2(\frac{1}{5}) + 3^2(\frac{1}{5}) + 4^2(\frac{1}{5}) + 5^2(\frac{1}{5}) - 3^2 = 2$$

Bài 20:

Tỷ lệ thuốc hỏng ở lô A là $p_A = 0,10$; lô B là $p_B = 0,08$; và lô C là $p_C = 0,15$. Giả sử các lô có rất nhiều lọ.

1) Lấy 3 lọ ở lô A. Tính xác suất có ít nhất 1 lọ hỏng. Lấy tối thiểu mấy lọ (ở lô A) để xác suất có ít nhất 1 lọ hỏng $\geq 0,95$.

2) Chọn một trong 3 lô rồi lấy từ đó ra 3 lọ. Tính xác suất có ít nhất 1 lọ hỏng.

3) Lấy ở mỗi lô 1 lọ. Gọi X là số lọ hỏng trong 3 lọ lấy ra. Tìm hàm mật độ của X.

4) Cửa hàng nhận 500 lọ ở lô A, 300 lọ ở lô B, 200 lọ ở lô C. Ta mua ở cửa hàng 1 lọ về dùng. Tính xác suất được lọ tốt.

Bài giải

Gọi X = số lọ hỏng trong các lọ lấy ra.

[1] $P(X \geq 1) = 1 - P(X = 0) = 1 - (0,90)^3 = 0,271$

Gọi n = số lọ lấy ra

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,90)^n \geq 0,95$$

$$1 - 0,95 \geq (0,90)^n \Rightarrow (0,90)^n \leq 0,05$$

$$n \cdot \ln 0,90 \leq 0,05 \Rightarrow \geq \frac{\ln 0,05}{\ln 0,90} = 28,43$$

Vậy phải quan sát ít nhất 29 lọ.

[2] $P(X \geq 1) = 1 - P(X) = 0$

$$P(X = 0) = P(X = 0|A)P(A) + P(X = 0|B)P(B) + P(X = 0|C)P(C)$$

$$P(X = 0) = \frac{1}{3}(0,90)^3 + \frac{1}{3}(0,92)^3 + \frac{1}{3}(0,85)^3$$

$$P(X = 0) = 0,7073$$

$$P(X \geq 1) = 1 - P(X = 0) = 0,2927$$

[3] Nếu lấy ở mỗi lô một lọ, và gọi X = số lọ hỏng trong các lọ lấy ra, ta có:

$$P(X = 0) = P(T_A T_B T_C) = (0,90)(0,92)(0,85) = 0,7038$$

$$P(X=1) = (0,10)(0,92)(0,85) + (0,90)(0,08)(0,85) + $$
$$+ (0,90)(0,92)(0,15) = 0,2636$$

$$P(X=2) = P(H_A B_B T_C) + P(H_A T_B H_C) + P(T_A H_B H_C)$$

$$P(X=2) = (0,10)(0,08)(0,85) + (0,10)(0,92)(0,15) +$$
$$+ (0,90)(0,08)(0,15) = 0,0314$$

$$P(X=3) = P(H_A H_B H_C) = (0,10)(0,08)(0,15) = 0,0012$$

X	0	1	2	3	
P	0,7038	0,2636	0,0314	0,0012	1

[4] $P(T) = P(T|A)P(A) + P(T|B)P(B) + P(T|C)P(C)$

$$P(T) = (0,90)\frac{500}{1000} + (0,92)\frac{300}{200} + (0.85)\frac{200}{1000} = 0,896$$

Bài 21:

Hộp A có 15 lọ thuốc = 3 lọ hỏng + 12 lọ tốt

Hộp B có 15 lọ thuốc = 4 lọ hỏng + 11 lọ tốt

Hộp C có 15 lọ thuốc = 5 lọ hỏng + 10 lọ tốt

1) Chọn ngẫu nhiên 1 trong 3 hộp, rồi lấy từ đó ra 3 lọ. Tính xác suất được 3 lọ tốt? được 2 lọ tốt và 1 lọ hỏng?

2) Kiểm tra từng lọ ở hộp A (không hoàn lại) cho đến khi phát hiện đủ 3 lọ hỏng thì dừng. Tính xác suất để việc kiểm tra dừng lại ở lần thứ 5.

3) Lấy ở mỗi hộp 1 lọ. Gọi X là số lọ hỏng trong 3 lọ lấy ra. Tìm hàm mật độ của X.

Bài giải

[1] Chọn một trong 3 hộp, vậy:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Gọi D = biến cố: "lấy được 3 lọ tốt".

E = biến cố: lấy được 2 lọ tốt, 1 lọ hỏng".

$$P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)$$

$$P(D) = \left(\frac{C_{12}^3}{C_{15}^3}\right)\frac{1}{3} + \left(\frac{C_{11}^3}{C_{15}^3}\right)\frac{1}{3} + \left(\frac{C_{10}^3}{C_{15}^3}\right)\frac{1}{3}$$

$$P(D) = \left(\frac{220}{455}\right)\frac{1}{3} + \left(\frac{165}{455}\right)\frac{1}{3} + \left(\frac{120}{455}\right)\frac{1}{3} = 0,37$$

$$P(E) = P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C)$$

$$P(E) = \left(\frac{C_{12}^2C_3^1}{C_{15}^3}\right)\frac{1}{3} + \left(\frac{C_{11}^2C_4^1}{C_{15}^3}\right)\frac{1}{3} + \left(\frac{C_{10}^2C_5^1}{C_{15}^3}\right)\frac{1}{3}$$

$$P(E) = \left(\frac{66\times 3}{455}\right)\frac{1}{3} + \left(\frac{55\times 4}{455}\right)\frac{1}{3} + \left(\frac{45\times 5}{455}\right)\frac{1}{3} = 0,47$$

[2] Gọi Y = số lần kiểm tra

$$P(Y = 5) = C_4^2\left(\frac{12}{15}\cdot\frac{11}{14}\cdot\frac{3}{13}\cdot\frac{2}{12}\cdot\frac{1}{11}\right) = 0,0132$$

[3] Hàm mật độ xác suất của X:

$$P(X = 0) = P(T_AT_BT_C) = \frac{12}{15}\times\frac{11}{15}\times\frac{10}{15} = 0,39$$

$$P(X = 1) = P(H_AT_BT_C) + P(T_AH_BT_C) + P(T_AT_BH_C)$$

$$P(X=1)=\left(\frac{3}{15}\times\frac{11}{15}\times\frac{10}{15}\right)+\left(\frac{12}{15}\times\frac{4}{15}\times\frac{10}{15}\right)+\left(\frac{12}{15}\times\frac{11}{15}\times\frac{5}{15}\right)$$

$$P(X=1)=0,44$$

$$P(X=2)=P(H_AH_BT_C)+P(H_AT_BH_C)+P(T_AH_BH_C)$$

$$P(X=2)=\left(\frac{3}{15}\times\frac{4}{15}\times\frac{10}{15}\right)+\left(\frac{3}{15}\times\frac{11}{15}\times\frac{5}{15}\right)+\left(\frac{12}{15}\times\frac{4}{15}\times\frac{5}{15}\right)$$

$$P(X=2)=0,16$$

$$P(X=3)=P(H_AH_BH_C)=\frac{3}{15}\times\frac{4}{15}\times\frac{5}{15}=0,02$$

Bài 22:

Tỉ lệ sốt rét trong dân số là 20% và lách to là 30%; trong số người bị sốt rét thì tỷ lệ lách to chiếm 80%.

1) Chọn ngẫu nhiên một người trong địa phương đó, thấy người này có lách to, tính khả năng người này không bị sốt rét.

2) Khám 10 người tại địa phương đó, tính xác suất có ít nhất:

   a) 1 người bị sốt rét.

   b) 1 người lách to.

   c) 1 người sốt rét có lách to.

3) Khám ít nhất bao nhiêu người để có ít nhất 1 người có lách to và không sốt rét với xác suất ít nhất 90%?

Bài giải

Ta có: $P(SR)=0,20$; $P(LT)=0,30$; $P(LT|SR)=0,80$

Do đó:

$$P(LT \,\&\, SR) = P(LT|SR)P(SR) = (0,80)(0,20) = 0,16$$

	SR	$\overline{SR}$	
LT	0,16	0,14	0,30
$\overline{LT}$	0,04	0,66	0,70
	0,20	0,8	1.00

[1] $$P(SR|LT) = \frac{P(LT|SR)P(SR)}{P(LT)} = \frac{(0,80)(0,20)}{0,30} = 0,53$$

$$P(\overline{SR}|LT) = 1 - P(SR|LT) = 0,47$$

[2] a) Gọi A = biến cố "có ít nhất một người bị sốt rét".

$$P(A) = 1 - P(\overline{A}) = 1 - (0,80)^{10} = 0,89$$

b) Gọi B = biến cố "có ít nhất một người lách to".

$$P(B) = 1 - P(\overline{B}) = 1 - (0,70)^{10} = 0,97$$

c) Gọi C = biến cố "có ít nhất một người bị sốt rét có lách to".

$$P(C) = 1 - P(\overline{C}) = 1 - (0,84)^{10} = 0,83$$

[3] $$P(LT \,\&\, \overline{SR}) = 0,14$$

Gọi D = biến cố "có ít nhất một người lách to và không sốt rét".

$$P(D) = 1 - P(\overline{D}) = 1 - (0.86)^n \geq 0,90$$

$$1 - 0,90 \geq (0,86)^n \Rightarrow (0,86)^n \leq 0,10$$

$$n \ln 0,86 \leq \ln 0,10 \Rightarrow n \geq \frac{\ln 0,10}{\ln 0,86} = 15,26$$

Do đó khám ít nhất có 16 người.

Bài 23:

Ba lô thuốc A, B, C gồm rất nhiều lọ, tỷ lệ hỏng ở mỗi lô lần lượt là:

$p_A = 0{,}10$; $p_B = 0{,}08$; $p_C = 0{,}05$

1) Lấy ở mỗi lô 1 lọ. Tính xác suất được 2 lọ tốt và 1 lọ hỏng.

2) Chọn 1 trong 3 lô rồi lấy từ đó ra 3 lọ. Tính xác suất được 2 lọ tốt và 1 lọ hỏng.

3) Lấy 5 lọ từ lô B. Tính xác suất có ít nhất 1 lọ hỏng. Lấy tối thiểu mấy lọ (ở lô B) để xác suất có ít nhất 1 lọ hỏng $\geq 0{,}90$?

4) Kiểm tra từng lọ ở lô B cho đến khi phát hiện được 2 lọ hỏng thì dừng. Tính xác suất để việc kiểm tra dừng lại ở lần thứ 10.

Bài giải

Gọi D là biến cố "lấy được 2 lọ tốt, 1 lọ hỏng".

[1] $P(D) = P(T_A T_B H_C) + P(T_A H_B T_C) + P(H_A T_B T_C)$

$$P(D) = (0,90)(0,92)(0,05) + (0,90)(0,08)(0,95) + $$
$$+ (0,10)(0,92)(0,95) \approx 0,20$$

[2] $P(D) = P(D|A).P(A) + P(D|B).P(B) + P(D|C).P(C)$

$$P(D) = C_3^2(0,09)^2(0,10) \times \frac{1}{3} + C_3^2(0,92)^2(0,80) \times$$
$$\times \frac{1}{3} C_3^2(0,95)^2(0,05) \times \frac{1}{3} \approx 0,19$$

[3] Gọi X = số lọ hỏng trong các lọ lấy ra:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,92)^5 = 0,34$$

Gọi n = số lọ lấy ra, ta có:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,92)^n \geq 0,90$$

$$1 - 0,90 \geq (0,92)^n \quad \Rightarrow (0,92)^n \leq 0,10$$

$$n \cdot \ln 0,92 \leq \ln 0,10 \Rightarrow n \geq \frac{\ln 0,10}{\ln 0,92} = 27,62$$

Vậy ta lấy ít nhất 28 lọ.

[4] Gọi Y = số lần kiểm tra.

$$P(Y = 10) = 9(0,92)^8(0,08)^2 = 0,03$$

Bài 24:

Tỷ lệ thuốc hỏng trong các lô thuốc A, B lần lượt là 0,10 và 0,07. Giả sử các lô thuốc này có rất nhiều lọ.

1) Lấy ngẫu nhiên 3 lọ ở lô thuốc A. Tính xác suất có ít nhất 1 lọ thuốc hỏng. Lấy tối thiểu mấy lọ để xác xuất có ít nhất 1 lọ hỏng ≥ 0,90?

2) Lấy ngẫu nhiên ở mỗi lô 1 lọ. Tính xác suất được 1 lọ tốt và 1 lọ hỏng.

3) Chọn ngẫu nhiên một trong hai lô rồi lấy từ đó ra 2 lọ. Tính xác suất được một lọ tốt và một lọ hỏng.

Bài giải

[1] Gọi X = số lọ hỏng lấy ra:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,90)^3 = 0,271 = 27,10\%$$

Gọi n = số lọ lấy ra:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,90)^n \geq 0,90$$

$$1 - 0,90 \geq (0,90)^n \Rightarrow (0,90^n \leq 0,10$$

$$n \cdot \ln(0,90) \leq \ln(0,10)$$

$$n \geq \frac{\ln 0,10}{0,90} = 21,85$$

$$n \geq 22 \text{ lọ}$$

[2] $P(T,H) = P(T_A H_B) + P(H_A T_B)$

$$= (0,90)(0,07) + (0,10)(0,93) = 0,1560$$

[3] Chọn ngẫu nhiên một lô, vậy:

$$P(A) = P(B) = \frac{1}{2}$$

Gọi C = biến cố "được một lọ tốt, một lọ hỏng".

$$P(C) = P(C|A)P(A) + P(C|B)P(B)$$

$$P(C) = 2(0,90)(0,10)\frac{1}{2} + 2(0,07)(0,93)\frac{1}{2} = 0,1551$$

Bài 25:

Cho biết tỷ lệ bệnh sốt rét tại một địa phương là $p = 0,08$

1) Khám ngẫu nhiên 3 người, tính xác suất để có ít nhất 1 người mắc bệnh sốt rét.

2) Khám tối thiểu mấy người để xác suất có ít nhất một người mắc bệnh $\geq 0,90$,

3) Dùng 3 loại thuốc A, B, C để điều trị. Tỷ lệ khỏi bệnh khi dùng từng loại thuốc để điều trị lần lượt là: 0,85;

0,90; 0,95. Nếu dùng cả 3 loại thuốc phối hợp điều trị thì tỷ lệ khỏi bệnh là bao nhiêu? (Bỏ qua sự tương tác giữa các loại thuốc).

Bài giải

[1] $P(X \geq 1)1 - P(X = 0) = 1 - (0,92)^3 = 0,22$

[2] Gọi n là số người được khám bệnh.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,92)^n \geq 0,90$$

$$1 - 0,90 \geq (0,92)^n \Rightarrow (0,92)^n \leq 0,10$$

$$n \ln 0,92 \leq \ln 0,10 \Rightarrow n \geq \frac{\ln 0,10}{\ln 0,92} = 27,61$$

$n \geq 28$ người.

[3] Gọi A, B, C lần lượt là các biến cố khỏi bệnh khi dùng thuốc A, B, C.

$$P(\text{Khỏi}) = P(A \cup B \cup C) = 1 - P(\overline{A} \cdot \overline{B} \cdot \overline{C})$$
$$P(\text{Khỏi}) = 1 - (0,15)(0,10)(0,05) = 0,99925$$

Bài 26:

Hộp A có 10 lọ thuốc = 8 lọ tốt + 2 lọ hỏng

Hộp B có 15 lọ thuốc = 11 lọ tốt + 4 lọ hỏng

Hộp C có 20 lọ thuốc = 15 lọ tốt + 5 lọ hỏng

1) Chọn ngẫu nhiên một hộp, rồi lấy ra một lọ. Tính xác suất được lọ tốt?

2) Lấy ngẫu nhiên ở mỗi hộp một lọ. Tính xác suất được 1 lọ hỏng và 2 lọ tốt.

3) Kiểm tra từng lọ (không hoàn lại) ở hộp A cho đến khi phát hiện được 2 lọ hỏng thì dừng. Tính xác suất để việc kiểm tra dừng lại ở lần thứ 4.

Bài giải

[1] Chọn một trong một hộp, vậy:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

$$P(T) = P(T|A)P(A) + P(T|B)P(B) + P(T|C)P(C)$$

$$P(T) = \left(\frac{8}{10}\right)\frac{1}{3} + \left(\frac{11}{15}\right)\frac{1}{3} + \left(\frac{15}{20}\right)\frac{1}{3} = 0,76$$

[2] Lấy ở mỗi hộp 1 lọ:

$$P(H, 2T) = P(H_A T_B T_C) + P(T_A H_B T_C) + P(T_A T_B H_C)$$

$$P(H, 2T) = \frac{2}{10}\cdot\frac{11}{15}\cdot\frac{15}{20} + \frac{8}{10}\cdot\frac{4}{15}\cdot\frac{15}{20} + \frac{8}{10}\cdot\frac{11}{15}\cdot\frac{5}{20} = 0,42$$

[3] Kiểm tra từng lọ ở hộp A (không hoàn lại) cho đến khi phát hiện được 2 lọ hỏng thì dừng. Gọi Y = số lần kiểm tra. Ta có:

$$P(Y = 4) = P(TTHH) + P(THTH) + P(HTTH)$$

$$P(Y = 4) = 3\left(\frac{8}{10}\cdot\frac{7}{9}\cdot\frac{2}{8}\cdot\frac{1}{7}\right) = 0,07$$

Bài 27:

Giả sử biết tỷ lệ viên thuốc bị sứt mẻ của máy dập A là p = 0,10.

1) Lấy ngẫu nhiên 5 viên từ máy dập A. Tính xác suất có ít nhất 1 viên bị sứt mẻ?

2) Quan sát tối thiểu mấy viên để xác suất có ít nhất 1 viên bị sứt mẻ $\geq 0,95$

Bài giải

[1] Gọi X = số viên bi sứt mẻ.

$$P(X \geq 1) = 1 - P(X) = 0) = 1 - (0,90)^5 = 0,41$$

[2] Gọi n là số lần quan sát, ta có:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,90)^n \geq 0,95$$

$$1 - \ 0,95 \geq (0,90)^n \Rightarrow (0,90)^n \leq 0,05$$

$$n.\ln 0,90 \leq \ln 0,05 \Rightarrow n \geq \frac{\ln 0,05}{\ln 0,90} = 28,43$$

$n \geq 29$ lần

Bài 28:

Hộp A = 15 lọ tốt + 5 lọ hỏng = 20 lọ.

Hộp B = 17 lọ tốt + 3 lọ hỏng = 20 lọ.

Hộp C = 10 lọ tốt + 10 lọ hỏng = 20 lọ.

1) Lấy ở mỗi hộp 1 lọ. Tính xác suất được 2 lọ tốt và 1 lọ hỏng.

2) Chọn 1 trong 3 hộp rồi lấy từ đó ra 3 lọ. Tính xác suất được 2 lọ tốt và 1 lọ hỏng.

3) Trộn chung 3 hộp rồi lấy từ đó ra 3 lọ. Tính xác suất được 2 lọ tốt và 1 lọ hỏng.

4) Kiểm tra từng lọ ở hộp B cho đến khi phát hiện đủ 3 lọ hỏng thì dừng. Tính xác suất để việc kiểm tra dừng lại ở lần thứ 5.

Bài giải

Gọi D là biến cố lấy được 2 lọ tốt và 1 lọ hỏng.

[1] Lấy ở mỗi hộp 1 lọ.

$$P(D) = P(T_AT_BH_C) + P(T_AH_BT_C) + P(H_AT_BT_C)$$

$$P(D) = \frac{15}{20}\cdot\frac{17}{20}\cdot\frac{10}{20} + \frac{15}{20}\cdot\frac{3}{20}\cdot\frac{10}{20} + \frac{5}{20}\cdot\frac{17}{20}\cdot\frac{10}{20}$$

$$P(D) = 0.48$$

[2] Chọn ngẫu nhiên 1 hộp rồi lấy ra 3 lọ:

$$P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)$$

$$P(D) = \left(\frac{C_{15}^2C_5^1}{C_{20}^3}\right)\frac{1}{3} + \left(\frac{C_{17}^2C_3^1}{C_{20}^3}\right)\frac{1}{3} + \left(\frac{C_{10}^2C_{10}^1}{C_{20}^3}\right)\frac{1}{3}$$

$$P(D) = \left(\frac{525}{1140}\right)\frac{1}{3} + \left(\frac{408}{1140}\right)\frac{1}{3} + \left(\frac{450}{1140}\right)\frac{1}{3}$$

$$P(D) = \frac{1383}{3420} = 0,40$$

[3] Trộn chung 3 hộp lại ta được 60 lọ = 42 lọ tốt + 18 lọ hỏng

$$P(D) = \frac{C_{42}^2C_{18}^1}{C_{60}^3} = \frac{15498}{34220} = 0,45$$

[4] Ta kiểm tra từng lọ ở hộp B.

Gọi Y = số lần kiểm tra để phát hiện 3 lọ hỏng.

$$P(Y = 5) = C_4^2\left(\frac{17}{20}\cdot\frac{16}{19}\cdot\frac{3}{18}\cdot\frac{2}{17}\cdot\frac{1}{16}\right) = 0,0053$$

Bài 29:

Có tài liệu cho biết tỷ lệ K phổi là p = 0,07.

1) Khám ngẫu nhiên 5 người. Tính xác suất có ít nhất 1 trường hợp K phổi. Khám tối thiểu mấy người để xác suất có ít nhất 1 trường hợp K phổi ≥ 0,80.

2) Khả năng kháng thuốc của vi trùng đối với từng loại thuốc A, B, C lần lượt là 0,10, 0,15, 0,12. Nếu dùng cả 3 loại thuốc để diệt vi trùng, hãy tính xác suất vi trùng bị diệt (bỏ qua sự tương tác của các loại thuốc).

Bài giải

[1] Tỷ lệ K phổi là: p = 0,70.

Vậy tỷ lệ không K phổi là: q = 1 – 0,07 = 0,93.

Gọi X = số trường hợp K phổi.

$$P(X \ge 1) = 1 - P(X = 0)$$
$$= 1 - (0,93)^5 = 0,30$$

Gọi n là số người được khám, ta có:

$$P(X \ge 1) = 1 - P(X = 0)$$
$$= 1 - (0,93)^n \ge 0,80$$

$$1 - 0,80 \ge (0,93)^n$$

$$(0,90)^n \le 0,20 \Rightarrow n \ln 0,93 \le \ln 0,20$$

$$n \ge \frac{\ln 0,20}{\ln 0,93} = 22,17$$

Vậy khám ít nhất 23 người.

[2] Gọi A, B, C lần lượt là biến cố vi trùng kháng được thuốc

A, B, C và K là biến cố vi trùng không bị diệt.

$$P(K) = P(ABC) = P(A)P(B)P(C)$$

$$P(K) = (0,10)(0,15)(0,12) = 0,0018$$

Xác suất vi trùng bị diệt:

$$P(\overline{K}) = 1 - P(K) = 1 - 0,0018 = 0,9982$$

Bài 30:

Một người đến khám vì ho ra máu. Theo tổng kết của phòng khám thì những bệnh nhân như vậy có thể do: lao phổi 40%, K phổi 30%, dãn phế quản 20%, còn lại là do các bệnh khác.

1) Cho làm xét nghiệm K(IDR), thấy người này có kết quả dương tính. Theo tổng kết của phòng xét nghiệm thì tỷ lệ IDR dương tính trong các bệnh trên thứ tự là: 0,80; 0,40; 0,20; 0,10. Tính khả năng người này bị lao phổi.

2) Cho làm xét nghiệm H (tìm BK), kết quả âm tính, nhưng phòng xét nghiệm cũng cho biết là với bệnh nhân có lao phổi thì H dương tính 80%, với bệnh nhân không lao phổi H cũng dương tính (giả) 10%. Tính khả năng người này bị lao phổi.

3) Nếu dùng ba loại thuốc A, B, C riêng rẽ để điều trị lao phổi thì tỷ lệ kháng thuốc của BK theo thứ tự là 15%, 20%, 25%. Dùng chung cả 3 loại thuốc thì khả năng kháng thuốc của BK là bao nhiêu?

4) Dùng riêng rẽ ba loại thuốc trên để điều trị thì tỷ lệ khỏi bệnh lần lượt là: 90%, 80%, 70%. Dùng phối hợp cả ba loại thì khả năng khỏi bệnh là bao nhiêu? (Bỏ qua sự tương tác giữa các loại thuốc)

Bài giải

Δ_1 = lao phổi; $P(\Delta_1) = 0{,}40$

Δ_2 = K–phổi; $P(\Delta_2) = 0{,}30$

Δ_3 = dãn phế quản; $P(\Delta_3) = 0{,}20$

Δ_4 = bệnh khác; $P(\Delta_4) = 0{,}10$

K = biến cố "IDR dương tính"

Bệnh	$P(\Delta_n)$	$P(K\|\Delta_n)$	$P(K\|\Delta_n)P(\Delta_n)$
Δ_1	0,40	0,80	0,32
Δ_2	0,30	0,40	0,12
Δ_3	0,20	0,20	0,04
Δ_4	0,10	0,10	0,01
			$P(K) = 0{,}49$

[1] $$P(\Delta_1|K) = \frac{P(K|\Delta_1)P(\Delta_1)}{P(K)}$$

$$P(\Delta_1|K) = \frac{(0,80)(0,04)}{0,49} = 0,65$$

[2] Gọi $\overline{H}$ là biến cố "BK âm tính".

$$P(H) = P(H|\Delta_1)P(\Delta_1) + P(H|\Delta_2)P(\Delta_2)$$

$$= (0,80)(0,40) + (0,10)(0,60) = 0,38$$

$$P(\overline{H}) = 1 - P(H) = 0,62$$

$$P(\Delta_1|\overline{H}) = \frac{P(\overline{H}|\Delta_1)}{P(\overline{H})} = \frac{(0,20)(0,40)}{0,62} = 0,13$$

[3] Gọi A, B, C lần lượt là biến cố BK kháng được thuốc A, thuốc B, thuốc C, ta có BK kháng được thuốc nghĩa là kháng được cả 3 loại thuốc A, B, C. Khả năng kháng thuốc của BK là:

$P(ABC) = P(A)P(B)P(C)$

$P(ABC) = (0,15)(0,20)(0,25) = 0,0075$

[4] Gọi A, B, C lần lượt là biến cố lành bệnh khi dùng riêng rẽ từng loại thuốc. Nếu dùng phối hợp thì khả năng lành bệnh là:

$$P(A \cup B \cup C) = 1 - P(\overline{A})P(\overline{B})P(\overline{C})$$
$$P(A \cup B \cup C) = 1 - (0,10)(0,20)(0,30) = 0,994$$

Bài 31:

Có một lô thuốc mà người giao hàng cho biết tỷ lệ xấu là 5%.

1) Lấy ra tối thiểu mấy lọ để xác suất có ít nhất 1 lọ xấu $\geq 0,90$.

2) Lấy ra ngẫu nhiên 20 lọ để kiểm tra thấy có 4 lọ xấu. Hỏi: lời nói của người giao hàng có đúng không? (Kết luận với $\alpha = 5\%$)

Bài giải

Gọi X = số lọ xấu trong n lọ lấy ra.

[1] $P(X \geq 1) = 1 - P(X = 0) = 1 - (0,95)^n \geq 0,90$

$$1 - 0,90 \geq (0,95)^n \Rightarrow (0,95)^n \leq 0,10$$

$$n\ln(0,95) \leq \ln 0,10 \Rightarrow \frac{\ln 0,10}{\ln 0,95} = 44,89$$

$n \geq 45$ lọ

Bài 32:

Một người đến khám vì sốt. Theo kinh nghiệm của BS trực đánh giá người này bị cúm là 40%, sốt rét 30%, thương hàn 10%, hoặc bệnh khác. Cho người này đi làm xét nghiệm máu thấy bạch cầu không tăng. Theo tổng kết của phòng xét nghiệm tại bệnh viện ấy, tỷ lệ bạch cầu tăng trong các bệnh trên theo thứ tự là 0,50; 0,40; 0,10 và 0,80.

Tính khả năng người này bị thương hàn?

Bài giải

Gọi: Δ_1 = Cúm ; $P(\Delta_1) = 0{,}40$

Δ_2 = Sốt rét ; $P(\Delta_2) = 0{,}30$

Δ_3 = Thương hàn ; $P(\Delta_3) = 0{,}10$

Δ_4 = Bệnh khác ; $P(\Delta_4) = 0{,}20$

Gọi K^+ = biến cố "bạch cầu tăng"

$$P(H^+) = \sum P(K^+|\Delta_n)P(\Delta_n)$$

$$= (0,5)(0,4) + (0,4)(0,3) + (0,1)(0,1) + (0,8)(0,2)$$

$$P(K^-) = 1 - P(K^+) = 1 - 0,49 = 0,51$$

$$P(\Delta_3|K^-) = \frac{P(K^-|\Delta_3)P(\Delta_3)}{P(K^-)}$$

$$= \frac{(0,9)(0,1)}{0,51} = 0,1765 = 17,65\%$$

Bài 33:

Người giao hàng cho biết là lô thuốc này có 10% số lọ xấu. Để kiểm tra ta lấy ngẫu nhiên 5 lọ.

1) Tính xác suất để được 3 lọ xấu.

2) Quả thật khi kiểm tra ta thấy có 3 lọ xấu. Nghĩ gì về tỷ lệ lọ xấu mà người giao hàng cho biết?

Bài giải

Gọi X = số lọ xấu trong các lọ lấy ra.

[1] $P(X = 3) = C_5^3(0,10)^3(0,90)^2 = 0,0081$

[2] Biến cố X là biến cố hiếm khi xảy ra. Theo nguyên lý của xác suất nhỏ, biến cố này không xảy ra khi ta quan sát một đôi lần.

Vậy mà khi ta quan sát 5 lọ thấy có 3 lọ hỏng, vậy tỷ lệ lọ hỏng mà người giao hàng nói là không đúng.

Bài 34:

Đo thị lực trên 200 sinh viên ở một trường Đại học, kết quả như sau:

Thị lực X	5/10	6/10	7/10	8/10	9/10	10/10
Số SV	10	16	50	60	50	14

1) Lập bảng phân phối tần suất của X?

2) Vẽ nhật đồ.

3) Một sinh viên thuộc trường Đại học ấy muốn được chấp thuận vào làm việc tại phòng xét nghiệm thì yêu cầu thị lực ít nhất là 8/10. Tính khả năng sinh viên này được

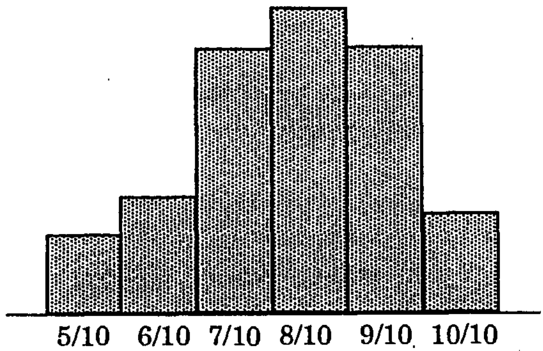
chấp thuận.

Bài giải

[1]

X	5/10	6/10	7/10	8/10	9/10	10/10
P	0,05	0,08	0,25	0,30	0,25	0,07

[2] Vẽ nhật đồ:



[3] $P(X \geq 8/10) = 0,30 + 0,25 + 0,07 = 0,62$

II. CÁC PHÂN PHỐI THƯỜNG DÙNG

A. BÀI TẬP

1) Một bình chứa 1 bi đỏ, 1 bi xanh, 2 bi trắng. Rút ngẫu nhiên 1 bi, nếu bi này đỏ thì được 3đ, xanh ta thua 2đ, trắng thua 1 đ. Ta rút 2 lần có hoàn lại, gọi X là tiền thu, được sau hai lần rút.

   a) Tìm phân phối xác suất của X, vẽ đồ thị hình que.

   b) Tính E(X), Var(X).

2) Gọi X là số con trai trong gia đình có 4 con, xác suất sanh trai là 0,52.

   a) Đồ thị phân phối của X.

   b) Tính E(X), var(X).

3) Một bài thi gồm 6 câu hỏi, mỗi câu có 5 cách trả lời trong đó có 1 cách đúng. Muốn đạt thí sinh phải trả lời đúng ít nhất 4 câu.

   a) Tính xác suất để 1 thí sinh không biết gì cả mà vẫn trả lời đúng 6 câu.

   b) Tính xác suất để 1 thí sinh như vậy mà lại đạt.

   c) Tính xác suất đạt của 1 thí sinh biết trả lời đúng 3 câu đầu.

   d) Tính xác suất đạt của 1 thí sinh bao giờ cũng có thể trả lời đúng 3 câu trong loại bài thi như vậy.

4) Một xí nghiệp sản xuất thuốc cho biết có 10% số chai không đúng tiêu chuẩn lấy 10 chai tính xác suất để:

   a) Có 1 chai không đúng tiêu chuẩn.

b) Ít nhất 1 chai không đúng tiêu chuẩn.

c) Nhiều nhất 1 chai không đúng tiêu chuẩn.

5) Một máy sản xuất sản phẩm với tỷ lệ phế phẩm là 7%.

a) Quan sát ngẫu nhiên 10 sản phẩm. Tính xác suất:

- Có 1 sản phẩm hỏng?
- Có ít nhất 1 sản phẩm hỏng?
- Có nhiều nhất 1 sản phẩm hỏng?

b) Quan sát tối thiểu mấy sản phẩm để xác suất có ít nhất 1 sản phẩm hỏng $\geq 0{,}90$.

6) Một trung tâm bưu điện nhận trung bình 3 cuộc điện thoại trong mỗi phút.

Tính xác suất để trung tâm này nhận được 1 lần? 2 lần? 3 lần gọi trong 1 phút, biết rằng số lần gọi điện thoại trong 1 phút $X \sim$ Poission.

7) Khi tiêm truyền một loại huyết thanh trung bình có một trường hợp bị phản ứng trên 1000. Ta lại dùng huyết thanh trên tiêm cho 2000 người. Tính xác suất để:

a) Có 3 ca bị phản ứng.

b) Nhiều nhất 3 ca bị phản ứng.

c) Hơn 3 ca bị phản ứng.

8) Tỷ lệ một bệnh bẩm sinh trong dân số là 1%.

Bệnh này cần được chăm sóc đặc biệt ngay từ lúc mới sinh. Một nhà bảo sinh thường có 20 ca sinh trong 1 tuần lễ. Tính xác suất để:

a) Không có ca nào cần sự chăm sóc đó.

b) Có 1 trường hợp cần.

c) Có hơn 1 trường hợp cần.

Tính bằng quy luật nhị thức, rồi bằng quy luật Poisson, so sánh kết quả.

9) Cho $X \approx B(n, P)$ với $E(X) = 2$, $Var(X) = \frac{4}{3}$. Tìm hàm mật độ.

10) Cho biến số ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & ; x > 0 ; \lambda > 0 \\ 0 & ; \text{nơi khác} \end{cases}$$

a) Tính trung bình μ và phương sai σ^2.

b) Tìm hàm gây moment M(t). Dùng hàm gây Moment để tính lại trung bình μ và phương sai σ^2.

11) Cho $X \sim N(0, 1)$: Tính xác suất:

$$P(0 < x < 1,42) ; P(-0,42 < x < 0) ; P(1,37 < x < 2,01)$$
$$P(-1,79 < x < -0,54) ; P(|X| \le 0,5)$$

Tìm C sao cho: $P(0 < X < C) = 0,423 ; P(X < C) = 0,797$

$$P(C < X < 2) = 0,100 ; P(-1 \le X \le C) = 0,8054$$

12) Cho $X \sim N(13, 16)$: Tính xác suất:

$$(P < 20) ; P(X) > 10) ; P(5 < X < 21)$$

13) Hãy chia phân phối N(2,36) thành 10 lớp có xác suất bằng nhau.

14) Đo 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 trung bình của Na^+ và độ lệch chuẩn của phép đo trên.

15) Đường kính của một chi tiết máy do một máy tiện tự động sản xuất có phận phối bình thường với trung bình $\mu = 50$mm và độ lệch chuẩn $\sigma = 0,05$mm.

Chi tiết máy được xem như đạt yêu cầu nếu đường kính không sai quá 0,10mm.

a) Tính tỷ lệ đạt yêu cầu.

b) Lấy ngẫu nhiên ba sản phẩm của máy tiện đó. Tính xác suất có ít nhất một sản phẩm không đạt yêu cầu.

16) Trọng lượng X (gam) của một loại trái cây có phân phối bình thường $N(\mu = 500; \sigma^2 = 16)$.

Trái cây thu hoạch được phân loại theo trọng lượng:

Loại 1: trên 505 gam.

Loại 2: 495 - 505 gam

Loại 3: dưới 495 gam.

Tính tỷ lệ mỗi loại.

17) Tỷ lệ lọ thuốc hỏng trong trong các lô thuốc A, B lần lượt là 0,10 và 0,70. Giả sử các lô thuốc này có rất nhiều lọ.

a) Lấy ngẫu nhiên 3 lọ ở lô thuốc A.

Tính xác suất có ít nhất một lọ thuốc hỏng. Lấy tối thiểu máy lọ trong lô thuốc A để xác suất có ít nhất 1 lọ hỏng $\geq 0,90$?

b) Chọn ngẫu nhiên 1 trong 2 lô rồi lấy từ đó ra 1 lọ.

- Tính xác suất để lọ lấy ra là hỏng.
- Biết lọ lấy ra là hỏng. Tính xác suất để lô thuốc lấy ra là lô A.

c) Lấy ngẫu nhiên 50 lọ ở lô thuốc A. Tính xác suất để có 3 lọ hỏng?

18) Cho biết trọng lượng trẻ sơ sinh phân phối Bình thường với kỳ vọng là 3,2kg và phương sai $0,16kg^2$. Một trẻ sơ sinh được gọi là bình thường nếu trọng lượng từ 2,688 $\rightarrow$ 3,721kg. Đo trọng một cách ngẫu nhiên trên 100 trẻ sơ sinh. Tính:

a) Xác suất để có 85 trẻ bình thường.

b) Xác suất để có ít nhất 75 trẻ bình thường.

19) Cho biết trọng lượng viên thuốc sản xuất tại một xí nghiệp là độc lập và có phân phối Bình thường với kỳ vọng là 250mg, phương sai là $8,1mg^2$. Thuốc được đóng thành vĩ, mỗi vĩ 10 viên. Mỗi vĩ gọi là đúng tiêu chuẩn khi trọng lượng từ 2490mg đến 2510mg (đã trừ bao bì). Lấy ngẫu nhiên 100 vĩ để kiểm tra. Tính xác suất để:

a) Có 80 vĩ đạt tiêu chuẩn.

b) Có từ 70 vĩ trở lên đạt tiêu chuẩn.

20) Khảo sát một lô thuốc viên, trọng lượng trung bình của một viên thuốc là μ = 252,6mg và có độ lệch chuẩn σ =

4,2mg. Giả sử trọng lượng phân phối theo quy luật bình thường.

a) Tính tỷ lệ viên thuốc có trọng lượng lớn hơn 260mg.

b) Tính trọng lượng x_0 sao cho có 30% viên thuốc nhẹ hơn x_0.

c) Theo dược điển, viên thuốc đúng tiêu chuẩn phải có trọng lượng xung quanh trọng lượng trung bình với độ gia giảm tối đa 5%. Tính tỷ lệ các viên thuốc đúng tiêu chuẩn của lô thuốc được khảo sát.

B. LỜI GIẢI

[1]

Ω	ĐĐ	ĐT	ĐX	XX	XĐ	XT	TT	TĐ	TX
P_i	1/16	1/8	1/16	1/16	1/16	1/8	1/4	1/8	1/8

X	−4	−3	−2	1	2	6
f(x)	1/16	4/16	4/16	2/16	4/16	1/16

$E(x) = -\frac{1}{2}$; $Var(X) = 7,37$

[2] X~B(4; 0,52)

$f(x) = C_4^x(0,52)^x(0,48)^{4-x}$ với x = 0, 1, 2, 3, 4.

X	0	1	2	3	4
f(x)	0,054	0,230	0,373	0,270	0,073

E(X) = 2,078 ; Var(X) = 1,002

[3] a) Gọi X là số câu hỏi trả lời đúng trong 6 câu.

Ta có: $X \sim B(6, \frac{1}{5})$

Xác suất để một thí sinh không biết gì cả mà trả lời đúng 6 câu:

$$P(X = 6) = \left(\frac{1}{5}\right)^6 = (6,4) \cdot 10^{-5}$$

b) Xác suất để một thí sinh không biết gì cả mà lại đạt:

$$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6) = 0,017$$

c) Gọi Y là số câu đúng trong 3 câu còn lại: $Y \sim B\left(3, \frac{1}{5}\right)$

xác suất đạt của một thí sinh biết trả lời đúng 3 câu đầu:

$$P(Y \geq 1) = 1 - P(X) = 0) = 0,49$$

d) $X \sim B\left(6, \frac{1}{5}\right)$. Ta tính xác suất:

$$P(X \geq 4 | X \geq 3) = \frac{P(X \geq 4 \cap X \geq 3)}{P(X \geq 3)} = \frac{P(X \geq 4)}{P(X \geq 3)} = 0,17$$

[4] Gọi X là số chai không đúng tiêu chuẩn trong 10 chai

Ta có. $X \sim B\left(10, \frac{1}{10}\right)$

a) $P(X = 1) = 0,387$

b) $P(X \geq 1) = 1 - P(X = 0) = 1 - 0,349 = 0,651$

c) $P(X \leq 1) = P(X = 0) + P(X = 1) = 0,736$

[5] a) Gọi X = số sản phẩm hỏng trong 10 sản phẩm quan sát:

Ta có: $X \sim B(n = 10; p = 0.07)$

Do đó hàm mật độ của X là:

$$f(x) = \begin{cases} C_{10}^{x}(0,07)^{x}(0,93)^{10-x} & ; x = 0,1,2,...,10 \\ 0 & ; \text{nơi khác} \end{cases}$$

$$P(X = 1) = C_{10}^{x}(0,07)(0,39)^{9} = 0,36$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0,93)^{10} = 0,52$$

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= (0,93)^{10} + C_{10}^{x}(0,07)(0,93)^{9} = 0,85$$

b) Gọi n là số lần quan sát, ta có:

$$P(X \geq 1) = 1 - P(X = 0)$$

$$= 1 - (0,93)^{n} \geq 0,90$$

$$1 - 0,90 \geq (0,93)^{n}$$

$$(0,93)^{n} \leq 0,10$$

$$n \ln(0,93) \leq \ln).10$$

$$n \geq \frac{\ln 0,10}{\ln 0,93} = 31,72$$

Vậy quan sát ít nhất 32 sản phẩm.

[6] Ta có $X \sim$ Poisson ($\lambda = 3$), do đó hàm mật độ xác suất của X là:

$$f(x) = \begin{cases} e^{-3} \cdot \frac{3^{x}}{x!} & ; x = 0,1,2,3,... \\ 0 & ; \text{nơi khác} \end{cases}$$

Vậy: $P(X = 1) = e^{-3} \cdot 3 = 0,1494$

$$P(X = 2) = e^{-3} \cdot \frac{3^2}{2!} = 0,2240$$

$$P(X = 3) = e^{-3} \cdot \frac{3^3}{3!} = 0,2240$$

[7] Gọi X là số ca bị phản ứng:

$$X \sim B\left(2000, \frac{1}{1000}\right) \sim P(2) = \frac{e^{-2}2^x}{x!}$$

a) $P(X = 3) = \dfrac{e^{-2}2^3}{3!} = \dfrac{4}{3e^2} \doteq 0,18$

b) $P(X = 0) = \dfrac{1}{e^2} = 0,135$

$$P(X = 1) = \frac{2}{e^2} = 0,270$$

$$P(X = 2) = \frac{2}{e^2} = 0,270$$

Ta có:

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$ 0,854

c) $P(X \geq 4) = 1 - P(X \leq 3) = 0,146$

[8]

Qui luật \ Xác suất	$P(X = 0)$	$P(X = 1)$	$P(X > 1)$	Nhận xét
$X \sim B\left(20, \frac{1}{100}\right)$	0,818	0,165	0,017	Hai kết quả # là do p nhỏ
$X \sim P\left(\frac{1}{5}\right)$	0,818	0,164	0,018	

$E(X) = np = 2$

[9] Ta có: $\begin{cases} E(X) = np = 2 \\ \sigma^2 = np(1-p) = \dfrac{4}{3} \end{cases} \Rightarrow \begin{cases} n = 6 \\ p = \frac{1}{3} \end{cases}$

Vậy: $X \sim B\left(6; \frac{1}{3}\right)$

Nên hàm mật độ: $f(x) = C_6^x \left(\dfrac{1}{3}\right)^x \left(\dfrac{2}{3}\right)^{6-x}$; $x = 0; 1; ...; 6$

[10] a) $\mu = \displaystyle\int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x\left(\frac{1}{\lambda}e^{-\frac{x}{\lambda}}\right)dx$

Đặt: $u = \dfrac{x}{\lambda} \Rightarrow x = \lambda u; dx = \lambda.du$, ta có:

$$\mu = \int_0^{\infty} ue^{-u}(\lambda du) = \lambda$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_0^{\infty} x^2\left(\frac{1}{\lambda}^{-\frac{x}{\lambda}}\right)dx - \lambda^2$$

$$= \lambda^2 \int_0^{\infty} u^2 e^{-u}du - \lambda^2 = 2\lambda^2 - \lambda^2 = \lambda^2$$

b) Tính hàm gây moment:

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$= \int_0^{\infty} e^{tx} \cdot \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = \frac{1}{\lambda}\int_0^{\infty} e^{-\left(\frac{1-\lambda t}{\lambda}\right)} dx$$

$$M(t) = \frac{1}{1 - \lambda t}$$

Do đó: $M'(t) = \dfrac{\lambda}{(1 - \lambda t)^2} \Rightarrow \mu = M'(0) = \lambda$

$$M''(t) = \frac{2\lambda^2}{(1 - \lambda t)^3} \Rightarrow \mu = M''(0) = 2\lambda^2$$

Phương sai: $\sigma^2 = M''(0) - M'(0)^2 = 2\lambda^2 - \lambda^2 = \lambda^2$

[11] $X \sim N(0, 1)$

a) $P(0 < X < 1{,}42) = P(X < 1{,}42) - P(X < 0) = 0{,}922 - 0{,}5 = 0{,}422$

b) $P(-0{,}42 < X < 0) = P(0 < X < 0{,}42) = 0{,}663 - 0{,}500 = 0{,}163$

c) $P(1{,}37 < X < 2{,}01) = 0{,}978 - 0{,}915 = 0{,}063$

d) $P(-1{,}79 < X < -0{,}54) = P(0{,}54 < X < 1{,}79) = 0{,}963 - 0{,}705 = 0{,}258$

e) $P|X| < 0{,}5) = P(-0{,}5 < X < 0{,}5) = 2P(X < 0{,}5) - 1 = 0{,}382$

f) $P(0 < X < C) = 0{,}423 \Rightarrow P(X < C) - P(X < 0) = 0{,}423$

$\Rightarrow P(X < C) = 0{,}923$

$\Rightarrow C = 1{,}425$

g) $P(X < C) = 0{,}797 \Rightarrow C = 0{,}83$

h) $P(C < X < 2) = 0{,}100 \Rightarrow P(X < 2) - P(X < C) = 0{,}100$

$\Rightarrow P(X < C) = P(X < 2) - 0{,}100 = 0{,}877$

$\Rightarrow C = 1{,}16$

i) $P(-1<X<C) = 0{,}8045 \quad \Rightarrow P(X<C) - P(X < -1) = 0{,}8054$

$$\Rightarrow P(X<C) = P(X < -1) + 0{,}8054$$

$$= P(X > 1) + 0{,}8054$$

$$= 1 - P(X < 1) + 0{,}8054$$

$$= 1 - 0{,}841 + 0{,}8054$$

$$= 0{,}9644$$

$$\Rightarrow C = 1{,}8$$

[12] $X \sim N(13,16) \Rightarrow = \dfrac{X-13}{4} \sim N(0,1)$

- $P(X < 20) = P\left(\dfrac{X-13}{4} < \dfrac{20-13}{4}\right) = P\left(U < \dfrac{7}{4}\right) = 0{,}960$
- $P(X > 10) = P\left(\dfrac{X-13}{4} > \dfrac{10-13}{4}\right) = P(U > -0.75)$
- $P(5 < X < 21) = P(-2 < U < 2) = P(U < 2) - P(U < -2)$

$$P(U < 2) - (1 - P(U < 2)) = 2P(U < 2) - 1$$

$$= 2 \times 0{,}977 - 1 = 0{,}954$$

[13] $X \sim N(2,36) \quad \Rightarrow U = \dfrac{X-2}{6} \sim N(0,1)$

- $P(X < a_1) = 0{,}100 \Rightarrow P\left(U < \dfrac{a_1 - 2}{6}\right) = 0{,}001$

$$\Rightarrow P\left(U < \frac{2-a_1}{6}\right) = 0{,}900$$

$$\Rightarrow \frac{2-a_1}{6} = 1{,}28$$

$$\Rightarrow a_1 = -5,68$$

- $P(X < a_1) = 0,200 \Rightarrow P\left(U < \frac{a_2 - 2}{6}\right) = 0,200$

$$\Rightarrow P\left(U < \frac{2 - a_2}{6}\right) = 0,800$$

$$\Rightarrow \frac{2 - a_2}{6} = 0,84$$

$$\Rightarrow a_2 = -3,04$$

- $P(X < a_3) = 0,300 \Rightarrow P\left(U < \frac{a_3 - 2}{6}\right) = 0,300$

$$\Rightarrow P\left(U < \frac{2 - a_3}{6}\right) = 0,700$$

$$\Rightarrow \frac{2 - a_3}{6} = 0,525$$

$$\Rightarrow a_3 = -1,15$$

- $P(X < a_4) = 0,400 \Rightarrow P\left(U < \frac{2 - a_4}{6}\right) = 0,600$

$$\Rightarrow \frac{2 - a_4}{6} = 0,25$$

$$\Rightarrow a_4 = 0,5$$

- $P(X < a_5) = 0,500 \Rightarrow P\left(U < \frac{a_5 - 2}{6}\right) = 0,500$

$$\Rightarrow \frac{a_5 - 2}{6} = 0$$

$$\Rightarrow a_5 = 2$$

Tương tự: $a_6 = 3,5$; $a_7 = 5,15$; $a_8 = 7,04$; $a_9 = 9,68$

[14] Đặt: X = [Na^+]mEa/L

$X \sim N(\mu,\sigma^2)$

$P(X > 143) = 1 - P(X < 143) = 0{,}2$

$$\Rightarrow P(X < 143) = 0{,}8$$

$$\Rightarrow P\left(U < \frac{143-\mu}{\sigma}\right) = 0{,}8$$

$$\Rightarrow \frac{143-\mu}{\sigma} = 0{,}84 \qquad (1)$$

$$P(X > 141) = P\left(U < \frac{141-\mu}{\sigma}\right) = 0{,}3$$

$$\Rightarrow P\left(U < \frac{\mu-141}{\sigma}\right) = 0{,}7$$

$$\Rightarrow \frac{\mu-141}{\sigma} = 0{,}52 \qquad (2)$$

Từ (1) và (2) cho ta: $\mu = 141{,}8$

$\sigma = 1{,}5$

[15] a) Gọi X = đường kính của chi tiết máy.

Ta có: $P(|X-\mu| \le 0{,}10) = P\left(-\frac{0{,}10}{0{,}05} \le \frac{X-\mu}{\sigma} \le \frac{0{,}01}{0{,}05}\right)$

$$= P(-2 \le U \le 2)$$

$$= 2P(U \le 2) - 1 = 0{,}9544$$

b) Gọi Y = số sản phẩm không đạt yêu cầu (trong 3 sản phẩm lấy ra). Ta có:

$P(Y \ge 0) = 1 - P(Y = 0)$

$$= 1 - (0{,}9544)^3 = 0{,}1307$$

[16] Ta có:

$$P_1 = (X > 505) = P\left(\frac{X-\mu}{\sigma} > \frac{505-500}{4}\right)$$
$$= (PU > 1,25) = 0,1056$$

$$P_2 = P(495 < X < 505) = P\left(\frac{495-500}{4} < U \le \frac{505-500}{4}\right)$$
$$= P(-1,25 < U < 1,25)$$
$$= 2P(U < 1,25) - 1 = 0,788$$

$$P_3 = P(X < 495) = P\left(\frac{X-\mu}{\sigma} < \frac{495-500}{4}\right)$$
$$= P(U < -1,25)$$
$$= P(U > 1,25) = 0,1056$$

[17] a) • P (có ít nhất 1 lọ hỏng) = 1– p(3 lọ đều tốt)

$$= 1 - (0{,}9)^3 = 0{,}271$$

• Gọi n là số lọ thuốc cần lấy:

P(có ít nhất 1 lọ hỏng) $= 1 - (0{,}9)^n \ge 0{,}90$

$$\Rightarrow (0{,}9)^n \le 0{,}10$$
$$\Rightarrow n\ln(0{,}9) \le \ln(0{,}10)$$
$$n \ge \frac{n\ln(0,10)}{\ln(0,9)} = 21,854$$
$$\Rightarrow n = 22$$

b) $P(H) = P(H|A)\cdot P(A) + P(H|B)\cdot P(B)$

$$= 0,10 \times \frac{1}{2} + 0,07 \times \frac{1}{2}$$

$= 0,085$

$$P(A|H) = \frac{P(H|A)\cdot P(A)}{P(H)} = \frac{0,10\times\frac{1}{2}}{0,085} = 0,588$$

c) Gọi X là số lọ hỏng trong 50 lọ lấy ra:

X ~ B(50; 0,1) → X ~ N(5; 4,5) do n = 50 lớn

$$P(X = 3) = P(2,5 < X < 3,5)$$

$$= P\left(\frac{2,5-5}{\sqrt{4,5}} < U < \frac{3,5-5}{\sqrt{4,5}}\right)$$

$$= P(-118 < U < -0,71)$$

$$= P(0,71 < U < 1,18)$$

$$= 0,881 - 0,761 = 0,12$$

[18] Gọi X là trọng lượng trẻ sơ sinh ⇒ X ~ N(3,2; 0,16). Xác suất đứa trẻ bình thường là:

$$P = P(2,688 \le X \le 3,712)$$

$$= P\left(\frac{2,688-3,2}{0,4} \le \frac{X-3,2}{0,4} \le \frac{3,712-3,2}{0,4}\right)$$

$$= P(-1,28 \le U \le 1,28)$$

$$= 2\cdot P(U \le 1,28) - 1 = 0,8$$

Gọi X là trọng lượng trẻ bình thường trong 100 trẻ quan sát

⇒ Y ~ B(100; 0,8)

⇒ Y ~ N(80; 16)

a) (PY = 85) = P(84,5 < Y < 85,5)

$$= P\left(\frac{84,5-80}{4} < \frac{Y-80}{4} < \frac{85,4-80}{4}\right)$$

$= P(1{,}13 < U < 1{,}38)$, với $U = N(0{,}1)$

$= P(U < 1{,}38) - P(U < 1{,}13)$

$= 0{,}916 - 0{,}871 = 0{,}045$

b) $P(Y \geq 75) = P(Y \geq 74{,}5)$

$$= \left(\frac{Y-80}{4} \geq \frac{74{,}5-80}{4}\right)$$

$= P(U > -1{,}38) = P(U < 1{,}38) = 0{,}916$

[19] $\begin{cases} X_i \text{ trọng lượng viên thuốc: } X_i \sim N(250;\ 8{,}1) \\ X \text{ Trọng lượng vĩ thuốc : } X \sim N(2500;\ 81) \end{cases}$

Gọi A: Biến cố vĩ thuốc đạt tiêu chuẩn. Ta có:

$$P(A) = P(2490 \leq X \leq 2510) = P\left(-\frac{10}{9} \leq U \leq \frac{10}{9}\right) = 0{,}74$$

Y là số vĩ đạt tiêu chuẩn trong 100 vĩ chọn kiểm tra:

$Y \sim B(100;\ 0{,}74) \sim N(74;\ 19{,}24)$

a) $P(Y = 80) = P(79{,}5 \leq Y \leq 80{,}5)$

$$= P\left(-\frac{79{,}5-74}{4{,}386} \leq U \leq \frac{80{,}5-74}{4{,}836}\right)$$

$= P(1{,}25 \leq U \leq 1{,}48) = 0{,}037$

b) $P(Y \geq 70) = P\left(U \leq \frac{69{,}5-74}{4{,}836}\right)$

$= P\ (U \geq -1{,}03)$

$= P(U \leq 1{,}03) = 0{,}849$

[20] Gọi X_i là trọng lượng viên thuốc: $X \sim N(252{,}6;\ (4{,}2)^2)$

a) $P(X > 260) = P\left(\frac{X-\mu}{\sigma} > \frac{260-252,6}{4,2}\right) = P(U > 1,76)$

$= 1 - P(U \le 1,76) = 1 - 0,961 = 0,039$

b) $P(X < x_0) = 0,3 \Leftrightarrow P\left(\frac{X-\mu}{\sigma} < \frac{x_0-\mu}{\sigma}\right) = 0,3$

$\Leftrightarrow P\left(U < \frac{x_0 - 252,6}{4,2}\right) = 0,3$

$\Leftrightarrow 1 - P\left(U < \frac{252,6 - x_0}{4,2}\right) = 0,3$

$\Leftrightarrow P\left(U < \frac{252,6 - x_0}{4,2}\right) = 0,7$

$\Leftrightarrow \frac{252,6 - x_0}{4,2} = 0,52$

$\Leftrightarrow x_0 = 250,4$ mg

c) Đặt $M_1 = 252,6 - \frac{5}{100} \times 252,6 = \mu - \frac{5}{100}\mu = 252,6 - 12,63$

$M_2 = 252,6 + \frac{5}{100} \times 252,6 = 252,6 + 12,63$

Xác suất viên thuốc đạt tiêu chuẩn là:

$P(M_1 \le X \le M_2) = P\left(\frac{M_1-\mu}{\sigma} \le U \le \frac{M_2-\mu}{\sigma}\right)$

$= P\left(-\frac{12,63}{4,2} \le U \le \frac{12,63}{4,2}\right)$

$= P(-3 \le U \le 3) = 0,998 = 99,8\%$

III. Bài tập LÝ THUYẾT MẪU

BÀI TẬP

[1] Đo lượng cholesterlemie (đơn vị: mg%) của một số người, ta được:

X(mg%)	150-160	160-170	170-180	180-190	190-200	200-210
Số người	2	4	5	6	4	3

a) Tính trung bình mẫu $\overline{X}$ và độ lệch tiêu chuẩn của S_X.

b) một mẫu thứ nhì cũng quan sát lượng cholesterlemie Y(mg%) của 30 người, tính được $\overline{Y} = 180$mg%, $S_Y = 16$mg%. Nhập hai mẫu lại, tính trung bình và độ lệch chuẩn của mẫu nhập.

[2] Có 3 mẫu quan sát sức nặng con người, kết quả ghi nhận được như sau:

	Lần quan sát	Trung bình	Độ lệch
Mẫu 1	70	55kg	8,30kg
Mẫu 2	75	57kg	8,60kg
Mẫu 3	95	54kg	8,50kg

Nhập chung 3 mẫu lại, tính trung bình và độ lệch mẫu nhập.

[3] Đo độ dài của 30 chi tiết được chọn ngẫu nhiên của một loại sản phẩm ta được bảng số liệu sau:

39	43	41	41	40	41	43	42	41	39	40	41
44	42	42	41	41	42	43	40	41	41	42	43
39	40	41	39	40	42						

a) Tìm các đặc trưng mẫu $\overline{X}, S^2$.

b) Tìm bảng phân phối thực nghiệm

c) Tìm các hàm phân phối thực nghiệm tương ứng với mẫu trên.

[4] Khi kiểm tra thể lực một nhóm sinh viên, ta có kết quả về cân nặng như sau:

X_i(kg)	42,5 47,5	47,5 52,5	52,5 57,5	57,5 62,5	62,5 67,5
Số sinh viên (n_i)	8	14	28	18	12

a) Tính $\overline{X}, S^2, S$.

b) Lập bảng phân phối thực nghiệm

c) Tìm hàm phân phối thực nghiệm

BÀI GIẢI

[1] a) Ta có: n – 24 người

$$\overline{X} = 181,25\text{mg\%}$$
$$S_X = 14,98\text{mg\%}$$

b) Nhập mẫu:

	Quan sát	Trung bình	Độ lệch
Mẫu 1	24	181,25mg%	14,98mg%
Mẫu 2	30	180mg%	16mg%

Số lần quan sát N = 24 + 30 = 54 người

- Nhập trung bình: $\overline{X} = \dfrac{\sum X}{n} \Rightarrow \sum X = n\overline{X}$

Mẫu 1: $\sum X = (24)(181,25) = 4350\text{mg\%}$

Mẫu 2: $\sum Y = (30)(180) \quad = 5400mg\%$

Mẫu nhập: $\sum Z = \sum X + \sum Y \quad = 9750mh\%$

Trung bình mẫu nhập:

$$\overline{Z} = \frac{\sum Z}{N} = \frac{9750}{54} \quad = 180,56mg\%$$

- Nhập độ lệch:

$$S_X^2 = \frac{\sum X^2 - n\overline{X}^2}{n-1}$$

$$(n-1)S_X^2 = \sum X^2 - n\overline{X}^2$$

$$\sum X^2 = (n-1)S_X^2 + n\overline{X}^2$$

Mẫu 1: $\sum X^2 = (23)(14,98)^2 + 24(181,25)^2$

$= 793598,71\ (mg\%)^2$

Mẫu 2: $\sum Y^2 = (29)(16)^2 + (30)(180)^2$

$= 979424\ (mg\%)^2$

Mẫu nhập: $\sum Z^2 = \sum X^2 + \sum Y^2 = 1773022,7092$

- Phương sai mẫu nhập:

$$S_Z^2 = \frac{\sum Z^2 - N\overline{Z}^2}{N-1} = \frac{1773022,7092 - 54(180,56)^2}{53} = 236,21$$

$S_Z = 15,37$ mg%

[2] • Số lần quan sát: N = 70 + 75 + 95 = 240 người

• Nhập trung bình: $\overline{X} = \dfrac{\sum X}{n} \Rightarrow \sum X = n\overline{X}$

Mẫu 1: $\left(\sum X\right)_1 = 70(55) = 3850$ kg

Mẫu 2: $\left(\sum X\right)_2 = 75(57) = 4275$ kg

Mẫu 3: $\left(\sum X\right)_3 = 95(54) = 5130$ kg

Mẫu nhập: $\sum X = \left(\sum X\right)_1 + \left(\sum X\right)_2 + \left(\sum X\right)_3 = 13255$ kg

• Trung bình mẫu nhập:

$$\overline{X} = \frac{\sum X}{N} = \frac{13255}{240} = 55,23\text{kg}$$

• Nhập độ lệch:

$$S^2 = \frac{\sum X^2 - n\overline{X}^2}{n-1} \Rightarrow \sum X^2 = (n-1)S^2 + n\overline{X}^2$$

Mẫu 1: $\left(\sum X^2\right)_1 = 69(8,30)^2 + 70(55)^2 = 216503,41$

Mẫu 2: $\left(\sum X^2\right)_2 = 74(8,60)^2 + 75(57)^2 = 249148,04$

Mẫu 3: $\left(\sum X^2\right)_3 = 94(8,50)^2 + 95(54)^2 = 283811,50$

Mẫu nhập: $\sum X^2 = \left(\sum X^2\right)_1 + \left(\sum X^2\right)_2 + \left(\sum X^2\right)_3$

$= 749462,95$

• Phương sai mẫu nhập:

$$S^2 = \frac{\sum X^2 - N\overline{X}^2}{N-1} = \frac{749462,95 - 240(55,23)^2}{269} = 72,71$$

$S = 8{,}53$ kg

[3] Bảng số liệu ban đầu có thể thu gọn, khi xét đến tần số của các giá trị quan sát, ta được bảng sau:

X_i	39	40	41	42	42	44
n_i	4	5	9	7	4	1

$\sum_{i=1}^{6} n_i = 30, n_i$ là tần số của x_i trong mẫu đã cho.

$$\overline{X} = \frac{1}{30}(4.39 + 5.40 + 9.41 + 7.42 + 4.43 + 44) = 41{,}17$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2 - n\overline{X}\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{k}\left(n_i X_i^2 - n\overline{X}^2\right) = \frac{1}{29}(50.893 - 30 \times 41{,}17^2) = 1{,}80$$

Độ lệch chuẩn: $S = \sqrt{1{,}80} = 1{,}34$

Bảng phân phối thực nghiệm:

X_i	39	40	41	42	43	44
$\frac{n_i}{n}$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{9}{30}$	$\frac{7}{30}$	$\frac{4}{30}$	$\frac{1}{30}$

Từ bảng phân phối thực nghiệm dễ dàng tìm được phân phối $F_{30}(x)$ theo quy tắc "cộng dồn sang trái":

$$F_{30}(x) = \begin{cases} 0 & \text{nếu } x < 39 \\ \dfrac{4}{30} & 39 \le x < 40 \\ \dfrac{9}{30} & 40 \le x < 41 \\ \dfrac{18}{30} & 41 \le x < 42 \\ \dfrac{25}{30} & 42 \le x < 43 \\ \dfrac{29}{30} & 43 \le x < 44 \\ 1 & 44 \le x \end{cases}$$

[4] Khi các giá trị mẫu $x_i \in (a_i, b_i)$ thì ta lấy $x_i = \dfrac{a_i + b_i}{2}$. Từ đấy ta có bảng số liệu thu gọn:

X(kg)	45	50	55	60	65
Số sinh viên $n = \sum n_i = 80$	8	14	28	18	12

$$\overline{X} = \frac{1}{n}\sum n_i x_i = 55,75$$

$$S^2 = \frac{1}{n-1}\sum X_i^2 - \overline{X} = 34,87$$

$$S = 5,91$$

Ta có bảng phân phối thực nghiệm:

X_i	45	50	55	60	65
P_i	$\frac{8}{80}$	$\frac{14}{80}$	$\frac{28}{80}$	$\frac{18}{80}$	$\frac{12}{80}$

Hàm phân phối thực nghiệm:

$$F(x) = \begin{cases} 0 & \text{nếu } x < 45 \\ \dfrac{8}{80} & \text{nếu } 45 \le x < 50 \\ \dfrac{22}{80} & \text{nếu } 50 \le x < 55 \\ \dfrac{50}{80} & \text{nếu } 55 \le x < 60 \\ \dfrac{68}{80} & \text{nếu } 60 \le x < 65 \\ 1 & \text{nếu } 65 \le x \end{cases}$$

IV. Bài tập LÝ THUYẾT ƯỚC LƯỢNG

BÀI TẬP

[1] Đo đường kính X (đơn vị: mm) của một chi tiết máy do một máy tiện tự động sản xuất, ta ghi nhận được số liệu như sau:

X (mm)	12,00	12,05	12,10	12,15	12,20	12,25	12,30	12,35	12,40
Số trường hợp	2	3	7	9	10	8	6	5	3

a) Tính trung bình mẫu X và độ lệch tiêu chuẩn của S.

b) Ước lượng đường kính trung bình $\mu(x)$ của độ tin cậy 0,95.

Nếu muốn sai số ước lượng không quá $\varepsilon = 0,02$ mm ở độ tin cậy 0,95 thì quan sát mẫu ít nhất mấy trường hợp?

[2] Đem cân một số trái cây vừa thu hoạch ở nông trường, ta được kết quả như sau:

X(gam)	200-210	210-220	220-230	230-240	240-250
Số trái	12	17	20	18	15

a) Tìm khoảng ước lượng của trọng lượng trung bình μ của trái cây ở nông trường ở độ tin cậy 0,95 và 0,99.

b) Nếu muốn sai số ước lượng không quá $\varepsilon = 2$g ở độ tin cậy 0,95 thí quan sát ít nhất mấy trường hợp?

c) Trái cây có trọng lượng $X \geq 230$ gam được xếp vào loại A.

Hãy tìm khoảng ước lượng tỷ lệ p của trái cây loại A ở độ tin cậy 0,95 và 0,99. Nếu muốn sai số ước lượng không quá

$\varepsilon = 0{,}04$ ở độ tin cậy 0,99 thì quan sát mẫu ít nhất mấy trường hợp?

[3] Người ta đo ion Na^+ trên một số người và ghi nhận được kết quả như sau (đơn vị: mEq/lít)

129, 132, 140, 141, 138, 143

133, 137, 140, 143, 138, 140.

Tính trung bình mẫu $\overline{X}$ và phương sai của mẫu S^2.

Hãy ước lượng trung bình μ và phương sai σ^2 của dân số ở độ tin cậy 0,95.

Nếu muốn sai số ước lượng trung bình không quá $\varepsilon = 1$ mEa/l với độ tin cậy 0,95 thì quan sát mẫu ít nhất mấy người?

[4] Quan sát tuổi thọ X(giờ) của một số bóng đèn do xí nghiệp A sản xuất, ta có:

X	1000	1100	1200	1300	1400	1500	1600	1700	1800
Số trường hợp	10	14	16	17	18	16	16	12	9

Tính trung bình mẫu $\overline{X}$ và độ lệch tiêu chuẩn của mẫu.

Hãy ước lượng tuổi thọ trung bình của bóng đèn ở độ tin cậy 0,95 và 0,99.

Nếu muốn sai số ước lượng không quá $\varepsilon = 30$ giờ, ở độ tin cậy 0,99 thì quan sát mẫu ít nhất mấy trường hợp?

[5] Ta muốn ước lượng tỉ lệ viên thuốc bị sức mẻ p trong một lô thuốc rất nhiều.

a) Nếu ta muốn sai số ước lượng không quá 0,01 ở độ tin cậy $\gamma = 0{,}95$ thì phải quan sát ít nhất mấy viên?

b) Quan sát ngẫu nhiên 200 viên, thấy có 18 viên bị sứt mẻ:

- Hãy ước lượng p độ tin cậy $\gamma = 0,95$.
- Trong trường hợp này, nếu sai số ước lượng không quá 0,01 ở độ tin cậy $\gamma = 0,95$ thì phải quan sát ít nhất mấy viên?

[6] Quan sát chiều cao X(cm) của một số người, ta ghi nhận được:

X(cm)	140-145	145-150	150-155	155-160	160-165	165-170
Số người	1	3	7	9	5	2

a) Tính $\overline{X}$ và S^2.

b) Hãy ước lượng μ và σ^2 ở độ tin cậy $\gamma = 0,95$.

[7] Một loại thuốc mới được đem thử điều trị cho 50 người bị bệnh B, kết quả có 40 người người khỏi bệnh.

a) Hãy ước lượng tỉ lệ p khỏi bệnh nếu dùng thuốc đó điều trị, ở độ tin cậy $\gamma = 0,95$ và 0,99.

b) Nếu ta muốn sai số ước lượng không quá 0,02 ở độ tin cậy $\gamma = 0,95$ thì phải quan sát ít nhất mấy trường hợp.

[8] Một loại bệnh có tỉ lệ tử vong là 0,10.

Muốn chứng tỏ một loại thuốc có hiệu nghiệm (nghĩa là hạ thấp được tỉ lệ tử vong) ở độ tin cậy $\gamma = 0,95$ thì phải thử thuốc đó ít nhất trên bao nhiêu người?

[9] Để đánh giá sức khoẻ các bé gái sơ sinh, người ta kiểm tra số đo trọng lượng các cháu gái sơ sinh trong một bệnh viện và có kết quả thống kê như sau:

X	1,7-2,1	2,1-2,5	2,5-2,9	2,9-3,3	3,3-3,7	3,7-4,1
Số người	4	20	21	15	2	3

a) Biết trọng lượng bé gái sơ sinh theo phân phối chuẩn $N(\mu, \sigma^2)$.

Hãy tìm các ước lượng đúng của μ và σ^2.

b) Ta qui định những bé gái sơ sinh nặng trên 2,9kg là bé khoẻ. Hãy ước lượng tỉ lệ bé gái khoẻ trong vùng với độ tin cậy 99%.

c) Hãy tìm khoảng tin cậy cho trọng lượng trung bình của bé gái sơ sinh với độ tin cậy 95%.

[10] Cho hai mẫu độc lập cùng cỡ 10, rút ra từ dân số có phân phối $N(\mu_1, \sigma^2)$ và $N(\mu_2, \sigma^2)$.

Nếu: $\overline{X}_1 = 4,8; S_1^2 = 8,65$

$\overline{X}_2 = 5,6; S_2^2 = 7,88.$

Hãy tìm khoảng tin cậy 95% dành cho $\mu_1 - \mu_2$.

BÀI GIẢI

[1] Ta có: n = 53 trường hợp

$$\overline{X} = 12,21 \text{ mm}$$
$$S = 0,10 \text{ mm}$$

Khoảng ước lượng của trung bình:

$$\mu(x) = X \pm C\frac{S}{\sqrt{n}}$$

$$= 12,21 \pm (1,96\frac{0,10}{\sqrt{53}}$$

$$= 12,21 \pm 0,03 \text{ mm}$$

Muốn sai số ước lượng không quá ε = 0,02 mm thì cỡ mẫu là:

$$n \geq \frac{C^2S^2}{\varepsilon^2} = \frac{(1,96)^2(0,10)^2}{(0,02)^2} = 101,61$$

$n \geq 102$ trường hợp

[2] Ta có: n = 82 trái

$$\overline{X} = 225,85 \text{ gam}$$

$$S = 13,26 \text{ gam}$$

$$\mu = \overline{X} \pm C\frac{S}{\sqrt{n}}$$

Với độ tin cậy 95% ta có:

$$\mu = 225,85 \pm (1,96)\frac{13,26}{\sqrt{82}} = 225,85 \pm 2,78\text{g}$$

Với độ tin cậy 99% ta có:

$$\mu = 225,85 \pm (2,58)\frac{13,26}{\sqrt{82}} = 225,85 \pm 3,78\text{g}$$

Nếu muốn sai số không quá 2 gam, thì cỡ mẫu là:

$$n \geq \frac{C^2S^2}{\varepsilon^2} = \frac{(1,96)^2(13,26)^2}{(2)^2} = 168,84$$

$n \geq 169$ trái

Tỉ lệ trái cây loại A ở mẫu là: $f = \dfrac{18+15}{82} = 0,40$

Khoảng ước lượng của p là:

$$p = f \pm \sqrt{\frac{f(1-f)}{n}}$$

Với độ tin cậy 0,95 ta có:

$$p = 0,40 \pm (1,96)\sqrt{\frac{(0,4)(0,6)}{82}} \doteq 0,40 \pm 0,11$$

Với độ tin cậy 0,95 ta có:

$$p = 0,40 \pm (2,58)\sqrt{\frac{(0,4)(0,6)}{82}} = 0,40 \pm 0,14$$

Nếu muốn sai số ước lượng không quá ε = 0,04 ở độ tin cậy 0,99 thì cỡ mẫu là:

$$n \geq \frac{C^2 f(1-f)}{\varepsilon^2} = \frac{(2,58)^2(0,4)(0,6)}{(0,04)^2} = 998,46$$

$n \geq 999$ trái

[3] Ta có: n = 12 người

$$\overline{X} = 137,92 \text{ mEq / lít}$$

$$S = 4,42 \text{ mEq / lít}$$

$$S^2 = 19,54(\text{nEq / lít})^2$$

Khoảng ước lượng của trung bình:

$$\mu = \overline{X} \pm C\frac{S}{\sqrt{n}}$$

$$= 137,92 \pm (1,96)\frac{4,42}{\sqrt{12}}$$

$$\doteq 137,92 \pm 2,50 \text{ mEq / l}$$

Khoảng ước lượng phương sai:

$$\frac{\sum(X_i - \overline{X})^2}{b} \le \sigma^2 \le \frac{\sum(X_i - \overline{X})^2}{a}$$

Ta có: $\sum(X_i - \overline{X})^2 = (n-1)S^2 = 11(19,54) = 214,19$

$$P(Y \ge b) = 0,025 \Rightarrow b = 21,92$$

$$P(Y \ge a) = 0,975 \Rightarrow b = 3,82$$

(Các giá trị a, b đọc ở bảng chi bình phương (11)).

Do đó: $\dfrac{214,92}{21,92} \le \sigma^2 \le \dfrac{214,92}{3,82} \Rightarrow 9,80 \le \sigma^2 \le 56,26$

Nếu muốn sai số không quá ε = 1 mElq/l ở độ tin cậy 0,95 thì cỡ mẫu là:

$$n \ge \frac{C^2S^2}{\varepsilon^2} = \frac{(1,96)^2(19,54)}{(1)^2} = 75,06$$

Vậy quan sát ít nhất 76 người.

[4] Ta có: n = 128 trường hợp

$\overline{X} = 1391,41$ giờ

$S = 234,45$ giờ

Khoảng ước lượng của trung bình:

$$\mu = \overline{X} \pm C\frac{C}{\sqrt{n}}$$

Với độ tin cậy 0,95, ta có:

$$\mu = 1391,41 \pm (1,96)\frac{234,45}{\sqrt{128}}$$

$$= 1391,41 \pm 40,62 \text{ giờ}$$

Với độ tin cậy 0,99, ta có:

$$\mu = 1391,41 \pm (2,58)\frac{234,45}{\sqrt{128}}$$

$$= 1391,41 \pm 53,46 \text{ giờ}$$

Nếu muốn sai số ước lượng không quá ε = 30 giờ, ở độ tin cậy 0,99 thì cỡ mẫu là:

$$n \geq \frac{C^2S^2}{\varepsilon^2} = \frac{(2,58)^2(234,45}{(30)^2} = 406,52$$

$n \geq 407$ trường hợp

[5] $n \geq \dfrac{C^2}{4\varepsilon^2} = \dfrac{(1,96)^2}{4(0,01)^2} = 9604$ viên

$$f = \frac{18}{200} = 0,09$$

- $p = f \pm C\sqrt{\dfrac{f(1-f)}{n}} = 0,09 \pm (1,96\sqrt{\dfrac{(0,09)(0,91)}{200}}$

  p = 0,09 ± 0,04

- $n \geq \dfrac{C^2f(1-f)}{\varepsilon^2} = (1,96)^2\dfrac{(0,09)(0,91)}{(0,01)^2} = 3146,27$

  $n \geq 3147$ viên

[6] n = 27 người

$\overline{X} = 156,20$ cm,

S = 6,14 cm,

$S^2 = 37,68\ (cm)^2$

$$\mu = X \pm C\frac{S}{\sqrt{n}} = 156,20 \pm (2,06)\frac{6,14}{\sqrt{27}}$$

$$\mu = 156{,}20 \pm 2{,}43$$

$$\frac{\sum (X_i - \overline{X})^2}{b} \le \sigma^2 \le \frac{\sum (X_i - \overline{X})^2}{b}$$

$$\sum (x_i - \overline{X})^2 = (n-1)S^2 = 26(37,28) = 979,63(\mathrm{cm})^2$$

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1=26)$$

$$\begin{cases} P(Y > b) = 0,025 \\ P(Y > a) = 0,975 \end{cases} \Rightarrow \begin{cases} b = 41,92 \\ a = 13,84 \end{cases}$$

$$\frac{979,63}{14,92} \le \sigma^2 \le \frac{979,63}{13,84}$$

$$23,37 \quad \le \sigma^2 \le 70,78\ \mathrm{cm}^2$$

[7] $$f = \frac{40}{50} = 0,80$$

$$P = f \pm C\sqrt{\frac{f(1-f)}{n}}$$

$$\gamma = 0,95 : P = 0,80 \pm (1,96)\sqrt{\frac{(0,8)(0,2)}{50}} = 0,80 \pm 0,11$$

$$\gamma = 0,99 : P = 0,80 \pm (2,58)\sqrt{\frac{(0,8)(0,2)}{50}} = 0,80 \pm 0,15$$

$$n \ge \frac{C^2 f(1-f)}{2} = \frac{(1,96)^2(0,8)(0,2)}{(0,02)^2} = 1536,64$$

$n \geq 1537$ người.

[8] $P_0 = 0{,}10$

Gọi P = tỉ lệ tử vong khi có dùng thuốc.

Dùng thuốc đặc trị một số người, theo dõi kết quả (lấy mẫu).

Đặt : $f = \dfrac{\sum X_i}{n}$ = tỉ lệ số vong ở mẫu.

- Nếu $f \geq 0{,}10$: Thuốc phản tác dụng (loại)
- Nếu $0 < f < 0{,}10$: Ước lượng P

$$P = f \pm C\sqrt{\frac{f(1-f)}{n}}$$

Thuốc có hiệu nghiệm nếu:

$$P_{max} = f + C\sqrt{\frac{f(1-f)}{n}} < 0{,}10$$

$$C\sqrt{\frac{f(1-f)}{n}} < 0{,}10 - f$$

$$\frac{C\sqrt{f(1-f)}}{0{,}10 - f} < \sqrt{n}$$

$$n \geq \frac{C^2 f(1-f)}{(0{,}10 - f)^2}$$

[9] a)

X	1,9	2,3	2,7	3,1	3,5	3,9
Số người	4	20	21	15	2	3

Ta có: $n = 65$

$\overline{X} = 2{,}7$

$S = 0{,}464$

- Ước lượng đúng của μ là $\overline{X} = 2{,}7$
- Ước lượng đúng của σ^2 là $S^2 = 0{,}215$

b) Ta cần tìm ước lượng khoảng của tỉ lệ

$$f = \frac{m}{n} = \frac{20}{65}$$

Với độ tin cậy 99% nên $C = 2{,}58$

Vậy khoảng ước lượng là:

$$\left[f - C\sqrt{\frac{f(1-f)}{n}}; f + C\sqrt{\frac{f(1-f)}{n}}\right]$$

$$= [0{,}159\ ;\ 0{,}455]$$

c) Ta có:

$$\mu = \overline{X} \pm \frac{CS}{\sqrt{n}} \text{ với } C = 1{,}96$$

$$= 2{,}7 \pm 0{,}11.$$

[10] Vì: $\dfrac{S_1^2}{S_2^2} < 4 \Rightarrow \dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\widehat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2)$

$$\Rightarrow \mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm C_\gamma \widehat{\sigma}\sqrt{\frac{1}{n} - \frac{1}{m}}$$

Với: $\widehat{\sigma^2} = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} = \dfrac{9(8,65) + 9(7,88)}{18} = 8,26$

$$\begin{aligned}
\Rightarrow \quad \widehat{\sigma} \quad &= 2,87 \\
C_\gamma \quad &= t_{0,025}(18) \\
&= 2,1 \\
\mu_1 - \mu_2 \quad &= (5,6 - 4,8) \pm 2,1(2,87)\sqrt{\frac{2}{10}} \\
&= 0,8 \pm 2,69
\end{aligned}$$

V. Bài tập PHÉP KIỂM THỐNG KÊ

- SO SÁNH TỈ LỆ - DÙNG PHÉP KIỂM u VÀ PHÉP KIỂM χ^2.

[1] Đầu năm bệnh viện A đưa ra một số cải tiến về phương pháp điều trị. Cuối năm tổng kết thấy số tử vong là 45 người. Trong khi đó những năm trước số tử vong trung bình là 65 người. Hỏi cải tiến về phương pháp điều trị có đem lại kết quả không? Biết rằng mỗi năm bệnh viện A điều trị khoảng 2000 ca.

[2] Khám ngẫu nhiên 150 người thấy có 12 người mắc bệnh K phổi. Hỏi quan sát này có phù hợp với tỉ lệ bênh K phổi là 7% trong cộng đồng?

[3] Bệnh X theo điều trị đã gây tử vong 15%. Một loại thuốc A dùng cho 250 bệnh nhân bị bệnh X thấy có 20 người tử vong. Hỏi hiệu quả của thuốc A trong việc điều trị bệnh X?

[4] Tại một địa phương tỉ lệ bệnh sốt rét là 20%. Dùng DDT để diệt muỗi. Khám 100 người thấy có 13 người bị sốt rét. Hỏi DDT có làm giảm tỉ lệ bệnh này không?

[5] Có khoảng 12% người bị huyết khối khi thay van tim trong vòng 4 năm. Người ta muốn xét xem Aspirin có ảnh hưởng tới bị huyết khối khi thay van tim không? Chọn ngẫu nhiên 188 bệnh nhân sau khi thay van tim, cho dùng 100 mg Aspirin/ngày suốt 4 năm liền, theo dõi thấy có 21 trường hợp bị huyết khối. Kết luận?

[6] Quan sát 100 người bị tâm thần phân liệt, ta có:

Mùa	Xuân	Hạ	Thu	Đông
Số BN	20	25	20	35

Hỏi tâm thần phân liệt có phụ thuộc theo mùa không?

[7] Một nghiên cứu trên bệnh án của 300BN nữ, trong độ tuổi sinh đẻ, bị chẩn đoán là tắt mạch máu không rõ nguyên nhân. Về PP ngừa thai có 90 người dùng thuốc; có 79 người đặt vòng; có 60 người dùng màn chắn; số còn lại không dùng gì. Hỏi bệnh tắt mạch máu không rõ nguyên nhân có liên quan đến PP ngừa thai không?

[8] Gọi t là thời gian chủng ngừa bệnh X cho đến khi bệnh này. Ta có kết quả thực nghiệm sau:

		t ≤ 10	t < 1 ≤ 25	25 < t
Bị bệnh	nặng	1	42	230
	vừa	6	114	347
	nhẹ	23	301	510

Hỏi mức độ bị bệnh X có phụ thuộc vào thời gian chủng ngừa không?

Hỏi mức độ bị bệnh X có phụ thuộc vào thời gian chủng ngừa không?

[9] Dùng 2 loại thuốc A và B để điều trị cho 2 nhóm người. Kết quả:

	khỏi	giảm	biến chứng	tử vong
A	84	39	16	11
B	41	36	9	14

Hỏi tác dụng của hai loại thuốc trên có như nhau không?

- SO SÁNH 2 SỐ TRUNG BÌNH - DÙNG PHÉP KIỂM u VÀ PHÉP KIỂM t.

[10] Đo lượng cholesterolemie (X mg%) trên một số người bình thường. Kết quả:

X	125-149	150-174	175-199	200-224	225-249	250-274	275-299	300-324
n	2	5	5	7	10	10	8	3

a) Cho hằng số sinh học trung bình về cholesterolemie là 225 mg%. Hỏi kết quả thực nghiệm trên có khác hằng số sinh học trung bình về cholesterolemie không?

b) Lượng cholesterolemie trung bình của 25 người bệnh B là 245 mg%, s = 50 mg%. Hỏi bệnh B có làm thay đổi lượng cholesterolemie trung bình của người bình thường không?

[11] Một mẫu 35 người bị K tiền liệt tuyến có di căn, hàm lượng trung bình PSA là 15 mg/ml; s = 1,5 mg/ml. Dùng PSA làm chất chỉ điểm có di căn trong bệnh K tiền liệt tuyến. Với bệnh K tiền liệt tuyến chưa di căn, hàm lượng trung bình PSA là 12 mg/ml. Hỏi PSA có thể làm cho chất chỉ điểm có di căn trong bệnh K tiền liệt tuyến được không?

[12] Một mẫu 10 bệnh nhân sốt rét, đo đường huyết trung bình là 0,8 g/l. Hằng số sinh học trên đường huyết là 1 ± 0,2 g/l, độ tin cậy 95%. Hỏi bệnh sốt rét có làm giảm đường huyết không?

[13] Quan sát trọng lượng của 32 trẻ sơ sinh trai ta có:

X(kg)	2,2	2,5	2,8	3,2	3,4	3,7
Số trẻ	1	1	6	13	8	3

Quan sát trọng lượng của 30 trẻ sơ sinh gái ta có trọng lượng trung bình 3kg; s = 0,3 kg. Hỏi có sự khác nhau về trọng lượng trung bình của trẻ sơ sinh trai và gái không?

[14] Một máy phân tích huyết học được gọi là không đạt yêu cầu, khi máy chạy hết công suất, trung bình máy phân tích 99 mẫu máu/ngày. Chọn một máy phân tích huyết học, cho chạy thử một tuần hết công suất, kết quả như sau:

Thứ	2	3	4	5	6	7	C.nhật
Số mẫu máu/ngày	104	93	97	101	105	95	105

Hỏi máy phân tích huyết học trên đạt yêu cầu chưa?

[15] Hàm lượng Na^+ trong máu của người có huyết áp bình thường và người có huyết áp cao, số liệu thực nghiện như sau:

Nhóm	cở mẫu	trung bình	độ lệch chuẩn
huyết áp bình thường	15	144	6,2
huyết áp cao	12	160	3,9

Kết luận?

• SO SÁNH 2 PHƯƠNG SAI - DÙNG PHÉP KIỂM F.

[16] Hãy so sánh 2 phương sai trong bài tập [13], [15].

[17] Quan sát trọng lượng của hai nhóm người như sau:

	n	Trọng lượng TB	s
Nhóm 1 cao trung bình 1,56m	40	58kg	8kg
Nhóm 2 cao trung bình 1,65m	25	60kg	10kg

Hãy so sánh hai phương sai và so sánh 2 số trung bình?

BÀI GIẢI

- SO SÁNH 2 TỈ LỆ - DÙNG PHÉP KIỂM u VÀ PHÉP KIỂM χ^2

[1] Giả thiết H: Số ca tử vong năm nay khác với số ca tử vong năm trước không có ý nghĩa.

	Tử vong	Không tử vong	Tổng
Năm nay	45(55)	1955(1945)	2000
Năm trước	65(55)	1935(1945	2000
	110	3890	4000

$Q = \sum \frac{(n_i - n'_i)^2}{n'_i} = 3,7391.$ Vì $Q < \chi^2_{0,005}(1) = 3,84$ nên chấp nhận H, nghĩa là số ca tử vong năm nay khác với số ca tử vong năm trước không có ý nghĩa.

[2] Giả thiết H: Tỉ lệ bệnh K phổi quan sát khác với tỉ lệ bệnh K phổi trong cộng đồng không có ý nghĩa.

Quan sát: $n_1 = 12$; $n_2 = 138$

$$\begin{cases} n'_1 = 150.7\% = 10,5 \\ n'_2 = 139,5 \end{cases}$$

$Q = \sum \frac{(n_i - n'_i)^2}{n'_i} = 0,230,$ vì $Q < \chi^2_{0,05}(1) = 3,84$ nên chấp nhận H, nghĩa là tỉ lệ bệnh K phổi quan sát khác với tỉ lệ K phổi trong cộng đồng không có ý nghĩa.

[3] Ta có: f = 20/250 = 0,08

Đặt giả thiết H: Sự khác biệt giữa f và p không có ý nghĩa.

Tính: $u = \frac{p - f}{\sqrt{\frac{p(1-p)}{n}}} = 3,18$

Vì $|u| > 2,58$ nên bác bỏ H, ngưỡng sai lầm 1%.

Kết luận: Sự khác biệt giữa f và p có ý nghĩa, ngưỡng sai lầm 1%.

[4] và [5] Làm tương tự như bài [3].

[6] Đặt giả thiết H: Bệnh tâm thần phân liệt không phụ thuộc theo mùa.

Mùa	Xuân	Hạ	Thu	Đông	
Số BN	20(25)	25(25)	20(25)	35(25)	100

Vì: $Q = \sum \frac{(n_i - n'_i)^2}{n'_i} = 6 < \chi^2_{0,05}(3) = 7,82$ nên chấp nhận H.

Bệnh tâm thần phân liệt không phụ thuộc theo mùa.

[7] Đặt giả thiết H: Bệnh tất mạch máu không rõ nguyên nhân không liện quan đến PPNT.

PPNT	dùng thuốc	đặt vòng	màn chắn	Không dùng gì	
Số BN	90(75)	79(75)	60(75)	71(75)	300

Vì: $Q = \sum \frac{(n_i - n'_i)^2}{n'_i} = 6,426 < \chi^2 0,05\ (3) = 7,82$ nên chấp nhận H.

Bệnh tất mạch máu không rõ nguyên nhân không liên quan đến PPNT.

[8] Đặt giả thiết H: Mức độ bệnh X không phụ thuộc vào thời gian chủng ngừa.

		$t \leq 10$	$10 < t \leq 25$	$25 < t$	tổng
Bị bệnh X	nặng	1	42	230	273
	vừa	6	114	347	467
	nhẹ	23	301	510	834
Tổng		30	457	1087	1574

	$t \leq 10$	$10 < t \leq 25$	$25 < t$
nặng	5,203	79,263	188,534
vừa	8,9	135,59	322,51
nhẹ	15,879	242,147	575,956

Vì: $Q = \sum \frac{(n_i - n_i')^2}{n_i'} = 61 > \chi^2 0,01\,(4) = 18,46$ nên bác bỏ H, ngưỡng 0,001.

Kết luận: Mức độ bệnh X có phụ thuộc vào thời gian chủng ngừa, ngưỡng 0,001.

[9] Đặt giả thiết H: Hai loại thuốc trên có tác dụng khác nhau không có ý nghĩa:

	Khỏi	Giảm	Biến chứng	Tử vong	
A	84	39	16	11	150
B	41	36	9	14	100
	125	75	25	25	250

	Khỏi	Giảm	Biến chứng	Tử vong
A	75	63	15	15
B	50	42	10	10

Vì: $Q = \sum \frac{(n_i - n'_i)^2}{n'_i} = 7,53 < \chi^2_{0,05}(3) = 7,82$ nên chấp nhận H.

Kết luận: Hai loại thuốc trên có tác dụng khác nhau không có ý nghĩa.

- SO SÁNH HAI SỐ TRUNG BÌNH - DÙNG PHÉP KIỂM u VÀ PHÉP KIỂM t.

[10] a)

X	137	162	187	212	237	262	287	312
n	2	5	5	7	10	10	8	3

$n = 50$; $\overline{X} = 234,5$ mg%; $s = 46,91$ mg%; $\mu = 225$ mg%.

Đặt giả thiết H: sự khác biệt giữa x và μ không có ý nghĩa.

Tính: $u = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = 1,43$

Vì $|u| < 1,96$ nên ta chấp nhận H.

Kết luận: Kết quả thực nghiện trên khác hằng số sinh học trung bình cholesterolemie không có ý nghĩa.

b) Đặt giả thiết H: sự khác biệt giữa μ_1; μ_2 không có ý nghĩa:

Ước lượng:

$$\sigma' = \sigma'_1 = \sigma'_2 = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 47,95$$

Ta có: $t = \dfrac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = -0,89$

Vì $|t| < t_{0,05}(n_1 + n_2 - 2 = 73) = 1,98$ nên ta chấp nhận H.

Kết luận: Bệnh B không làm thay đổi lượng cholesterolemie trung bình của người bình thường.

[11] Làm giống như bài [10] a).

[12] Khoảng tin cậy 95% của cá thể là: $x = \mu \pm 1,96\sigma$

Theo giả thiết suy ra: $\mu = 1$và σ 0,2/1,96 = 0,102.

Đặt giả thiết H: $\overline{x} < \mu$ không có ý nghĩa.

Tính: $u = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = -6,211$

Vì $u < -1,64$ nên ta bác bỏ H, ngưỡng 0,05.

Kết luận: Bệnh sốt rét làm giảm đường huyết, P = 0,05.

• **Chú ý**: Phép kiểm u nói trên là phép kiểm u – 1 đuôi.

$$\begin{cases} P(u < -C = 0,05 \Rightarrow P(u > C) & = 0,05 \\ \qquad\qquad\qquad\quad \Rightarrow 0,5 - \Phi_0(C) & = 0,05 \\ \qquad\qquad\qquad\quad \Rightarrow \Phi_0(C) & = 0,45 \\ \text{Tra bảng ta tìm được } C = 1,64 & \end{cases}$$

[13] Ta nói: $\overline{x_1} = 3,128\text{kg}; s_1 = 0,335\text{kg}; \overline{x_1} = 3\text{kg}; s_2 = 0,3\text{kg}.$

Đặt giả thiết H: trọng lượng trung bình của trẻ sơ sinh trai và gái khác nhau không có ý nghĩa.

Vì $n_1 \geq 30, n_2 \geq 30$ ta có thể xem $\sigma_1 \approx s_1, \sigma_2 \approx s_2$.

Tính được: $u = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = 1,588$

Vì $|u| < 1,96$ nên ta chấp nhận H.

Kết luận: Trọng lượng trung bình của trẻ sơ sinh trai và gái khác nhau không có ý nghĩa.

[14] Ta có: $n = 7$; $\overline{x} = 100$; $s = 5$.

Đặt giả thiết H: sự khác biệt giữa $\overline{x}$ và μ không có ý nghĩa.

Tính: $t = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = 0,53$

Vì $|t| < t_{0,05}(n - 1 = 6) = 2,447$ nên ta chấp nhận H.

Kết luận: máy phân tích huyết học trên không đạt yêu cầu.

[15] Đặt giả thiết H: sự khác biệt giữa μ_1; μ_2 không có ý nghĩa.

Ước lượng: $\sigma \cong \sigma_1 \cong \sigma_2 \cong \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 5,31$

Ta có: $t = \dfrac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = -7,77$

Vì $|t| > t_{0,01}(n_1 + n_2 - 2 = 25) = 2,787$ nên ta bác bỏ H, ngưỡng 0,01.

Kết luận: Hàm lượng Na^+ trong máu của người có huyết áp bình thường và người có huyết áp cao khác nhau có ý nghĩa.

- SO SÁNH 2 PHƯƠNG SAI - DÙNG PHÉP KIỂM F

[16] Trong bài [13], ta có: $s_1^2 = 0,112; n_1 = 32$ và $s_2^2 = 0,09$; $n_2 = 30$.

Đặt giả thiết H: sự khác biệt của s_1^2 và s_2^2 không có ý nghĩa.

Tính: $f = \frac{s_1^2}{s_2^2} = 1,244$

Vì $F < F_{0,95}(31;29) = 1,6$ nên ta chấp nhận H.

Kết luận: Sự khác biệt của s_1^2 và s_2^2 không có ý nghĩa.

Trong bài [15], làm tương tự.

[17] Ta có: $s_1^2 = 100; n_1 = 25$ và $s_2^2 = 64; n_2 = 40$.

Đặt giả thiết H: sự khác biệt của s_1^2 và s_2^2 không có ý nghĩa.

Tính: $F = \frac{s_1^2}{s_2^2} = 1,562$

Vì $F < F_{0,95}(24;39) = 1,8$ nên ta chấp nhận H.

Kết luận: Sự khác biệt của s_1^2 và s_2^2 không có ý nghĩa.

So sánh 2 số trung bình làm như các bài so sánh 2 trung bình trên.

BÀI TẬP
(Tự giải)

[18] Cây GENTIANA BELLIDIFOLIA có 2 giống: 1 giống cho hoa cái và 1 giống cho hoa lưỡng phái. lấy ngẫu nhiên 109 cây, người ta đếm được 82 cây cho hoa lưỡng phái và 27 cây cho hoa cái. Khảo sát sự phù hợp của phân phối này với giả thuyết $\frac{1}{2}/\frac{1}{2}$ và giả thuyết $\frac{3}{4}/\frac{1}{4}$ cho mỗi loại hoa.

[19] Khảo sát sự phù hợp của phân phối sau đây với một phân phối POISSON.

Số hồng cầu k	0	1	2	3	4	5	6	7	8	9	≥10
Số ô = n_k	1	6	16	21	20	15	11	5	3	1	1

(Số hồng cầu đếm được trong mỗi ô của hồng cầu kế)

[20] Khảo sát sự phù hợp của phân phối sau đây với một phân phối chuẩn.

Trọng lượng (mg)	≤ 327	327 - 328	328 - 329	329 - 330	330 - 331	≥ 331
Số viên	6	8	14	9	7	15

[21] Trong số 152 người sống chung trong 1 trại, có 93 người dùng DDT để ngừa một bệnh ngoài da. Sau một thời gian người ta ghi nhận số người bị bệnh ngoài da như sau:

– Trong số 93 người dùng DDT, có 29 người bị bệnh.

– Trong số 59 người không dùng DDT, có 23 người bị bệnh. Hỏi DDT có tác dụng không?

[22] Sự lai giống đến thế hệ thứ hai của hai đặc tính cho 4 loại tượng hình (theo quy luật Mendel) với tỷ lệ 9/16, 3/16, 3/16, 1/16. Một mẫu 453 cá thể có thế hệ thứ hai được phân phối như sau: 243 / 94 / 91 / 25. Cách phân phối này có phù hợp với quy luật Mendel không?

[23] 42 người dùng một loại thuốc A. Sau một thời gian nhất định có 15 người xuống cân 17 người không đổi và 10 người lên cân. Thuốc A có ảnh hưởng gì trên trọng lượng không?

[24] 72 người uống cafein: có 30 người mất ngủ.

73 người uống placebo (viện giả): có 9 người mất ngủ.

Cafein có ảnh hưởng đến sự mất ngủ không?

[25] Kiểm nghiệm độ trong của 4 lô thuốc tiêm, ta được kết quả:

Lô	Đúng tiêu chuẩn	Cần kiểm lại	Loại bỏ	Tổng số
A	203	150	6	359
B	266	112	1	379
C	258	126	2	386
D	196	168	17	381
Tổng số	923	556	26	1505

Độ trong của 4 lô có như nhau không?

[26] Ba nhân viên kiểm nghiệm X, Y, Z được giao mỗi người 200 lọ pénicilline lấy ngẫu nhiên trong một lô. Họ có nhiệm vụ phân chia thành 4 loại, tùy theo mức độ biến thành dịch treo của chất bột khi thêm chất lỏng vào:

a) Rất tốt b) Khá

c) Tạm được d) Tồi

Kết quả được ghi nhận trong bảng sau đây:

	a	b	a	d	Tổng số
X	125	53	18	4	200
Y	112	59	22	7	200
Z	128	57	10	5	200
Tổng số	365	169	50	16	600

[27] Hai máy tự động A và B đếm viên cho vào chai. Trong một tháng hoạt động, người ta kiểm tra 254 chai (máy A) và nhận thấy có 46 chai không đúng số viên qui định đồng thời có 33 chai không đúng số viên qui định trên tổng số 303 chai của máy B được kiểm tra. Kết luận?

[28] Trong một xí nghiệp sản xuất thuốc kháng sinh, 4 máy lên men hoạt động trong những điều kiện như nhau. Sau 3 năm, số lần hỏng của 4 máy là 4, 3, 8, 9. Có thể kết luận rằng 2 máy sau cùng xấu hơn 2 máy đầu không?

[29] Có 2 phương pháp chữa một bệnh B, 100 người bệnh chữa theo phương pháp 1, có 6 người chết, 100 người bệnh chữa theo phương pháp 2, có 3 người chết. Dùng phép kiểm χ^2 để so sánh 2 tỷ lệ này.

a) Không dùng hiệu chỉnh Yates.

b) Dùng hiệu chỉnh Yates.

Kết luận.

[30] Trên 10.000 trẻ sơ sinh, người ta quan sát thấy 5241 con trai. Tỷ lệ này có phù hợp với quy luật Mendel ($p = q = \frac{1}{2}$) không? Nếu chỉ quan sát được 52 con trai trên tổng số 100 trẻ sơ sinh, thì kết luận thế nào.

[31] Một giống chuột đặc biệt bị một loại ung thư với tỷ lệ 20%. Người ta thử nghiệm một phương pháp chống ung

thư trên 100 con, và nhận xét có 14 con bị ung thư. Phương pháp này có hiệu nghiệm không? Dùng phép kiểm u "2 đuôi" và "1 đuôi" để thử giải bài toán này.

[32] Trong một dân số nhất định, có 13 bé trai và 23 bé gái mắc bệnh B. Có thể kết luận rằng bé gái dễ mắc bệnh hơn bé trai không?

[33] Trong số 13 076 em bé được chủng ngừa BCG, có 152 em chết, và trong số 11 881 em bé không chủng ngừa BCG (tất cả các điều kiện khác đều như nhau) cò 255 em chết. BCG có hiệt lực không?

[34] Một vườn cây thuốc có 102 cây, được xịt thuốc ngừa một bệnh B. Cò 20 cây chết vì bệnh này. Một vườn khác có 98 cây không xịt thuốc ngừa, có 29 cây chết. Kết luận?

[35] Điều tra về nguyên nhân ung thư phổi, người ta hỏi:

- 300 người bị ung thư phổi: có 3 người không hút thuốc lá.
- 300 người không bị ung thư phổi: có 30 người không hút thuốc lá.

Hai tỷ lệ này thật sự có khác nhau không? Có thể kết luận rằng hút thuốc lá là một nguyên nhân gây ung thư phổi không?

[36] Người ta khảo sát độc tính đối với chuột, sống cô lập hoặc chung cả nhóm.

Thí nghiệm 1: chích amphetamin

- Chuột cô lập: chết 8 trên 84.
- Chuột nhóm: chết 58 trên 84.

Thí nghiệm 2: chích reserpin ngay sau khi chích amphetamin

– Chuột cô lập: chết 1 trên 30.

– Chuột nhóm: chết 15 trên 30.

Reserpin có tác dụng không? Hai kết luận trên đây (2 trường hợp chuột cô lập và chuột nhóm) có giá trị thống kê không?

[37] Hai loại thuốc tiêm được thử trên 100 người để biết ý kiến của từng người là thuốc "đau" hay "không đau". Mỗi người được thử cả hai loại, thứ tự trước sau được quyết định bằng cách tút thăm. Kết quả:

Thuốc A	Thuốc B	Số người
Đau	Đau	45
Đau	Không đau	15
Không đau	Đau	5
Không đau	Không đau	35
	Tổng số:	100

[38] Bốn loại thuốc chữa cùng một bệnh B được thử trên 4 nhóm bệnh nhân:

	Số người khỏi	Số người không khỏi	Tổng số
1	123	28	151
2	95	19	114
3	152	63	215
4	132	53	185
Tổng số	502	163	665

Kết luận?

[39] Năm 1946, thuốc streptomycin được thử tác dụng chống lao.

– 55 người điều trị bằng streptomycin : 4 người chết

– 52 người chứng : 14 người chết

Kết luận?

[40] Một loại thuốc tiêm có tiêu chuẩn về nồng độ hoạt chất được quy định là 250 ± 10mg ($x = \mu \pm 1,9\sigma$) (khoảng tin cậy 95% của các x). Lấy một mẫu 10 ống trong lô thuốc 7341 để kiểm nghiệm, ta được:

243 248 245 252 250 241 243 240 238 240

Lô thuốc 7341 có đúng tiêu chuẩn không?

[41] Một bảo sanh viện cung cấp các dữ kiện về các trẻ sơ sinh như sau:

– 41 con trai có trọng lượng trung bình 3,400kg và độ lệch chuẩn $s_1 = 0,380$kg.

– 65 con gái có trọng lượng trung bình 3,360kg và độ lệch chuẩn $s_2 = 0,360$kg.

Có thể kết luận con trai nặng hơn con gái được không?

[42] Đo chiều dài của loại giun móc ANKYLOSTOMA BRAZILIENSE, người ta được các dữ kiện:

– 126 con đực có chiều dài trung bình 10,67mm và phương sai 1,48mm^2.

– 121 con cái có chiều dài trung bình 11,74mm và phương sai 1,29mm^2.

Con cái có dài hơn con đực không?

[43] Hai loại thuốc ngủ A và B được tiêm cho hai lô vật thử nghiệm và ghi nhận thời gian ngủ (tính bằng phút).

– Lô 1 (thuốc ngủ A) gồm 10 con vật:

170 175 197 180 190 165 175 174 173 181

– Lô 2 (thuốc ngủ B) gồm 12 con vật:

155 160 164 150 160 159 154 156 160 167
153 158

Tác dụng của A và B có khác nhau không?

[44] Người ta tiêm 2 liều esrin khác nhau cho 2 lô chuột và ghi nhận thời gian t từ lúc tiêm đến lúc chuột bắt đầu run:

– Lô 1 gồm 10 con được tiên 30μg, và tính ra khoảng tin cậy 95% của thời gian t trung bình là 8,40 ± 0,72 phút $= \bar{x} \pm t_{0,05} \frac{s}{\sqrt{n-1}}$

– Lô 2 gồm 10 con được tiêm 60μg, khoảng tin cậy 95% là 7,70 ± 0,80 phút.

Hai liều này có ảnh hưởng trên thời gian t không?

[45] Người ta đo tính cương cơ trước và sau khi tác dụng một chất A trên 10 vật thử nghiệm. Kết quả:

Trước	75	96	32	41	50	39	59	45	30	33
Sau	53	67	21	29	35	27	37	30	21	10

Chất A có tác dụng trên tính cương cơ không?

[46] Để khảo sát tác dụng tăng lực của một chất A, người ta ghi nhận thời gian chuột bơi – lúc bắt đầu bơi đến lúc kiệt sức. Kết quả được hai chuỗi thống kê (thời gian tính bằng phút).

(1)	30	24	32	38	21	35
(2)	42	31	41	42	31	35

a) Giả sử 2 chuỗi này là kết quả thử nghiệm trên 2 lô chuột, mỗi lô 6 con, lô 1 làm chứng, lô 2 được uống chất A. Kết luận?

b) Giả sử 2 chuỗi trên là kết quả thử nghiệm trên 1 lô 6 con chuột, chuỗi 1 là kết quả trước khi uống chất A, chuỗi 2 là kết quả sau khi uống chất A. Kết luận?

[47] Sau đây là phân phối thống kê trọng lượng của trẻ sơ sinh chia thành 2 nhóm con so và con rạ:

Lớp (kg)	Số con so n_{Ai}	Số con ra n_{Bi}
[2,0 ; 2,2[	2	1
[2,2 ; 2,4[	4	1
[2,4 ; 2,6[	6	3
[2,6 ; 2,8[	4	9
[2,8 ; 3,0[	10	6
[3,0 ; 3,2[	18	12
[3,2 ; 3,4[	21	20
[3,4 ; 3,6[	17	19
[3,6 ; 3,8[	5	12
[3,8 ; 4,0[	4	10
[4,0 ; 4,2[	3	6
[4,2 ; 4,4[	0	4
[4,4 ; 4,6[	1	1
[4,6 ; 4,8[	0	0
[4,8 ; 5,0[	0	0
[5,0 ; 5,2[	0	1

Dựa vào các số liệu trên đây, có thể kết luận con rạ nặng hơn con so không?

[48] Một phản ứng kết tinh chất A trong một qui trình sản xuất cho trung bình 34kg chất A khô, với độ lệch chuẩn 0,75kg.

Thay đổi một chi tiết của qui trình và thử 10 lần, ta được trung bình 34,35kg chất A khô.

a) Chi tiết được thay đổi có làm thay đổi trọng lượng chất A sản xuất không?

b) Chi tiết này có làm tăng trọng lượng chất A không?

[49] Vẫn bài toán trên đây, giả sử ta chỉ biết μ = 34kg, không biết độ lệch chuẩn σ. Thay đổi một chi tiết của qui trình và thử 10 lần, ta được 10 số liệu:

34,30 34,60 34,00 34,85 33,60 34,20 34,10

33,40 34,10 36,40

Kết luận?

[50] Chất poniciphinat procain được định phân bằng hai phương pháp iod-kế và triền quang kế. Để so sánh kết quả, người ta định phân 9 lô thuốc:

Lô số	1	2	3	4	5	6	7	8	9
Phương pháp 1	985	1015	980	1000	1010	1005	995	1020	1045
Phương pháp 2	990	980	995	995	990	965	985	995	1005

Kết quả theo phương pháp 1 có kết quả cao hơn kết quả phương pháp 2 không?

[51] Nhiều người được cho uống một xirô có chứa aslicylic acid. Người ta định phân chất này trong máu 1 giờ và 2 giờ sau khi uống.

Kết quả (đơn vị: mg/l):

Người số	1	2	3	4	5	6	7	8	9	10
Sau 1 giờ	128	54	166	138	130	90	130	132	110	192
Sau 2 giờ	170	76	146	130	118	85	110	170	78	202

Lượng salisylic acid có thay đổi tùy 2 thời gian định phân trên đây?

[52] Để kiểm nghiệm một thuốc A, người ta định phân chất Cl_2 trong nước tiểu trước và sau khi tiêm thuốc A. Kết quả (đơn vị: mEa/giờ)

Bệnh nhân số	1	2	3	4	5	6	7	8	9	10
Trước khi tiêm	3,32	2,32	1,9	4,73	4,64	1,74	4,20	0,73	3,08	4,61
Sau khi tiêm	5,29	8,96	9,83	5,91	0,59	7,42	9,00	17,71	3,05	6,42

Sự bài tiết chất Cl_2 có bị thay đổi bởi thuốc A không?

VI. Bài tập PHÂN TÍCH PHƯƠNG SAI

- So sánh nhiều số trung bình - ANOVA

[1] Làm sinh thiết gan trên 22 BN đã được chẩn đoán bệnh A, B, C, D, E để đo hàm lượng GGTP (gọi là X; đơn vị μg). Kết quả:

	A	B	C	D	E
	27,7	45,9	85,3	39,6	41,8
X	25,8	39	64,1	41,1	46,3
	38,1	40,4	74,4	35,3	52,7
	39,6	34	78,2	32,6	57,2

Hỏi lượng GGTP trung bình trong 5 bệnh trên khác nhau có ý nghĩa không? Nếu lượng GGTP trong 5 bệnh trên khác nhau ý nghĩa, hãy so sánh từng cặp số trung bình?

[2] Nghiên cứu về hiệu quả của 3 loại thuốc A, B, C dùng điều trị chứng suy nhược thần kinh, 12 người bị suy nhược thần kinh chia làm 3 nhóm dùng thuốc A, B, C; trong mội nhóm có đủ các mức độ bệnh. Sau 1 tuần điều trị (đánh giá bằng thang điểm), kết quả như sau:

Thuốc A	25	15	20	14
Thuốc B	20	16	18	25
Thuốc C	25	15	20	20

Hãy đánh giá hiệu quả các loại thuốc A, B, C. Nếu hiệu quả các loại thuốc A, B, C khác nhau có ý nghĩa, hãy so sánh từng cặp thuốc.

[3] So sánh hiệu quả giảm đau của 4 loại thuốc A, B, C, D bằng cách chia 20 bệnh nhân thành 4 nhóm, mỗi nhóm

dùng một loại thuốc giảm đau. kết quả (mức độ giảm đau được chấm theo thang điểm).

A	B	C	D
82	80	77	65
89	70	69	75
77	72	67	67
72	90	65	55
	78		63
	92		70

Hỏi hiệu quả giảm đau của 4 loại thuốc có khác nhau không? Nếu hiệu quả của các loại thuốc A, B, C, D khác nhau có ý nghĩa, hãy so sánh từng cặp thuốc.

[4] Nghiên cứu về hiệu quả của 3 loại thuốc A, B, C dùng điều trị chứng suy nhược thần kinh. 12 người bị suy nhược thần kinh được chia làm 4 nhóm tùy theo mức độ bệnh 1, 2, 3, 4; trong mỗi nhóm chia ra để dùng thuốc A, B, C. Sau 1 tuần điều trị (đánh giá bằng thang điểm), kết quả như sau:

	Mức độ bệnh	1	2	3	4
	A	25	40	25	30
Thuốc	B	30	39	25	31
	C	28	42	27	29

Hãy đánh giá hiệu quả các loại thuốc A, B, C.

[5] 10 người bị bệnh suyễn tham gia thử nghiệm để đánh giá hiệu quả của 3 loại thuốc A, B, C bằng cách đo FEV (lít) sau 2 giờ dùng thuốc:

Bệnh nhân		1	2	3	4	5	6	7	8	9	10
	A	0	0,04	0,02	0,02	0,04	0,03	0,05	0,02	0	0,12
Thuốc	B	0,13	0,17	0,2	0,27	0,11	0,18	0,21	0,23	0,24	0,08
	C	0,26	0,23	0,21	0,19	0,36	0,25	0,32	0,38	0,3	0,3

Đánh giá hiệu quả các loại thuốc A, B, C trên khác nhau không?

[6] Điện thế trên da (mV) của 8 người khi có trạng thái tâm lý: lo sợ, vui mừng, trầm cảm, bình tĩnh như sau:

Bệnh nhân	1	2	3	4	5	6	7	8
Lo sợ	24	55	10	22	11	38	18	22
Vui mừng	23	53	9	19	14	37	15	23
Trầm cảm	21	54	11	20	13	36	16	21
Bình tĩnh	22	52	8	21	12	35	17	20

Hỏi trạng thái tâm lý có làm thay đổi điện thế trên da không?

BÀI GIẢI

- **So sánh nhiều số trung bình - ANOVA**

[1] Đặt giả thiết H: lượng GGTP trung bình trong 5 bệnh trên khác nhau không có ý nghĩa.

Ta có: $\overline{y}_1 = 32,8$; $\overline{y}_2 = 39,825$; $\overline{y}_3 = 75,5$; $\overline{y}_4 = 37,15$; $\overline{y}_5 = 49,5$.

$s_1^2 = 49,78$; $s_2^2 = 23,49$; $s_3^2 = 78,16$; $s_4^2 = 15,24$; $s_5^2 = 46,35$.

• **Phép kiểm COCHRAN**

Đặt giả thiết H: sự khác biệt của các phương sai s_1^2, s_2^2, s_3^2, s_4^2, s_5^2 khác nhau không có ý nghĩa.

$$\text{Tính } G = \frac{\max\{s_1^2\}}{\sum s_1^2} = 78,16 / 213,47 = 0,366$$

Tra bảng COCHRAN, $G_{0,99}(m = 3; n = 5) = 0,695$

Vì G < 0,695 nên chấp nhận H. Nghĩa là các s_1^2, s_2^2, s_3^2, s_4^2, s_5^2 khác nhau không có ý nghĩa.

Nguồn	TBPĐL	df	Phương sai	F
Giữa các mức của A	4674,532	4	1168,633	27,36942
Trong từng mức	640,4775	15	42,6985	

Do $F > F_{0,99}(4; 15) = 5,5$ nên bác bỏ H, ngưỡng 0,01.

Kết luận: Lượng GGTP trung bình trong 5 bệnh trên khác nhau có ý nghĩa.

• ***Phép kiểm DUCAN***:

a) Xếp thứ tự các số trung bình từ nhỏ đến lớn:

$\bar{y}_1$	$\bar{y}_4$	$\bar{y}_2$	$\bar{y}_5$	$\bar{y}_3$
32,8	37,15	39,82	49,5	75,5

b) Tính: $s = \sqrt{\frac{s_r^2}{n}} = \sqrt{\frac{42,6985}{4}} = 3,267.$

c) Tra bảng DUCAN, độ tự do = N – k = 15, ngưỡng 0,05.

	2	3	4	5
Hạng r	3,01	3,16	3,25	3,31
rs	9,833	10,323	10,617	10,813

d) Kết quả so sánh:

$\bar{y}_3 - \bar{y}_1 = 42,7 > 10,813$

$\Rightarrow$ $\bar{y}_1$ và $\bar{y}_3$ khác nhau có ý nghĩa, ngưỡng 0,05.

$\bar{y}_3 - \bar{y}_4 = 38,35 > 10,617$

$\Rightarrow$ $\bar{y}_3$ và $\bar{y}_4$ khác nhau có ý nghĩa, ngưỡng 0,05.

$\bar{y}_3 - \bar{y}_2 = 35,68 > 10,323$

$\Rightarrow$ $\bar{y}_3$ và $\bar{y}_2$ khác nhau có ý nghĩa, ngưỡng 0,05.

$\bar{y}_3 - \bar{y}_5 = 26 > 9,833$

$\Rightarrow$ $\bar{y}_3$ và $\bar{y}_5$ khác nhau có ý nghĩa, ngưỡng 0,05.

$\bar{y}_5 - \bar{y}_1 = 16,7 > 10,617$

$\Rightarrow$ $\bar{y}_5$ và $\bar{y}_1$ khác nhau có ý nghĩa, ngưỡng 0,05.

$\bar{y}_5 - \bar{y}_4 = 12,35 > 10,323$

$\Rightarrow$ $\bar{y}_4$ và $\bar{y}_5$ khác nhau có ý nghĩa, ngưỡng 0,05.

$\bar{y}_5 - \bar{y}_2 = 9,68 < 9,833$

$\Rightarrow$ $\bar{y}_5 - \bar{y}_2$ = khác nhau không có ý nghĩa.

$\bar{y}_2 - \bar{y}_1 = 7,02 < 10,323$

$\Rightarrow \bar{y}_2 - \bar{y}_1 =$ khác nhau không có ý nghĩa.

$\bar{y}_2 - \bar{y}_4 = 2{,}67 < 9{,}833$

$\Rightarrow \bar{y}_2 - \bar{y}_4 =$ khác nhau không có ý nghĩa.

$\bar{y}_4 - \bar{y}_1 = 4{,}35 < 9{,}833$

$\Rightarrow \bar{y}_4 - \bar{y}_1 =$ khác nhau không có ý nghĩa.

[2] Làm như bài [1].

[3] Đặt giả thiết H: hiệu quả giảm đau của 4 loại thuốc A, B, C, D khác nhau không có ý nghĩa.

Ta có: $\bar{y}_1 = 80$; $\bar{y}_2 = 80,8$; $\bar{y}_3 = 69,5$; $\bar{y}_4 = 65,83$

$s_1^2 = 52,66$; $s_2^2 = 101,2$; $s_3^2 = 27,66$; $s_4^2 = 45,76$

- ***Phép kiểm BARTLETT:***

Đặt giả thiết H: sự khác biệt của các phương sai: s_1^2; s_2^2; s_3^2; s_4^2 không có ý nghĩa.

Gọi $V_1 = 3$, $V_2 = 5$, $V_3 = 3$, $V_4 = 5$ là độ tự do trong các phương sai s_1^2; s_2^2; s_3^2; s_4^2.

Tính: $s^2 = \dfrac{v_1 \cdot s_1^2 + \ldots + v_4 \cdot s_4^2}{v_1 + \ldots + v_4} = 60,98; v = \sum\limits_i v_i = 16$

$$B = 2,303\left(v \cdot \lg s^2 - \sum_i v_i \cdot \lg s_1^2\right)$$

$$= 2,303(16 \times 1,785 - 27,8) = 1,75$$

$$C = 1 + \frac{1}{3(n-1)}\left(\sum_i \frac{1}{v_1} - \frac{1}{v} i\right)$$

$$= 1 + \frac{1}{3(4-1)}(1,066 - 0,062) = 1,111$$

Vì B/C = 1,575 < $\chi^2_{0,05}$(n – 1 = 3) = 7,82 nên chấp nhận H.

Nghĩa là sự khác biệt của các phương sai $s_1^2; s_2^2; s_3^2; s_4^2$ không có ý nghĩa.

Nguồn	TBPĐL	df	Phương sai	F
Giữa các nhóm	852,1035	3	284,0345	4,8712
Trong mỗi nhóm	874,6333	15	58,3089	

Do F > $F_{0,95}$ (3; 15) = 3,25 nên bác bỏ H, ngưỡng 0,05.

Kết luận: Hiệu quả giảm đau của 4 loại thuốc A, B, C, D khác nhau có ý nghĩa.

- ***So sánh từng cặp số trung bình***

Dùng phép kiểm t để so sánh từng cặp số trung bình trong các số trung bình $\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4$.

Trong trường hợp này ta lấy: $\sigma^2 \equiv s_r^2$ = 53,3089 và độ tự do v = 16.

So sánh $\bar{y}_1$ và $\bar{y}_2$: $|T| = \dfrac{|\bar{y}_1 - \bar{y}_2|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 0,169 < t_{0,95}(16) = 2,12$

Suy ra $\bar{y}_1$ và $\bar{y}_2$ khác nhau không có ý nghĩa.

So sánh $\overline{y}_1$ và $\overline{y}_3$: $|T| = \dfrac{|\overline{y}_3 - \overline{y}_1|}{\sigma\sqrt{\dfrac{1}{n_3} + \dfrac{1}{n_1}}} = 2{,}034 < t_{0,95}(16) = 2{,}12$

Suy ra $\overline{y}_1$ và $\overline{y}_3$ khác nhau không có ý nghĩa.

So sánh $\overline{y}_1$ và $\overline{y}_4$: $|T| = \dfrac{|\overline{y}_1 - \overline{y}_4|}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_4}}} = 3{,}009 > t_{0,95}(16) = 2{,}12$

Suy ra $\overline{y}_1$ và $\overline{y}_4$ khác nhau có ý nghĩa, ngưỡng 0,05.

So sánh $\overline{y}_2$ và $\overline{y}_3$: $|T| = \dfrac{|\overline{y}_3 - \overline{y}_2|}{\sigma\sqrt{\dfrac{1}{n_3} + \dfrac{1}{n_2}}} = 3{,}559 > t_{0,95}(16) = 2{,}12$

Suy ra $\overline{y}_2$ và $\overline{y}_3$ khác nhau có ý nghĩa, ngưỡng 0,05.

So sánh $\overline{y}_2$ và $\overline{y}_4$: $|T| = \dfrac{|\overline{y}_4 - \overline{y}_2|}{\sigma\sqrt{\dfrac{1}{n_4} + \dfrac{1}{n_2}}} = 3{,}559 > t_{0,95}(16) = 2{,}12$

Suy ra $\overline{y}_2$ và $\overline{y}_4$ khác nhau có ý nghĩa, ngưỡng 0,05.

So sánh $\overline{y}_3$ và $\overline{y}_4$: $|T| = \dfrac{|\overline{y}_4 - \overline{y}_3|}{\sigma\sqrt{\dfrac{1}{n_4} + \dfrac{1}{n_3}}} = 0{,}779 < t_{0,95}(16) = 2{,}12$

Suy ra $\overline{y}_3$ và $\overline{y}_4$ khác nhau không có ý nghĩa.

[4] Đặt giả thiết H_1: Yếu tố mức độ bệnh (B) không ảnh hưởng đến kết quả điều trị.

Đặt giả thiết H_2: Yếu tố thuốc (A) không ảnh hưởng đến kết quả điều trị.

Nguồn	TBPĐL	df	Phương sai	F
Thuốc (A)	5,116	2	2,583	0,905
Mức độ bệnh (B)	382,916	3	127,638	45,503
Ngẫu nhiên	16,833	6	2,805	

Kết luận: – Vì $F_B > F_{0,99}$ (c – 1 = 3; (h – 1)(c – 1) = 6) = 9,8 nên bác bỏ H_1, mức độ bệnh có ảnh hưởng đến kết quả điều trị, ngưỡng sai lầm α = 0,01.

– Vì $F_A > F_{0,95}$ (h – 1 = 2 ; (h – 1)(c – 1) = 6) = 5,1 nên chấp nhận H_2, các loại thuốc A, B, C không ảnh hưởng đến kết quả điều trị.

[5] Đặt giả thiết H_1: Yếu tố bệnh nhân (B) không ảnh hưởng đến FEV.

Đặt giả thiết H_2: Yếu tố thuốc (A) không ảnh hưởng đến FEV.

Nguồn	TBPĐL	df	Phương sai	F
Thuốc (A)	0,3067	2	0,1533	43,553
Bệnh nhân (B)	0,0158	9	0,0017	0,498
Ngẫu nhiên	0,0633	18	0,0035	

Kết luận: – Vì $F_B > F_{0,95}$ (c – 1 = 9; (h – 1)(c – 1) = 18) = 2,6 nên chấp nhận H_1, yếu tố bệnh nhân không ảnh hưởng đến FEV.

– Vì $F_A > F_{0,99}$ (h – 1 = 2; (h – 1)(c – 1) = 18) = 6 nên bác bỏ H_2, các loại thuốc A, B, C có ảnh hưởng đến FEV, ngưỡng sai lầm α = 0,01.

[6] Đặt giả thiết H_1: Yếu tố bệnh nhân (B) không ảnh hưởng đến điện thế trên da.

Đặt giả thiết H_2: Yếu tố trạng thái tâm lý (A) không ảnh hưởng đến điện thế trên da.

Nguồn	TBPĐL	df	Phương sai	F
Tâm lý (A)	10,75	3	3,583	2,572
Bệnh nhân (B)	5783,5	7	826,214	593,179
Ngẫu nhiên	29,25	21	1,392	

Kết luận: – Vì $F_B > F_{0,99}$ (c – 1 = 7; (h – 1)(c – 1) = 21) = 3,5 nên bác bỏ H_1, yếu tố bệnh nhân có ảnh hưởng điện thế trên da.

– Vì $F_A < F_{0,95}$ (h – 1 = 3; (h – 1)(c – 1) = 21) = 2,9 nên chấp nhận H_2, yếu tố bệnh nhân không ảnh hưởng đến điện thế trên da.

BÀI TẬP *(Tự giải)*

[7] Một dung dịch chất adenin được đo bằng quang phổ kế ở hai bước sóng 250nm và 260nm, phép đo được lặp lại 10 lần cho mỗi bước sóng, để so sánh phương pháp về mặt chính xác. Kết quả:

$\lambda_1 = 250$nm	120	164	153	148	143	132	155	142	169	144
$\lambda_2 = 260$nm	172	210	206	199	192	184	204	190	218	198

Tính chính xác có khác nhau không?

[8] Người ta muốn so sánh hàm lượng hoạt chất của một dược liệu trồng tại 2 vùng A và B. Một mẫu 12 cây được chọn trong mỗi vùng và định phân kết quả:

Vùng A	11,0	12,0	12,7	12,7	11,6	11,8
Vùng B	15,8	14,3	16,3	14,9	14,2	14,7

Vùng A	13,3	13,8	12,3	11,4	14,0	14,2
Vùng B	15,4	14,5	15,0	16,6	16,9	16,8

Kết luận?

[9] Để khảo sát ảnh hưởng của nồng độ K^+ trên sự tăng trưởng của mô thực vật người ta thực hiện 8 loại thử nghiệm với 8 nồng độ K^+ khác nhau. Kết quả được tóm tắt sau đây:

Số thứ tự	1	2	3	4	5	6	7	8
n_i	9	8	9	9	9	7	8	8
$\overline{y}_i$	187	198	215	241	258	245	250	203

Với $\sum y_{ij}^2 = 3,487.135$ và $\overline{y} = 124,3$.

Kết luận?

[10] Người ta muốn khảo sát ảnh hưởng của 5 môi trường cấy BCG. Muốn vậy, BCG được cấy vào 10 ống cho mỗi loại môi trường và sự tăng trưởng của BCG được tính theo số khuẩn lọc đếm được. Kết quả:

Môi trường	A	B	C	D	E
Số khuẩn lọc	10	11	7	12	7
	12	18	14	9	6
	8	12	10	11	10
	10	15	11	10	7
	6	13	9	7	7
	13	8	10	8	5
	9	15	9	13	6
	10	16	11	14	7
	8	9	7	10	9
	9	13	9	11	6

Kết luận?

[11] Người ta muốn khảo sát ảnh hưởng của chất ouabain trên sự phóng thích nor-adrenalin ở cơ tim chuột.

Kết quả:

Lô 1 (chứng)	Lô 2 (ouab: 0,25mg/kg)	Lô 3 (ouab: 0,50)	Lô 4 (ouab: 1,00)
0,49	0,63	0,51	0,66
0,60	0,93	0,53	0,48
0,59	0,48	0,28	0,25
0,62	0,34	0,70	0,30
0,76	0,83	0,43	0,35
0,57	0,44	0,40	0,61
0,62	0,86	0,46	0,45
0,53			0,26
1,03			0,41

a) Chất ouabain có hàm lượng nor-adrenalin phóng thích không? (Số liệu trang bảng là lượng NA phóng thích)

b) Biểu diễn trên một đồ thi số trung bình của mỗi lô kèm theo khoảng tin cậy 95% theo lượng ouabain.

[12] Người ta muốn khảo sát ảnh hưởng của hai yếu tố trên số lượng dược liệu thu hoạch được tại một nông trường: yếu tố 1 là số lần xịt thuốc trừ sâu và yếu tố 2 là dược liệu được thu hoạch vào lúc nào trong năm.

Kết quả:

Số lần xịt thuốc trừ sâu	thu hoạch 1	thu hoạch 2	thu hoạch 3
9	1060 (kg)	1228	262
6	949	1262	489
5	684	1116	650
4	473	1190	737
3	314	1024	862

Kết luận?

V.II. Bài tập HỒI QUI TƯƠNG QUAN

[1] Trên một mẫu n = 100 sản phụ, ta đo chiều cao bụng X(cm) và trọng lượng trẻ mới sinh Y(gam) rồi rút gọn số liệu qua phép biến số:

$$x = X - 30 \,; y = \frac{Y - 3200}{200}$$

Ta tính được: $\sum x = 54 \,; \sum y = -60$

$$\sum x^2 = 560 \,; \sum xy = 270 \,; \sum y^2 = 350$$

1) Tính hệ số tương quan giữa X và Y. Có nhận xét gì?

2) Tìm phương trình đường thẳng hồi quy của Y theo X. Một sản phụ đang chuyển dạ, có chiều cao bụng 36cm thì con có trọng lượng nhẹ nhất cũng là bao nhiêu với độ tin cậy 95%?

BÀI GIẢI

$$\bar{x} = \frac{\sum x}{n} = \frac{54}{100} = 0,54$$

$$S_x^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{560 - 100(0,54)^2}{99} = 5,3620$$

$$S_x = \sqrt{5,620} = 2,3156$$

$$\bar{X} = 30 + \bar{x} = 30,54$$

$$S_x^2 = S_x^2 = 5,3620$$

$$S_X = \sqrt{5,3629} = 2,3156$$

$$\bar{y} = \frac{\sum y}{n} = \frac{-60}{100} = -0,60$$

$$S_y^2 = \frac{\sum y^2 - n\overline{y}^2}{n-1} = \frac{350 - 100(-0,60)^2}{99} = 3,1717$$

$$S_y = \sqrt{3,1717} = 1,7809$$

$$\overline{Y} = 300 + 200 \cdot \overline{y} = 3080g$$

$$S_y^2 = 200^2 \cdot S_y^2 = 126868.69$$

$$S_Y = \sqrt{126869,69} = 356.1863g$$

1) Tính hệ số tương quan thực nghiệm:

$$R(X,Y) = R(x,y) = \frac{\sum xy - n \cdot \overline{x} \cdot \overline{y}}{(n-1)X_x S_y}$$

$$R(X,Y) = \frac{270 - 100(0,54)(-0,60)}{99(2.3156)(1.7809)} = 0,74$$

Nhận xét rằng sự tương quan của (X, Y) là tương quan thuận và chặc.

2) Phương trình hồi quy:

$$a = R\frac{S_Y}{S_X} = (0,74)\left(\frac{356,1863}{2,3156}\right) = 113,83$$

$$b = \overline{Y} - a\overline{X} = 3080 - (113,83)(30,54) = -396,37$$

Phương trình đường thẳng hồi quy: Y = (113,83)X – 396,37

Với X_0 = 36cm, ta có: Y_e = (113,83)(36) – 396,37 = 3701,51g

Khoảng dự trù cho Y ở độ tin cậy 95%:

$$Y = Y_e \pm CS_{Y \cdot X}\sqrt{1 + \frac{1}{n} + \frac{(X - X_0)^2}{(n-1)S_X^2}}$$

$$Y = 3701,51 \pm (1,96)(240,79)\sqrt{1 + \frac{1}{100} + \frac{(36 - 30,54)^2}{99(5,3620)}}$$

$$Y = 3701,51 \pm 487,31\text{g}$$

Trọng lượng nhẹ nhất của đứa con là:

$$Y\text{min} = 3701,51 - 487,31 = 3214,20\text{g} \approx 3214\text{g}$$

[2] Số lượng sau đây là thể trọng P(kg) và đường huyết H (mg/mℓ) trên 16 người lớn khoẻ mạnh:

Người	1	2	3	4	5	6	7	8
P	64	75,3	73	82,1	76,2	95,7	59,4	93,4
H	108	109	104	102	105	121	79	107

Người	9	10	11	12	13	14	15	16
P	82,1	78,9	76,7	82,1	83,9	73	64,4	77,6
H	101	85	99	100	108	104	102	87

1) Có phải thể trọng càng nặng thì đường huyết càng cao?

2) Những người nặng 75kg thì đường huyết trung bình của họ ít nhất là bao nhiêu với độ tin cậy 95%?

BÀI GIẢI

Ta có:

$$\sum P = 1238 \; ; \; \sum P^2 = 97178,60$$

$$\sum H = 1621 \; ; \; \sum H^2 = 165801 \; ; \; \sum PH = 126128,10$$

Vậy: $\overline{P} = 77,36\text{kg}$; $\overline{H} = 101,31(\text{mg}/100\text{m}\ell)$

$$S_P^2 = 94,62(kg)^2 \quad ; S_H^2 = 104,90(mg/100m\ell)^2$$

$$S_P = 9,73kg \quad ; S_H = 10,24(mg/100m\ell)$$

1) Ta có bài toán kiểm định:

$$\begin{cases} H : \rho = 0 \\ K : \rho > 0 \end{cases}$$

Ta có:

$$R = \frac{\sum PH - n \cdot \overline{P} \cdot \overline{H}}{(n-1)S_p \cdot S_H}$$

$$R = \frac{126128,10 - 16(77,36)(101,31)}{15(9,73)(10,24)} = 0,49$$

$$T = \frac{R}{\sqrt{1 - R^2}}\sqrt{n-1} \sim \text{Student } (n - 2 = 16 - 2 = 14)$$

$$T = \frac{0,49}{\sqrt{1 - (0,49)^2}}\sqrt{14} = 2,10$$

Nếu H đúng thì T ~ Student (14), vậy:

$$\alpha = 0,05 = P(T > C|H) \Rightarrow C = 1,76$$

Kết luận: T > C nên từ chối H, nghĩa là ρ > 0, vậy thể trọng càng nặng thì đường huyết càng cao.

2) Ta có: $a = R\frac{S_H}{S_P} = 0,49\left(\frac{10,24}{9,73}\right) = 0,52$

$$b = \overline{H} - a \cdot \overline{P} = 101,31 - (0,52)(77,36) = 61,08$$

Phương trình hồi quy của đường huyết thẹo thể trọng là:

$$\boxed{H = 0,52)P + 61,08} \ mg/100m\ell$$

Với: $P_0 = 95kg$, ta có:

$$H = (0,52)95 + 61,08 = 110,48 (mg/100m\ell)$$

Ta có: $$S^1_{H \cdot P} = \frac{n-1}{n-2}(1 - R^2)S^2_H$$

$$S^2_{H \cdot P} = \frac{15}{14}(1 - (0,49)^2)(104,90) = 85,41$$

$$\Rightarrow S_{H \cdot P} = 9,24$$

Khoảng ước lượng của đường huyết trung bình:

$$\mu_{H|P_0} = H_0 \pm C \cdot S_{H \cdot P} \times \sqrt{\frac{1}{n} + \frac{(P_0 - \overline{P})^2}{(n-1)S^2_P}}$$

$$\mu_{H|P_0} = 110,48 \pm (2,14)(9,24)\sqrt{\frac{1}{16} + \frac{(95 - 77,36)^2}{15(94,62)}}$$

$$\mu_{H|P_0} = 110,48 \pm 10,50 (mg/100m\ell)$$

Đường huyết trung bình ít nhất của những người có thể trọng 95kg là:

$$(\mu_{H|P_0})_{min} = 99,98 (mg/100\ell)$$

[3] Nghiên cứu về ảnh hưởng của stress lên huyết áp. Chín con vật thử nghiệm có cùng huyết áp ban đầu, được gây stress bằng sốc điện với cường độ X, sau 2 phút, đo huyết áp Y:

Cường độ X	30	30	30	50	50	50	70	70	70
Huyết áp Y	125	139	120	150	145	160	175	180	180

1) Vẽ đồ thị phân tán.

2) Xác định đường thẳng hồi quy của huyết áp Y theo cường độ.

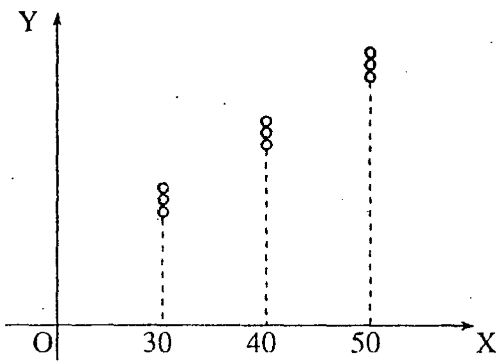
   Với cường độ là 60 thì huyết áp trung bình tối đa là bao nhiêu với độ tin cậy 95%.

3) Có phải cường độ stress gia tăng thí huyết áp cũng gia tăng?

BÀI GIẢI

X	30	30	30	50	50	50	70	70	70
Y	125	139	120	150	145	160	175	180	180

1) Vẽ đồ thị phân tán:



2) $n = 9$; $\sum X = 450$; $\sum Y = 1365$

$\sum X^2 = 24900$; $\sum XY = 71450$; $\sum Y^2 = 211475$

$$\overline{X} = \frac{\sum X}{n} = \frac{450}{9} = 50$$

$$S_X^2 = \frac{\sum X^2 - n\overline{X}^2}{n-1} = \frac{24900 - 9(50)^2}{8} = 300$$

$S_X = 17,32$

$$\overline{Y} = \frac{\sum X}{n} = \frac{1365}{9} = 151,67$$

$$S_Y^2 = \frac{\sum Y^2 - n\overline{Y}^2}{n-1} = \frac{211475 - 9(151,67)^2}{8} = 556,25$$

$S_Y = 23,58$

$$R = \frac{\sum XY - n\overline{X} \cdot \overline{Y}}{(n-1)S_Y S_X} = \frac{71450 - 9(50)(151,67)}{8(17,32)(23,58)} = 0,9789$$

$$a = R\frac{S_Y}{S_X} = (0,98)\frac{23,58}{17,32} = 1,33$$

$$b = \overline{Y} - a\overline{X} = 151,67 - (1,33)(50) = 85,03$$

Phương trình hồi quy: $Y = (1,33)X + 85,03$

<u>$X = 60$</u>: $Y_e = (1,33)(60) + 85,03 = 164,83$

$$\mu_{Y|X_0} = Y_e \pm C \cdot S_{Y \cdot X}\sqrt{\frac{1}{2} + \frac{(X_0 - X)^2}{(n-1)S_X^2}}$$

$$S_{Y \cdot X}^2 = \frac{8}{7}(1 - 0,98^2)(556,25) = 25.17$$

$S_{Y \cdot X} = 5,02$

$$\mu_{Y|X_0} = 164,83 \pm (2,36)(5,02)\sqrt{\frac{1}{9} + \frac{(50-60)^2}{8(300)}}$$

$$\mu_{Y|X_0} = 164,83 \pm 4,63 \text{ mmHg}$$

$$\left(\mu_{Y|X_0}\right)_{max} = 169,46 \text{ mmHg}$$

3) Ta có bài toán kiểm định:

$$\begin{cases} H: \rho = 0 \\ K: \rho > 0 \end{cases}$$

$$T = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} = \frac{0,98}{\sqrt{1-(0,98)^2}}\sqrt{7} = 13,03$$

Nếu H đúng thì T ~ Student(7), vậy:

$$\alpha = 0,05 = P(T > C|H) \Rightarrow C = 1,89$$

Kết luận: T > C vậy từ chối H, nghĩa là cường độ stress gia tăng thì huyết áp cũng tăng.

[4] Một thí nghiệm nghiên cứu ảnh hưởng của việc gia tăng liều dùng X(mg/kg) của một loại Barbiturate trên thời gian ngủ Y (giờ).

X	2	2	2	3	3	3	5	5	5
Y	4	6	5	9	8	7	13	11	9

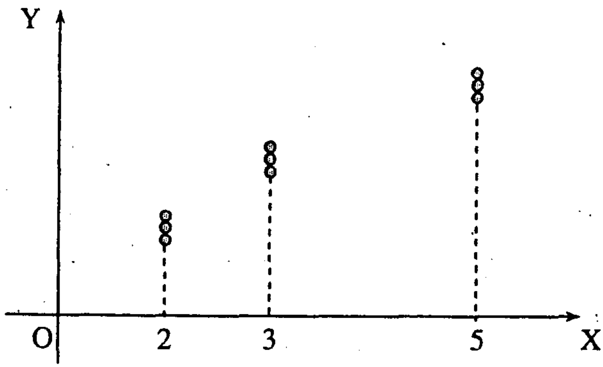
1) Vẽ đồ thị phân tán. (scatter diagram).

2) Tính hệ số tương quan R. Tìm KTC 95% của ρ_{xy}.

3) Tính đường thẳng hồi quy của Y theo X. Với liều dùng 4 mg/kg thì thời gian ngủ lâu nhất tiên đoán là bao nhiêu với độ tin cậy 95%.

BÀI GIẢI

1) Vẽ đồ thị phân tán:



2) Tính hệ số tương quan thực nghiệm:

$n = 9 \quad \overline{X} = 3,33 \quad S_X = 1,2329$

$\overline{Y} = 8,00 \quad S_Y = 2,8723$

$\sum XY = 267$

$$R = \frac{\sum XY - n \cdot \overline{X} \cdot \overline{Y}}{(n-1)S_X S_Y}$$

$$R = \frac{267 - 9(3,33)(8)}{8(1,3229)(2,8723)} = 0,89$$

$$Z_R = \frac{1}{2}\ln\left(\frac{1+R}{1-r}\right) = \frac{1}{2}\ln\left(\frac{1+0,89}{1-0,89}\right) = 1,4134$$

Khoảng ước lượng của $Z = \dfrac{1}{2}\ln\left(\dfrac{1+\rho}{1-\rho}\right)$ là:

$$Z = Z_R \pm \frac{C}{\sqrt{n-3}} = 1,4131 \pm \frac{1,96}{\sqrt{6}}$$

$$Z = 1,4131 \pm 0,8002$$

$$0,6129 \leq Z \leq 2,2133$$

Do đó: $0,5562 \leq \rho \leq 0,9764$

3) Phương trình đường thẳng hồi quy:

$$a = \frac{S_Y}{S_X} = (0,89)\left(\frac{2,8723}{1,3229}\right) = 1,93$$

$$b = \overline{Y} - \overline{X} = 1,56$$

Phương trình đường thẳng hồi quy:

$$Y = (1,93)X + 1,56$$

Với X_0 = 4mg/kg; ta có: $Y_e = (1,93)(4) + 1,56 = 9,28$

$$S^2_{Y \cdot X} = \frac{n-1}{n-2}(1-R)^2 S^2_Y$$

$$S^2_{Y \cdot X} = \frac{8}{7}(1-(0,89)^2)(2,8723)^2 = 1,9602$$

$$S_{Y \cdot X} = 1,40$$

Khoảng ước lượng của Y:

$$Y = Y_e \pm CS_{Y \cdot X}\frac{\sqrt{1 + \frac{1}{n} + (\overline{X} - X_0)^2}}{(n-1)S^2_X}$$

$$Y = 9.28 \pm (2,36)(1,40)\sqrt{1 + \frac{1}{9} + \frac{(4-3,33)^2}{8(1,3229)^2}}$$

$Y = 9{,}28 \pm 3{,}54$ giờ

Do đó: $Y_{max} = 9{,}28 + 3{,}54 = 12{,}82$ giờ

[5] Một mẫu gồm 32 bé, gọi X(oz) là trọng lượng mới sinh, Y là độ gia tăng trọng lượng lúc 3 tháng tuổi biểu thị bằng tỉ lệ % so với X. Cho:

$$\sum X = 3576; \quad \sum x^2 = 409880$$

$$\sum Y = 2281; \quad \sum Y^2 = 179761 \quad \sum XY = 246032$$

1) Tính hệ số tương quan Pearson của X và Y. Có phải X và Y là tương quan nghịch?

2) Tìm phương trình đường thẳng hồi quy của Y theo X. Một bé lúc mới sanh năng 100 oz, tìm khoảng tin cậy 95% trọng lượng bé khi được 3 tháng tuổi (1 oz = 28,35g)

BÀI GIẢI

1) $$\overline{X} = \frac{\sum X}{n} = \frac{3576}{32} = 111{,}75$$

$$S_X^2 = \frac{\sum X^2 - n\overline{X}^2}{n-1} = \frac{409880 - 32(111{,}75)^2}{31} = 331{,}03$$

$$S_X = 18{,}19$$

$$\overline{Y} = \frac{\sum Y}{n} = \frac{2281}{32} = 71.28$$

$$S_Y^2 = \frac{\sum Y^2 - n\overline{Y}^2}{n-1} = \frac{179761 - 32(71{,}28)^2}{31} = 553{,}82$$

$$S_Y = 23{,}53$$

$$R = \frac{\sum XY - n\overline{X}\cdot\overline{Y}}{(n-1)S_X \cdot S_Y} = \frac{246032 - 32(111,75)(71,28)}{31(18,19)(23,53)} = -0,67$$

Ta có bài toán kiểm định:

$$\begin{cases} H : \rho = 0 \\ K : \rho < 0 \end{cases}$$

$$T = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} = \frac{-0,67}{1-0,67^2}\sqrt{30} = -4,94$$

Nếu H đúng thì T ~ Student (30), do đó:

$$\alpha = 0,05 = P(T < C|H) \Rightarrow C = -1,65$$

Kết luận: Vì T< C nên từ chối H nghĩa là ρ < 0.

2) $$a = R \times \frac{S_Y}{S_X} = -0,67 \times \frac{23,53}{18,19} = -0,87$$

$$b = \overline{Y} - a\overline{X} = 71,28 - (-0,87)(111,75) = 168,13$$

Phương trình hồi quy: $Y = (-0,87)X + 168,13$

<u>$X_0 = 100$:</u>

$Y_0 = (0,87)(100) + 168,13 = 81,13$

$$S^2_{Y\cdot X} = \frac{n-1}{n-2}(1-R^2)S^2_Y = \frac{31}{30}(1-0,67^2)(553,82) = 315,38$$

$$S_{Y\cdot X} = 17,76$$

$$Y = Y_0 \pm C \cdot S_{Y\cdot X}\sqrt{1 + \frac{1}{n} + \frac{(\overline{X} - X_0)^2}{(n-1)S^2_X}}$$

$$Y = 81,13 \pm (2,04)(17,76)\sqrt{1 + \frac{1}{32}\frac{(111,75-100)^2}{31(331,03)}}$$

$$Y = 81,13 \pm 37,07 \Leftrightarrow 44,06 \le Y \le 118,20$$

Ta có: $\dfrac{\Delta X}{X} = \dfrac{Y}{100} \Rightarrow \dfrac{X+\Delta X}{X} = \dfrac{100+Y}{100}$

$$X + \Delta X = \left(\frac{100+Y}{100}\right) \times X$$

$$P_1 = \left(\frac{100+44,06}{100}\right) \times 100 = 144,06(\text{oz})$$

$$P_2 = \left(\frac{100+118,20}{100}\right) \times 100 = 218,20(\text{oz})$$

$$144,06 \le p \le 218,20(\text{oz})$$

[6] Nghiên cứu thải trừ một số loại thuốc, số liệu như sau:

(X = giờ, Y = nồng độ thuốc: μg/ml)

X	0,5	0,5	1	1	2	2	3	3	4	4
Y	0,42	0,45	0,35	0,33	0,25	0,22	0,20	0,20	0,15	0,17

1) Vẽ đồ thị phân tán.

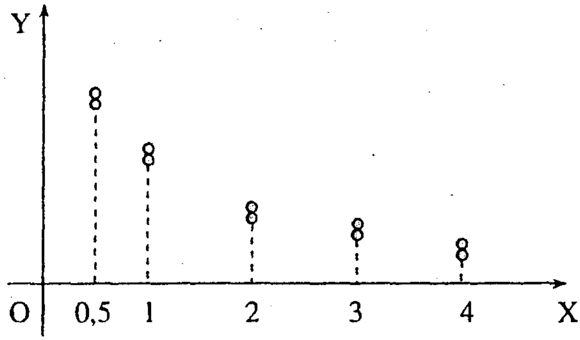
2) Tìm phương trình đường thẳng hồi quy của nồng độ theo thời gian.

3) Kiểm định giả thiết: "Không có sự liên quan giữa hai biến X và Y" (kết luận với $\alpha = 5\%$).

4) Sau 2 giờ nồng độ thuốc là bao nhiêu?

BÀI GIẢI

1) Vẽ đồ thị phân tán:



X	Y	X^2	XY	Y^2
0,5	0,42	0,25	0,21	0,1764
0,5	0,45	0,25	0,225	0,2025
1	0,35	1,00	0,35	0,1225
1	0,33	1,00	0,33	0,1089
2	0,25	4,00	0,50	0,0625
2	0,22	4,00	0,44	0,0484
3	0,20	9,00	0,60	0,0400
3	0,20	9,00	0,60	0,0400
4	0,15	16,00	0,60	0,0225
4	0,17	16,00	0,68	0,0289
21	2,74	60,50	4,5350	0,8526

2) Phương trình hồi quy:

Hệ phương trình xác định hệ số hồi quy:

$$\begin{cases} a\left(\sum X\right) + nb = \sum Y \\ a\left(\sum X^2\right) + b\left(\sum X\right) = \sum XY \end{cases}$$

$$\begin{cases} 21 \cdot a & + & 10b & = 2,74 \\ (60,5)a & + & 21b & = 4,5350 \end{cases}$$

$$\begin{cases} (41,1) \cdot a & + & (21) \cdot b & = 5,7540 \\ (60,5) \cdot a & + & 21 \cdot b & = 4,5350 \end{cases}$$

$$(16,40).a = -1,2190 \Rightarrow a = -0,07 \ ; \ b = 0,43$$

Phương trình hồi quy: $Y = -(0,07)X + 0,43$

3) Tính hệ số tương quan thực nghiệm:

Ta có: $S_X^2 = \dfrac{\sum X^2 - n \cdot \overline{X}^2}{n-1}$

$$S_X^2 = \frac{60,50 - 10(2,10)^2}{9} = 1,8222$$

$$S_X = \sqrt{1,8222} = 1,35$$

$$S_Y^2 = \frac{\sum Y^2 - n \cdot \overline{Y}^2}{n-1}$$

$$S_Y^2 = \frac{0,8526 - 10(0,724)^2}{9} = 0,0113$$

$$S_Y = \sqrt{0,0113} = 0,1063$$

$$R(X,Y) = \frac{\sum XY - n \cdot \overline{X} \cdot \overline{Y}}{(n-1)S_X S_Y}$$

$$R(X,Y) = \frac{4.5350 - 10(2,10)(0,274)}{9(1,35)(0,1063)} = -0,94$$

Ta có bài toán kiểm định:

$$\begin{cases} H : \rho = 0 \\ K : \rho \neq 0 \end{cases}$$

$$T = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} = \frac{-0,94}{\sqrt{1-(0,94)^2}}\sqrt{8} = -7,79$$

Nếu H đúng, T ~ Student (8), do đó:

$$\sigma = 0,05 = P(|T| > C|H) \Rightarrow C = 2,31.$$

Kết luận: Vì $|T| > C$ nên từ chối H, nghĩa là (X, Y) thực sự có tương quan.

Với $X_0 = 2$ giờ, ta có: $Y_e = -(0,07)(2) + 0,43 = 0,29$.

[7] Đo chiều cao X (cm) và chiều dài chi dưới Y (cm) của 8 thanh niên ở TP. HCM ta có:

X	156	158	160	162	164	166	168	170
Y	72	74	77	78	79	82	83	85

1) Tính hệ tương quan thực nghiệm R(X, Y)

2) Lập phương trình hồi quy của Y theo X. Tính phương sai hồi quy $S^2_{Y \cdot X}$

3) Một người có chiều cao 161cm, tìm chiều dài chi dưới của người đó ở độ tin cậy 95%.

BÀI GIẢI

Ta có: n = 8; $\overline{X} = 163$cm $\quad S_X = 4,90$cm

$\overline{Y} = 78,75$cm $\quad S_Y = 4,46$cm

$\sum XY = 102842$

1) Tính hệ số tương quan thực nghiệm:

$$R(X,Y) = \frac{\sum XY - n\cdot\overline{X}\cdot\overline{Y}}{(n-1)S_Y}$$

$$R(X,Y) = \frac{102842 - 8(163)(78,75)}{7(4,90)(4,46)} = 0,99$$

2) Phương trình hồi quy:

$$a = R\frac{S_Y}{S_X} = (0,99)\left(\frac{4,46}{4,90}\right) = 0,90$$

$$b = \overline{Y} - a\cdot\overline{X} = 78,75 - (0,90)(163) = -67,95$$

Phương trình đựpờng thẳng hồi quy: Y = (0,90)X – 67,95

Phương sai hồi quy:

$$S^2_{Y\cdot X} = \frac{n-1}{n-2}\left(1 - R^2\right)S^2_Y$$

$$S^2_{Y\cdot X} = \frac{7}{6}\left(1 - (0,99)^2\right)(4,46)^2 = 0,4618$$

$$S_{Y\cdot X} = 0,68$$

3) Khoảng ước lượng của Y:

Với: X_0 = 161cm, ta có:

Y_0 = (0,90)(161) – 67,95 = 96,95cm

$$Y = Y_e \pm CS_{Y\cdot X}\sqrt{1 + \frac{1}{n} + \frac{(X - X_0)^2}{(n-1)S^2_X}}$$

$$Y = 76,95 \pm (2,45)(0,68)\sqrt{1 + \frac{1}{8} + \frac{(163-161)^2}{7(4,9)^2}}$$

(Giá trị C đọc trong bảng Student (6), cột 0,05)

$Y = 76,95 \pm 1,79cm$

[8] Đo chiều cao bụng X(cm) và trọng lượng trẻ mới sinh Y(gam) của 100 sản phụ, rồi rút gọn số liệu qua phép đổi biến số:

$$x = X - 30 \; ; \; y = \frac{Y - 3200}{200}$$

Ta tính được: $\sum x = 54$; $\sum y = -60$

$$\sum x^2 = 560 \; ; \; \sum xy = 270 \; ; \; \sum y^2 = 350$$

1) Người ta cho rằng trọng lượng trung bình trẻ sơ sinh là 3000gam. Hỏi: kết quả quan sát có phù hợp với giá trị cho biết không? (Kết luận với $\alpha = 5\%$)

2) Tính hệ số tương quan thực nghiệm R(X, Y) giữa chiều cao bụng và trọng lượng trẻ sơ sinh. (X, Y) thực sự có tương quan không? Kết luận với $\alpha = 5\%$).

3) Tìm phương trình đường thẳng hồi quy của Y theo X.

4) Theo Johnson thì phương trình đường thẳng hồi quy của Y theo X là:

   $Y = 155(X - n)$, n lấy giá trị 11 hay 12 tùy theo độ lọt.

5) Một sản phụ đang chuyển dạ có bề cao bụng là 36cm thì đứa con có trọng lượng nhẹ nhất là bao nhiêu với độ tin cậy 99%.

BÀI GIẢI

$$\bar{x} = \frac{\sum x}{n} = \frac{54}{100} = 0,54$$

$$S_x^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{560 - 100(0,54)^2}{99} = 5,3620$$

$$S_x = \sqrt{5,3620} = 2,3156$$

$$\overline{X} = 30 + \bar{x} = 30,54$$

$$S_X^2 = S_x^2 = 5,3620$$

$$S_X = \sqrt{5,3620} = 2,3156$$

$$\bar{y} = \frac{\sum y}{n} = \frac{-60}{100} = -0,60$$

$$S_y^2 = \frac{\sum y^2 - n\bar{y}^2}{n-1} S_y^2 = \frac{350 - 100(0,60)^2}{99} = 3,1717$$

$$S_y = \sqrt{3,1717} = 1,7809$$

$$\overline{Y} = 3000 + 200 \cdot \bar{y} = 3080g$$

$$S_Y^2 = 200^2 \cdot S_y^2 = 126868,6869$$

$$S_Y = \sqrt{126868,6869} = 356,1863g$$

1) Ta có bài toán kiểm định:

$$\begin{cases} H : Mu_y = 3000g \\ H : \mu_Y \neq 3000g \end{cases}$$

$$T = \frac{\overline{Y} - \mu_0}{\frac{S_Y}{\sqrt{n}}}$$

$$T = \frac{3080 - 3000}{\frac{356,19}{\sqrt{100}}} = 2.25$$

Nếu H đúng thì T ~ Student (99). Do đó:

$$\alpha + 0,05 = P(|T| > C|H) \Rightarrow C = 1,96$$

Kết luận: Vì $|T| > C$ nên từ chối H, nghĩa là $\mu_Y \neq 3000g$

2) Tính hệ số tương quan thực nghiệm:

$$R(X, Y) = R(x, y) = \frac{\sum xy - n \cdot \overline{x} \cdot \overline{y}}{(n-1)S_x S_y}$$

$$R(x, Y) = \frac{270 - 100(0,54)(-0,60)}{99(2,32)(1,78)} = 0,74$$

3) Phương trình hồi quy:

$$a = R\frac{S_Y}{S_X} = (0,74)\left(\frac{356,1863}{2,3256}\right) = 113,83$$

$$b = \overline{Y} - a \cdot \overline{X} = 3080 - (113,83)(30,54) = -396,37$$

Phương trình đường thẳng hồi quy: Y = (113,83)X − 396,37

4) Ta có bài toán kiểm định:

$$\begin{cases} H : \alpha = 155 \\ K : \alpha \neq 155 \end{cases}$$

$$S^2_{X \cdot Y} = \frac{n-1}{n-2}(1 - R^2)S^2_Y$$

$$= \frac{99}{98}\left(1-(0,74)^2\right)(126868,69) = 57981,06$$

$$S_{Y \cdot X} = \sqrt{57981,06} = 240,79$$

$$T = \left(\hat{a} - a_0\right) S_X \frac{\sqrt{n-1}}{S_{Y \cdot X}}$$

$$T = \frac{(113,83-155)(2,3156)\sqrt{99}}{240,79} = -3,94$$

Nếu H đúng thì T ~ Student (98), do đó:

$$\alpha = 0,05 = P(|T| > C|H) \Rightarrow C = 1,96$$

Kết luận: Vì $|T| > C$ nên từ chối H, nghĩa là $a \neq 155$, do đó phương trình hồi quy tìm được khác với phương trình Johnson.

Với $X_0 = 36$cm, ta có: $Y_e = (113,72)(36) - 393 = 3700,92$g

Khoảng dự trù cho Y ở độ tin cậy 99%:

$$Y = Y_e \pm C S_{Y \cdot X} \sqrt{1 + \frac{1}{n} + \frac{(X - X_0)^2}{(n-1)S_X^2}}$$

$$Y = 3700,92 \pm (2,58)(240,79)\sqrt{1 + \frac{1}{100} + \frac{(36-30,54)^2}{99(5,3620)}}$$

$$Y = 3700,92 \pm 641,46\text{g}$$

Trọng lượng nhẹ nhất của đứa con là:

$$Y_{min} = 3700,92 - 641,46 = 3059,46 \approx 3059\text{g}$$

VIII. Bài tập XÁC SUẤT TRONG CHẨN ĐOÁN

[1] Một bà 70 tuổi đến khám vì mệt, đau ở tay và xương gối, thỉnh thoảng đau nhói ở ngực. Khám LS không có gì đáng chú ý, các khớp không sưng. BS trực nghi ngờ bị SLE (Systemic lupus erythematosus). Vấn đề là có nên cho làm ANA test hay không (Antinuclear antibody). Năm 1982 Tan et al báo cáo là ANA test có sens 95% và Spec 50% vì rằng ANA dương tính trong nhiều bệnh về mô liên kết và dương tính tăng theo tuổi.

a) Cho rằng trong đa số bà này có khã năng bị SLE là 2%. Nếu ANA dương tính thì khả năng SLE là bao nhiêu?

b) Giả sử BN này thêm các khớp bị sưng, khi ấy khả năng SLE trên nhóm BN này tăng lên 20%. Tính khả năng người này bị SLE khi ANA^+, ANA^-.

BÀI GIẢI

$$LR^+ = \frac{\text{sens}}{1-\text{spec}} = \frac{0,95}{1-0,50} = 1,9$$

$$LR^- = \frac{1-\text{sens}}{\text{spec}} = \frac{1-0,95}{0,50} = 0,1$$

Áp dụng công thức:

$$od(posttest^+) = od(pretest).LR^+)$$

a) $$od(SLE \mid ANA^+) = od(SLE).LR^+$$

$$= \frac{P(SLE)}{1-P(SLE)} \cdot LR^+$$

$$= \frac{0,02}{1-0,02} \cdot 1,9$$

$= 0{,}0388$

$$P(SLE \mid ANA^{+}) = \frac{0d(SLE \mid ANA^{+})}{1 - od(SLE \mid ANA^{+})}$$

$$= \frac{0,0388}{1 + 0,0388}$$

$$= 3{,}7\%$$

b) Lập luận tương tự với od(Pretest) = od(SLE) = $\frac{1}{4}$

Ta được kết quả:

$P(SLE \mid ANA^{+}) = 0{,}32$

$P(SLE \mid ANA^{-}) = 0{,}025$

[2] Một ông 43 tuổi đến khám tổng quát. Phân tích nước tiểu thấy có đường niệu. Bạn vừa biết có một test T dùng phát hiện bệnh tiểu đường như sau: 138 cho T^{+} trên 150 người bị tiểu đường và 24 cho T^{+} trên 150 người không bị tiểu đường.

a) Tìm độ nhạy, độ chuyên, dương giả, âm giả của T.

b) Bằng phương pháp thử máu (T') có độ nhạy 0,80; độ chuyên 0,96, đem áp dụng cho 2 nhóm trên thì giá trị đoán trước là bao nhiêu nếu T' dượng tính.

c) Với ông này, nếu T dương tính, bạn cho là có khả năng 90% bị tiểu đường. Còn nếu T' cũng dương tính thì khả năng bị tiểu đường là bao nhiêu?

BÀI GIẢI

a) Độ nhạy $= \text{sens} = \dfrac{138}{150} = 0{,}92$; Âm giả = 0,08

Độ chuyên $= \text{spec} = \dfrac{126}{150} = 0{,}84$; Dương giả = 0,16

b) Dùng bảng 2 × 2:

	B^+	B^-	
T'^+	a = 120	b = 6	126
T'^-	c = 30	d = 144	174
	150	150	300=a+b+c+d

$$\text{sens} = 0{,}80 = P(T^+|B^+) = \frac{P(T^+B^+)}{P(B^+)}$$

$$\Rightarrow P(B^+T^+) = 0{,}08 \cdot P(B^+)$$

$$= 0{,}80\left(\frac{150}{300}\right) = 0{,}40$$

$$\Rightarrow a = 0{,}40\ (a + b + c + d)$$

$$= 0{,}40(300) = 120$$

Tương tự: $\Rightarrow b = 6$; $c = 30$; $d = 144$

Nếu T' dương tính thì

$$P(B^+|T'^+) = \frac{P(B^+T'^+)}{P(T'^+)} = \frac{120}{126} = 95{,}2\%$$

c) Nếu T dương tính $P(B^+) = 0{,}90 \Rightarrow P(B^+|T'^+) = 0{,}994$

[3] Một bà 22 tuổi đến khám vì hồi hộp. Khám LS không phát hiện được gì. Theo kinh nghiệm, trong trường hợp này bạn cho là khả năng 20 – 30% bị sa van tim (MVP) (mitral valve prolapse). bạn biết là Echocardiogram có độ nhạy 90% và độ chuyên 95%.

a) Nếu bà này có Echocardiogram dương tính thì khả năng bị MVP là bao nhiêu?

b) Nếu Echocardiogram âm tính thì bạn tin bà này không bị NVP là bao nhiêu?

BÀI GIẢI

Chọn $P(B^+) = 0{,}20 \Rightarrow od(B^+) = \dfrac{0{,}20}{0{,}80} = \dfrac{1}{4}$

$B^+ = MVP$

$$LR^+ = \frac{\text{sens}}{1-\text{spec}} = \frac{0{,}90}{1-0{,}95} = 1{,}8$$

$$LR^- = \frac{1-\text{sens}}{\text{spec}} = \frac{1-0{,}90}{0{,}95} = 0{,}105$$

T = Echocardiogram

a) $od(B^+|T^+) = od(B^+)\cdot LR^+ = \dfrac{1}{4}\cdot 18 = \dfrac{9}{2}$

$$\Rightarrow P(B^+|T^+) = \frac{od(B^+|T^+)}{1+od(B^+|T^+)} = \frac{\frac{9}{2}}{1+\frac{9}{2}} = \frac{9}{11} = 0{,}82$$

Vậy: $P(MVP|T^+) = 0{,}82$

b) $od(B^+|T^-) = od(B^+)\cdot LR^-$

$$= \frac{1}{4}(0,105) = 0,02625$$

$$\Rightarrow P(B^+|T^-) = \frac{0,02615}{1+0,02615} = 0,0256$$

$$\Rightarrow P(B^-|T^-) = 0,9744$$

[4] Một xét nghiệm chẩn đoán có độ nhạy 90%, độ chuyên 95%. Tính giá trị tiên đoán trong bảng sau đây:

Prevalence	Predictive value of positive test	Predictive value of negative test
0,1%	0,018	0,001
1%	0,154	0,001
2%	0,269	0,002
5%	0,486	0,005
10%	0,667	0,012
20%	0,828	0,026
50%	0,947	0,095
80%	0,986	0,296

BÀI GIẢI

Prevalence = $P(B^+)$

Predictive value of positive test = $P(B^+|T^+) = PV^+$

Predictive value of negative test = $P(B^+|T^-) = PV^-$

$$PV^+ = P(B^+|T^+)$$

$$= \frac{P(B^+)\cdot P(T^+|B^+)}{P(T^+)}$$

$$= \frac{P(B^+) \cdot sens}{P(B^+) \cdot sens + (1 - P(B^+) \cdot (1 - spec)}$$

$$PV^+ = \frac{1}{1 + \frac{1+P(B^+)}{P(B^+)} \cdot \frac{1-spec}{sens}}$$

$$PV^- = P(B^+ | T^-)$$

$$= \frac{P(B^+) \cdot P(T^- | B^+)}{P(T^-)}$$

$$= \frac{P(B^+)(1 - sens)}{P(B^+)(1 - sens) + (1 - P(B^+)) \cdot spec}$$

$$PV^- = \frac{1}{1 + \frac{1-P(B^+)}{P(B^+)} \cdot \frac{spec}{1-sens}}$$

Thay số vào ta sẽ có kết quả trong bảng.

[5] Trong một dân số phụ nữ có 4% bị K vú, 20% có hút thuốc, 3% có hút thuốc và bị K vú. Một phụ nữ đến từ dân số đó, tính xác suất người này bị K vú hoặc có hút thuốc hoặc có cả hai.

BÀI GIẢI

Đặt: K : biến cố bị K vú

H : biến cố có hút thuốc.

$P(K) = 0{,}04$

$P(H) = 0{,}20$

$P(KH) = 0{,}03$

Bài toán yêu cầu tính $P(K \cup H)$

$$P(K \cup H) = P(K) + P(H) - P(KH)$$
$$= 0{,}04 + 0{,}02 - 0{,}03$$
$$= 0{,}21$$
$$P(K \cup H) = 0{,}21$$

[6] Một người đến khám vì sốt. Theo tổng kết của phòng khám thì những bệnh nhân như vậy có thể do: cảm cúm 30%, thương hàn 10%, sốt rét 20%, viêm đường hô hấp 30%, còn lại là do bệnh khác.

a) Tính khả năng người này bị sốt rét?

b) Khám LS thấy người này có lách to. Biết rằng tỷ lệ lách to trong sốt rét là 70% còn trong những bệnh không phải sốt rét là 20%. Tính khả năng người này bị sốt rét khi khám LS thấy có dấu hiệu lách to?

c) Cho làm xét nghiệm (XN) K (CTM), kết quả K dương tính (bạch cầu tăng). Tỷ lệ bạch cầu tăng trong các bệnh trên theo thứ tự là 0,5; 0,2; 0,4; 0,8; 0,7. Tính khả năng người này bị sốt rét khi dựa vào kết quả K?

d) Cho hàm XN H (Tìm KSTST), kết quả H dương tính. Giả sử phòng XN cho biết, Tỉ lệ H dương tính trong các bệnh trên theo thứ tự là: 3%; 5%; 80%; 5%; 4%. Tính khả năng người này bị SR khi biết thông tin này.

e) Nếu kết hợp LS và XN thì khả năng người này SR là bao nhiêu? Cho rằng các thông tin trên LS và CLS là độc lập.

BÀI GIẢI

a) Khi chưa có thông tin gì về LS và CLS, chỉ phán đoán theo tổng kết của phòng khám: P(SR) = 0,20

b) Ta có: $P(SR) = 0,20$; $P(LT|SR) = 0,70$

$$P(LT|\overline{SR}) = 0,20$$

$$P(SR|LT) = \frac{P(SR)\cdot P(LT|SR)}{P(LT)}$$

$$= \frac{P(SR)\cdot P(LT|SR)}{P(SR)\cdot P(LT|SR) + P(\overline{SR})\cdot P(LT|\overline{SR})}$$

$$= \frac{0,2(0,7)}{0,2(0,7)+0,8(0,2)}$$

$$= \frac{0,14}{0,30} = 0,47$$

$$P(SR|LT) = 0,47$$

c) $$P(SR|K) = \frac{P(SR)\cdot (K|SR)}{P(K)}$$

Bệnh	D_i	$P(D_i)$	$(P(K\|D_i)$	$P(D_i).\ P(K\|D_i)$
Cảm	D_1	0,30	0,50	0,15
Thương hàn	D_2	0,10	0,20	0,20
Sốt rét	D_3	0,20	0,40	0,08
Viêm hô hấp	D_4	0,30	0,80	0,24
Bệnh khác	D_5	0,10	0,70	0,07
				0,56

$$P(K) = \sum P(Di)\cdot P(K|Di)$$

$$= 0,56$$

$$P(SR|K) = \frac{P(SR)\cdot P(K|SR)}{P(K)} = \frac{0,20(0,40)}{0,56} = \frac{0,08}{0,56} = \frac{0,08}{0,56} = 0,143$$

d) Tương tự ta tính:

$$P(SR|H) = \frac{P(SR)\cdot P(H|SR)}{P(H)}$$

$$= \frac{0,2(0,80)}{0,193}$$

$$= 0,829$$

e) Đặt: A = SR|LT ⇒ P(A) = 0,470

B = SR|K ⇒ P(B) = 0,143

C = SR|H ⇒ P(C) = 0,829

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + (P(ABC)$$

$$= 0,47 + 0,43 + 0,829 - (0,47)(0,143) - (0,47)(0,829) - (0,143)(0,829) + (0,47)(0,143)(0,729)$$

$$= 0,992$$

[7] Trong một dân số tỷ lệ bệnh B là 0,23; tỳ lệ có triệu chứng C là 0,20. Một người đến từ dân số không có triệu chứng C, tính khả năng người này bị bệnh B khi biết thêm:

a) Tỷ lệ có triệu chứng C và bệnh B là 0,18.

b) Trong số những người bị bệnh B, tỷ lệ C chiếm 78%.

c) Những người không bị bệnh B và không có C chiếm tỷ lệ 75% trong dân số.

BÀI GIẢI

Ta có: $P(B) = 0,23$; $P(\overline{B}) = 0,77$

$P(C) = 0,20$; $P(\overline{C}) = 0,80$

Yêu cầu bài toán: $P(B|\overline{C}) = ?$

a) $P(CB) = 0,18$

$$P(B|\overline{C}) = \frac{P(B\overline{C})}{P\overline{C}} = \frac{P(B)\cdot P(\overline{C}|B)}{P(\overline{C})}$$

$$= \frac{P(B)}{P(\overline{C})}\left(1 - P(C|B)\right) = \frac{P(B)}{P(\overline{C})}\left(1 - \frac{P(BC)}{P(B)}\right)$$

$$= \frac{P(B) - P(C)}{P(\overline{C})} = \frac{0,23 - 0,18}{0,80}$$

$$= 0,0625$$

$$P(B|\overline{C}) = 0,0625$$

b) $P(C|B) = 0,78$

$\Rightarrow P(\overline{C}|B) = 0,22$

$$P(B|\overline{C}) = \frac{P(B\overline{C})}{P(\overline{C})} = \frac{P(B)\cdot P(\overline{C}|B)}{P(\overline{C})}$$

$$= \frac{0,23(0,22)}{0,8}$$

$\Rightarrow P(B|\overline{C}) = 0,063$

c) $P(\overline{B}\,\overline{C}) = 0,75 = P(\overline{C})\cdot P(\overline{B}|\overline{C})$

$$P(\overline{B}\,\overline{C}) \quad = P(\overline{C})\cdot P(\overline{B}|\overline{C}) \quad = P(\overline{C})(1 - P(B|\overline{C}))$$

$$\Rightarrow P(B\,\overline{C}) = 1 - \frac{P(\overline{B}\,\overline{C})}{P(\overline{C})} \quad = 1 - \frac{0,75}{0,80}$$

$$P(B|\overline{C}) = 0,0625$$

[8]a) Tính tỷ lệ sinh trai là ½ . Sanh bao nhiêu lần thì khả năng sinh trai ít nhất là 90%.

b) Bạn có ý định "sinh được cháu trai thì dừng!". Trong một dân số, nếu các cặp vợ chồng đều có ý định như vậy thì số cặp không chịu "dừng lại ở 1 hoặc 2 con" chiếm tỷ lệ là bao nhiêu?

c) Bạn cần 10 cặp vợ chồng chịu hợp tác để nghiên cứu về 1 chế độ "Ăn uống nhằm sinh con trai" (?). Trong điều kiện hạn chế sinh đẻ thì tỷ lệ hợp tác là khá cao: 60%. Tính XS bạn phải mời 20 cặp mới được 10 cặp chịu hợp tác?

BÀI GIẢI

a) $T_1, T_2, \ldots, T_k$ biến cố sinh lần thứ k có con trai, $k \in N$

$\overline{A} = \overline{T}_1, \overline{T}_2 \ldots \overline{T}_k$ biến cố sinh k lần không có con trai.

$$P(\overline{A}) = P(\overline{T}_1, \overline{T}_2 \ldots \overline{T}_k) = \left(\frac{1}{2}\right)^k$$

A: biến cố có con trai trong k lần sinh.

$$P(A) = 1 - P(\overline{A}) = 1 - \left(\frac{1}{2}\right)^k$$

Để cho:

$$P(A) \geq 0{,}90 \quad \Rightarrow 1 - \left(\frac{1}{2}\right)^k \geq 0{,}90$$

$$\Rightarrow \left(\frac{1}{2}\right)^k \leq 0{,}10$$

$$k \cdot \lg(0{,}5) \leq \lg(0{,}10)$$

$$-0{,}3k \leq -1$$

$k \geq 4$

b) X: Số lần sinh cho tới lúc được con trai

$$X = 1 \Rightarrow P(X = 1) = \frac{1}{2}$$

$$X = 2 \Rightarrow P(\overline{X}X) = P(\overline{X}) \cdot P(X) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

X	1	2	3	4	...	
P(X)	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	...	$\sum P)X) = 1$

$$\left(\sum_{1}^{n} P(X) = \frac{1}{2} + \frac{1}{4} + \cdots = \frac{1}{2} \cdot \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \frac{1}{2}} = 1\right)$$

$$P(X \leq 2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

$$P(X > 2) = 1 - P(X \leq 2) = 1 - \frac{3}{4} = \frac{1}{4}$$

Tỉ lệ những cặp vợ chồng không chịu dừng lại 1, 2 con chiếm tỷ lệ 25%.

c) X số cặp chịu hợp tác khi mới 20 cặp:

$$X \sim B(20;\,0{,}6) \quad \Rightarrow \quad \mu_X = np = 12$$

$$\sigma_X^2 = npq = 4{,}8$$

$$\Rightarrow \quad \sigma_X = 2{,}19$$

$$P(X = 10) \quad \# \quad P\left(\frac{9{,}5-12}{2{,}19} \le \frac{X-12}{2{,}19} \le \frac{10{,}5-12}{2{,}19}\right)$$

$$= \quad (-1{,}14 \le u \le -0{,}68)$$

$$= \quad \Phi(1{,}14) - \Phi(0{,}68)$$

$$= \quad 0{,}873 - 0{,}752$$

$$= \quad 0{,}12$$

Khả năng có đúng 10 cặp chịu hợp tác khi mời 20 cặp là 12%.

[9] Số bệnh nhân đến phòng khám với mức độ trung bình 4 người/ giờ.

a) Tính khả năng có đúng 10 người đến khám trong khoảng từ 8 đến 10 giờ sáng?

b) Nếu BS bận điểm tâm trong 30 phút, thì khả năng không có bệnh nhân nào đến khám trong 30 phút đó là bao nhiêu?

c) Dự đoán xem có bao nhiêu bệnh nhân đến khám lúc BS bận điểm tâm?

BÀI GIẢI

X Số người đến khám trong khoảng 8 đến 10 giờ.

a) Thời gian: 10 – 8 = 2 giờ.

Trung bình số người đến khám trong 2 giờ:

$\lambda_1 = 4 \times 2 \quad = 8$

$$X_1 \sim \mathcal{P}(\lambda_1) \quad = \frac{e^{-8}8^x}{x!}; x = 1, 2, \ldots$$

$$P(X_1 = 10) \quad = \frac{e^{-8}8^{10}}{10!} = \frac{1}{e^8}\frac{8^{10}}{10!} = 0,099$$

b) Thời gian 30 phút = $\frac{1}{2}$ giờ $\Rightarrow \lambda_1 = 4 \cdot \frac{1}{2} = 2$

$$X_2 \sim \mathcal{P}(\lambda_2) \quad = \frac{c^{-2}2^x}{x!}$$

$$P(X_2 = 0) \quad = \frac{1}{e^2} = \frac{1}{7,389} = 0,135$$

c) $E(X_2) = \lambda_2 = 2$

Có hai người đến khám trong thời gian BS điểm tâm.

[10] Số n/b đến khám thường bị 1 trong 3 bệnh M_1, M_2, M_3, với xác suất 0,4; 0,4; 0,2. Để Δ, BS dùng 3 XN E_1, E_2, E_3, mà kết quả sẽ là + hay – (ký hiệu là 1 hay 0). Ta gọi Profil symptomatique S của 1 b/n, chẳng hạn $E_1 = 0$, $E_2 = 1$, $E_3 = 0$ là S (010).

1) Số profil có thể có là bao nhiêu?

2) Cho biết $P(S|M_1)$ như sau:

$$P(101|M_1) = 0{,}5; P(001|M_1) = 0{,}4 \; P(010|M_1) = 0{,}1$$

$$P(100|M_2) = 0{,}8; P(000|M_2) = 0{,}2; P(111|M_3) = 1$$

Hỏi hai xác suất $P(100/M_1)$, $P(110/M_2)$ có giá trị là bao nhiêu?

3) Theo ý Anh, Chị liên tiếp thực hiện 3 XN này có cho phép Δ chắc chắn (không nhầm) hay không? tại sao?

4) Một b/n đến khám có E_1 âm, khả năng người này bị M_3, M_2, M_1 là bao nhiêu?

Khả năng 3 bệnh đó là bao nhiêu nếu b/n này có XN thứ 2 là dương?

5) Theo ý Anh, Chị liệu có một Δ chắc chắn bằng cách chỉ cần 2 XN? Nếu có, 2 cái nào? Nếu không, tại sao?

BÀI GIẢI

1) Dạng tổng quát của một profil: $S(E_1, E_2, E_3)$ trong đó E_1 lấy hai giá trị 0,1. Do đó bộ ba E_1, E_2, E_3 lấy $2 \times 2 \times 2 = 8$ profil có thể có:

000, 100, 010, 001, 110, 101, 011, 111

Mỗi profil là một biến cố về kết quả xét nghiệm.

2) Ta có:

- $P(101|M_1) = 0{,}5$; $P(001|M_1) = 0{,}4$; $P(010|M_1) = 0{,}1$

Hay:

M_1	101	001	010	
P	0,5	0,4	0,1	1

- $P(100|M_2) = 0{,}8$; $P(000|M_2) = 0{,}2$

M_2	100	000	
P	0,8	0,2	1

- $P(111|M_3) = 1$

M_3	111	
P	1	1

Tổng X/S trong mỗi không gian mẫu đều bằng 1.

Do vậy: $P(100|M_1) = 0$ và $P(110|M_2) = 0$

3) Bảng chẩn đoán Mi dựa vào $S(E_1, E_2, E_3)$:

Δ \ S	000	100	010	001	110	101	011	111
M_1	0	0	0,1	0,4	0	0,5	0	0
M_2	0,2	0,8	0	0	0	0	0	0
M_3	0	0	0	0	0	0	0	1
Kết luận	M_2	M_2	M_1	M_1		M_1		M_3

Với kết quả 3 xét nghiệm E_1, E_2, E_3 ta có kết luận người này có bệnh M_1 hay không. Không có sự nhầm lẫn hay chồng chéo trong chẩn đoán: Mỗi profil S cho ta một chẩn đoán cụ thể.

4) Yêu cầu câu toán: $P(M_i|E_1 = 0) = ?$

a)

- $$P(M_1|E_1 = 0) = \frac{P(M_1) \cdot P(E_1 = 0|M_1)}{P(E_1 = 0)}$$

$$= \frac{P(M_1) \cdot P(0**|M_1)}{P(E_1 = 0)}$$

$P(M_1) = 0,40$

$$P(0**|M_1) = P(010|M_1) + P(001|M_1)$$
$$= 0,1 + 0,4 = 0,5$$

$$P(E_1 = 0) = P(M_1) \cdot P(E_1 = 0|M_1) + P(M_2) \cdot P(E_1 = 0|M_2) + P(M_3) \cdot P(E_1 = 0|M_3)$$
$$= 0,4(0,5) + 0,4(0,2) + 0,2(0) = 0,28$$

$$P(M_1|E_1 = 0) = \frac{0,4(0,5)}{0,28} = 0,714$$

- $P(M_2|E_1 = 0) = \dfrac{P(M_2) \cdot P(E_1 = 0|M_2)}{P(E_1 = 0)} = \dfrac{0,4(0,2)}{0,28}$

  $= 0,268$

- $P(M_3|E_1 = 0) = \dfrac{P(M_3) \cdot P(E_1 = 0|M_3)}{P(E_1 = 0)} = \dfrac{0,4(0)}{0,28}$

  $= 0$

b) Yêu cầu tính: $P(M_i|E_1 = 0, E_2 = 1) = P(M_i|01)$:

- $P(M_1|01 = 0) = \dfrac{P(M_1) \cdot P(01*|M_1)}{P(01)}$

$$P(01*|M_1) = P(010|M_1) + P(011|M_1) = 0,1 + 0 = 0,1$$

$$P(01*|M_2) = P(010|M_2) + P(011|M_2) = 0 + 0 = 0$$

$$P(01*|M_3) = P(010|M_3) + P(010|M_3) = 0 + 0 = 0$$

$$P(01) = P(M_1) \cdot P(01|M_1) + P(M_2) \cdot P(01|M_2) + P(M_3) \cdot P(01|M_3)$$

$$= 0,4(0,1) + 0,4(0) = 0,2(0)$$
$$= 0,04$$

$$P(M_1|01) = \frac{P(M_1) \cdot P(01|M_1)}{P(01)} = \frac{0,4(01)}{0,04} = 1$$

$$P(M_2|01) = \frac{P(M_2) \cdot P(01|M_2)}{P(01)} = \frac{0,4(0)}{0,04} = 0$$

$$P(M_3|01) = 0$$

4) Theo giả thiết bài toán, ta có:

Xét nghiệm			Chẩn đoán
E_1	E_2	E_3	Δ
0	0	0	M_2
0	0	1	M_1
0	1	0	M_1
1	0	0	M_2
0	1	1	Không bị M
1	0	1	M_1
1	1	0	Không bị M
1	1	1	M_3

Ta có:

– Chẩn đoán là M_1 nếu $E_2 = 0$ $E_3 = 1$ $E_1 = 0$

$E_2 = 1$ $E_3 = 0$ $E_1 = 1$

$E_2 = 0$ $E_3 = 1$ $E_1 = 1$

hay $E_2 \neq E_3$. E_1 tùy ý

– Chẩn đoán là M_2 nếu $E_2 = 0$ $E_3 = 1$ $E_1 = 0$

$E_2 = 0$ $E_3 = 0$ $E_1 = 1$

hay $E_2 = 0$ $E_3 = 0$ E_1 = tùy ý

– Chẩn đoán là M_3 nếu $E_2 = 1$ $E_3 = 1$ $E_1 = 1$

hay $E_2 = 1$ $E_3 = 1$ E_1 = tùy ý

(Vì $P(E_2 = 1\ E_3 = 1\ E_1 = 0) = 0$)

Vậy chỉ cần 2 xét nghiệm E_2, E_3 là đủ để Δ M thuộc type nào:

$$P(M_1 | E_2 \neq E_3) = 1$$
$$P(M_2 | E_2 = E_3 = 0) = 1$$
$$P(M_3 | E_2 = E_3 = 1) = 1$$

[11] Bệnh A có 3 thể A_1, A_2, A_3 với tỷ lệ là $1/2$; $1/6$; $1/3$ trong dân số. Để chẩn đoán bệnh A ta dùng XN E_1. E_1 cho dương tính 10% nếu là A_1, 20% nếu A_2, 90% nếu A_3

1) Khả năng E_1 cho âm tính đối với một người bị bệnh A là bao nhiêu?

2) Một người bị bệnh A với kết quả E_1 dương tính thì khả năng bị A_1, A_2, A_3 là bao nhiêu?

3) Nếu E_1 âm tính, làm tiếp E_2, E_2 độc lập với E_1 bất chấp A ở thể nào. E_2 dương tính là 5% nếu A_1, 80% nếu A_2, 3% nếu A_3. Hỏi E_1, E_2 có độc lập với bệnh A hay không?

4) Người ta chọn luật quyết định như sau: Nếu E_1^+ ta chọn A_3; E_1^- và E_2^+ chọn A_2; E_1^- và E_2^- ta chọn A_1. Bảng sao đây cho biết thiệt hại gây ra do sai sót trong chẩn đoán. Hỏi thiệt hại trung bình gây ra do quyết định đó là bao nhiêu?

		FORME VRAIE		
		A_1	A_2	A_3
	A_1	0	1	2
DIAGNOSTIC	A_2	1	0	1
	A_3	2	1	0

BÀI GIẢI

1) Tính $P(E_1^-)$:

$$P(E_1^-) = 1 - P(E_1^+)$$

$$P(E_1^+) = P(A_1)\cdot P(E_1^+|A_1) + P(A_2)\cdot P(E_2^+|A_2) + P(A_3)\cdot P(E_1^+|A_3)$$

$$= \frac{1}{2}(0,1) + \frac{1}{6}(0,2) + \frac{1}{3}(-,9)$$

$$= 0,383$$

$$P(E_1^-) = 1 - 0,383 = 0,617$$

2) $$P(A_1|E_1^+) = \frac{P(A_1)\cdot P(E_1^+|A_1)}{P(E_1^+)}$$

$$= \frac{\frac{1}{2}(0,1)}{0,038} = 0,13$$

$$P(A_2|E_1^+) = \frac{\frac{1}{6}(0,2)}{0,038} = 0,087$$

$$P(A_3|E_1^+) = \frac{\frac{1}{3}(0,9)}{0,038} = 0,783$$

3) E_2 độc lập với E_1 bất chấp A ở thể nào

$$\Leftrightarrow P(E_1^-E_2^+|A_1) = P(E_1^-|A_1)\cdot P(E_2^+|A_1)$$

Ta có: $P(E_1^- E_2^+ | A_1)$ $= P(E_1^- | A_2) \cdot P(E_2^+ | A_1)$

$= (1 - = P(E_1^+ | A_1) P(E_2^+ | A_1)$

$= (1 - 0,1)(0,05) = 0,045$

Tương tự: $P(E_1^- E_2^+ | A_2)$ $= P(E_1^- | A_2) P(E_2^+ | A_2)$

$= 0,8(0,8) = 0,64$

$P(E_1^- E_2^+ | A_3)$ $= 0,1(0,03)$
$= 0,003$

Để chứng minh E_1, E_2 có độc lập hay không, ta chứng minh:

E_1^-, E_2^+ độc lập $\Leftrightarrow P(E_1^- E_2^+) = P(E_1^-) P(E_2^+)$

Vì rằng A, B độc lập $\Leftrightarrow \overline{A}, \overline{B}$ độc lập

$\Leftrightarrow A, \overline{B}$ độc lập

$\Leftrightarrow \overline{A}, B$ độc lập

$P(E_1^- E_2^+)$ $= \sum P(A_i) \cdot P(E_1^-, E_2^+ | A_i)$

$= \frac{1}{2}(0,045) + \frac{1}{6}(0,64) + \frac{1}{3}(0,003)$

$= 0,13$

Còn: $P(E_1^-) \cdot P(E_2^+)$ $= 0,617\left(\sum P(A_i) \cdot P(E_2^+ | A_1)\right)$

$= 0,617(0,618) = 0,104$

Vậy: $P(E_1^- E_2^+) \neq P(E_1^-) \cdot P(E_2^+)$ nên E_1, E_2 không độc lập.

4) Dựa vào luật quyết định của đề bài, ta có:

- Nếu $E_1^+ \Rightarrow$ chọn $A_3 \Rightarrow$ sai sót có thể là A_1, A_2

Do đó thiệt hại là:

$$2P(A_1 \cdot E_1^+) + 1P(A_2 E_1^+)$$
$$= 2P(A_1) \cdot P(E_1^+ | A_1) + 1P(A_2)P(E_1^+ | A_2)$$
$$= 2\left(\frac{1}{2}\right)(0,1) + 1\left(\frac{1}{6}\right)(0,2) = 0,113 \text{ đơn vị thiệt hại.}$$

- Nếu $E_1^- \, E_2^+$ chọn $A_2 \Rightarrow$ sai sót A_1, A_3

Thiệt hại là:

$$1P(A_1 \cdot E_1^- \, E_2^+) + 1P(A_3 \cdot E_1^- \, E_2^+)$$
$$= 1.P(A_1) \cdot P(E_1 \, E_2^+ | A_1) + 1P(A_3) \cdot P(E_1^- \, E_2^+ | A_3)$$
$$= 1 \cdot \left(\frac{1}{2}\right) \cdot 0,045 + 1\left(\frac{1}{3}\right) \cdot 0,003 = 0,0235 \text{ đv}$$

- Nếu $E_1^- \, E_2^-$ chọn $A_1 \Rightarrow$ sai sót A_2, A_3.

Thiệt hại là:

$$1 \cdot P(E_1^- \, E_2^- \cdot A_2) + 2 \cdot P(E_1^- \, E_2^- A_3)$$
$$= 1.P(A_2) \cdot P(E_1^- E_2^- | A_2) + 2 \cdot P(A_3)((E_1^- \, E_2^- | A_3)$$
$$= 1 \cdot \left(\frac{1}{6}\right) \cdot P(E_1^- | A_2) \cdot P(E_2^- | A_2) + 2\left(\frac{1}{3}\right) \cdot P(E_1^- | A_3) \cdot P(E_2^- | A_3)$$
$$= 1 \cdot \left(\frac{1}{6}\right) \cdot (0,80) \cdot (0,20) + 2\left(\frac{1}{3}\right) \cdot (0,10) \cdot (0,97)$$
$$= 0,0913$$

Tổng thiệt hại: 0,133 + 0,0235 + 0,0913 = 0,248 đv.

[12] Trong dân số bệnh M., BS tự hỏi liệu 2 triệu chứng S_1 và S_2 liên quan hay không? BS quan sát ngẫu nhiên 300 người bệnh M, kết quả như sau:

S1:		0	1
S2:	0	45	105
	1	72	78

1) Làm cách nào BS test sự liên quan 2 triệu chứng đó của bệnh này?

2) Với test đã chọn, những giá trị nào của ngưỡng ý nghĩa α có thể giúp kết luận?

3) Tìm KTC 95% về tỷ lệ b/n M có cả 2 triệu chứng S_1 và S_2 trong dân số.

4) Một đồng nghiệp cho rằng sự liên hệ S_1, S_2 xảy ra do sự kiện là bệnh M có 2 type khác nhau là M_1 và M_2, nên ông cho rằng nếu lưu ý điều này thì sự liên hệ S_1, S_2 sẽ không còn.

Để c/m, ông ta dựa vào mẫu vừa rồi, phần M ra 2 dạng M_1, M_2 và theo S_1, S_2.

		M_1				M_2	
S_1		0	1	S_1		0	1
S_2	0	27	93	S_2	0	18	12
	1	11	35		1	61	43

hãy giúp ông ta thực hiện phép kiểm định và kết luận.

5) Dựa vào quan sát trên, người ta kết hợp cả 3 biến S_1, S_2 và M, chẳng hạn khả năng 1 b/n không có S_1, S_2 và bị M_2 là 18/300.

Nếu dùng S_1, S_2 để chẩn đóan bệnh M_1, M_2 khi biết là bị M. Anh, Chị cho biết khả năng bị M_1 hoặc M_2 khi:

Không có S_1 lẫn S_2. Có S_1 không S_2.

Không S_1 có S_2. Có cả S_1 và S_2.

Xác suất sai sót trong Δ theo cách này là bao nhiêu?

6) Mặt khác người ta để ý tới hàm lượng cholesterol (X) trong máu, thấy có sự khác nhau giữa M_1 và M_2.

Cho biết: X_1 (M_1) ~ N(1,5; 0,25); X_2 (M_2) ~ N(2,5; 0,25). Hãy nêu 1 cách Δ thích hợp khi dựa vào hàm lượng của X.

Chẩn đoán theo cách này có tốt hơn Δ theo S_1, S_2 trên đây không?

7) Anh, Chị cho rằng còn một cách Δ tốt hơn nữa bằng cách dựa vào cả 3: S_1; S_2 và X?

Để có cách Δ dựa vào cả 3 biến này thì ta cần biết những yếu tố khác mà không có trong bài toán này. Yếu tố nào?

BÀI GIẢI

1) Dùng test $\chi^2(1)$ để đánh giá sự độc lập của hai triệu chứng: S_1, S_2.

2) Với test $\chi^2(1)$ nêu trên, ta có $Q_1 = 10,21$.

Với bảng χ^2 ta có: $10,8 > 10,21 > 6,8$

$$\Rightarrow 0,001 \leq P - \text{value} < 0,01$$

Với ngưỡng $\alpha = 0,05$ hoặc $\alpha = 0,01$ thì P – value < α nên bác bỏ $H_0 \Rightarrow S_1, S_2$ có phụ thuộc.

3) Với $f = \dfrac{78}{300} \quad \Rightarrow \quad P = 0,26 \pm 1,96\sqrt{0,26 \cdot 0,74/300}$

$$P = 0,26 \pm 0,05$$

4) Dùng test $\chi^2(1)$ cho từng dạng M_1, M_2 ứng với tần số quan sát đầu bài đã cho.

Tính ra ta có:

$$
\begin{aligned}
Q_1 = 0,038 \quad &\Rightarrow \quad 0,016 < 0,038 < 0,45\\
&\Rightarrow \quad 0,50 < \text{p value} < 0,90\\
&\Rightarrow \quad \text{chấp nhận } H_0\\
Q_2 = 0,017 \quad &\Rightarrow \quad 0,016 < 0,017 < 0,45\\
&\Rightarrow \quad 0,50 < \text{P–value} < 0,90\\
&\Rightarrow \quad \text{chấp nhận } H_0
\end{aligned}
$$

S_1, S_2 độc lập trên từng dạng bệnh M_1, M_2.

5) Dùng 3 biến S_1, S_2, M; mỗi biến có giá trị là $S_1(0, 1)$, $S_2(0, 1)$, $M(M_1, M_2)$.

$$
\begin{aligned}
P(S_1^-, S_2^-, M_2) &= P(0,0,M_2)\\
&= \frac{18}{300}
\end{aligned}
$$

Khi biết là bị M, ta dùng bảng số liệu đầu tiên để tính các xác suất liên quan S_1, S_2.

Với (S_1, S_2): (0,0), (0,1), (0,1), (11) ứng với một bộ giá trị (S_1, S_2) ta tính $P(M_1|S_1 S_2)$ và $P(M_2|S_1 S_2)$ chọn M_i nào ứng với xác suất cao nhất.

- Không có S_1 lẫn S_2 $\Leftrightarrow$ $S_1^- S_2^- = (0,0)$

$$P(M_1|0,0) = \frac{P(M_1)\cdot P(0,0|M_1)}{P(0,0)}$$

$$= \frac{P(M_1;0,0)}{P(0,0)}$$

$$= \frac{\frac{27}{300}}{\frac{45}{300}} = \frac{27}{45}$$

$$= 0{,}60$$

$$P(M_2|0,0) = \frac{P(M_2;0,0)}{P(0,0)} = \frac{18}{45}$$

$$= 0{,}40$$

$\Rightarrow$ Chọn M_1.

Tương tự:

- $S_1^+ S_2^-$: $P(M_1|1,0) = \dfrac{P(M_1;1,0)}{P(1,0)} = \dfrac{93}{105}$

$$= 0{,}88$$

$$P(M_2|1,0) = \frac{P(M_2;10)}{P(1,0)} = \frac{12}{105}$$

$$= 0{,}11$$

$\Rightarrow$ Chọn M_1.

- $S_1^- S_2^+ : P(M_1|0,1) = \dfrac{P(M_1;01)}{P(01)} = \dfrac{11}{72}$

$$= 0{,}15$$

$$P(M_2|01) = \dfrac{P(M_2;01)}{P(01)} = \dfrac{61}{72}$$

$$= 0{,}85$$

$\Rightarrow$ Chọn M_2.

- $S_1^+ S_2^+ : P(M_1|11) = \dfrac{P(M_1;11)}{P(11)} = \dfrac{35}{78}$

$$= 0{,}45$$

$$P(M_2|11) = \dfrac{P(M_2;11)}{P(11)} = \dfrac{43}{78}$$

$$= 0{,}55$$

$\Rightarrow$ Chọn M_2.

Bảng chọn M_i theo giá trụ S_1 S_2:

S_1 S_2	Chọn
00	M_1
10	M_1
01	M_2
11	M_2

Vậy chọn M_1 khi $S_2 = 0$ và M_2 khi $S_2 = 1$ chấp nhận S_1: Chỉ cần một xét nghiệm S_2 đủ để chẩn đoán M_i.

Xác suất sai theo cách chẩn đoán này:

S2	Δ	Sự thật	X/S sai
0	M_1	M_2	α_2
1	M_2	M_1	α_1

$$\begin{aligned}\alpha_2 &= P(M_1|M_2)?\\ &= P(S_2 = 0|M_2)\\ &= \frac{P(S_2 = 0, M_2)}{P(M_2)}?\\ &= \frac{30}{134}\\ &= 0{,}224\end{aligned}$$

$$\begin{aligned}\alpha_1 &= P(M_2|M_1)?\\ &= P(S_2 = 1|M_1) = \frac{P(S_2 = 1, M_1)}{P(M_1)}\\ &= \frac{46}{166}\\ &= 0{,}28\end{aligned}$$

Xác suất sai α_i là quá lớn, xấp xỉ 25%.

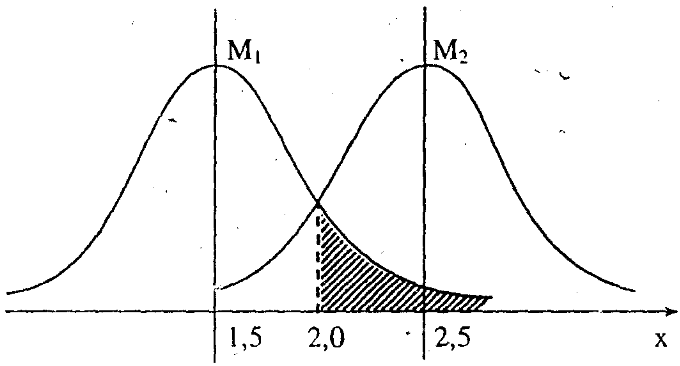
6) Để Δ là M_1 hay M_2 có thể dựa vào X

Do tính đối xứng của hàm mật độ X trong hai bệnh M_1, M_2 nên xác suất sai sót là như nhau và bằng:

$$\begin{aligned}\alpha_1 = \alpha_2 &= P(X \geq 2)\\ &= P\left(\frac{x-1,5}{0,5} \geq \frac{2-1,5}{0,5}\right)\end{aligned}$$

$$= P(u \geq 1) = 0{,}159 \neq 16\%$$

Sai sót trong cách chẩn đoán dựa vào hàm lượng X nhỏ hơn cách chẩn đoán trên.



$$X \geq 2 \Rightarrow M_2$$

$$X < 2 \Rightarrow M_1$$

7) Chẩn đoán dựa vào (S_1, S_2, X) trên lý thuyết là tốt hơn (ít sai sót hơn).

Nhưng cần phải biết hàm phân phối xác suất ba biến (S_1, S_2, X) hoặc hàm phân phối (S_1, S_2) và phân phối của X độc lập với phân phối (S_1, S_2).