Mục lục

Chương I

Biến cố và xác suất của các biến cố

1 Bài tập chương 1

Bài tập 1.1. Cho 3 xạ thủ, mỗi người bắn một phát vào mục tiêu. Gọi A_i là biến cố người thứ i $(i = \overline{1,3})$ bắn trúng. Hãy biểu diễn các biến cố sau qua các biến cố A_i
a) Có một người bắn trúng.
b) Có ít nhất một người bắn trúng.
c) Người thứ nhất bắn trúng.
d) Người thứ hai và người thứ ba cùng bắn trúng.
e) Người thứ nhất bắn trúng hoặc người thứ hai và người thứ ba cùng bắn trúng.
f) Chỉ 2 người bắn trúng.

Lời giải

a) $A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3$
b) $A_1 \cup A_2 \cup A_3$
c) A_1
d) $A_2 \cdot A_3$
e) $A_1 \cup A_2 \cdot A_3$
f) $A_1 \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot A_3 \cup A_1 \cdot \overline{A_2} \cdot A_3$

Bài tập 1.2. Trong một kho hàng có 10.000 sản phẩm có 500 phế phẩm. Lấy ngẫu nhiên 1 sản phẩm từ kho. Tìm xác suất để sản phẩm là phế phẩm. (Với phép thử lấy ngẫu nhiên một phần tử thuộc tổng thể thì xác suất để một phần tử ngẫu nhiên của tổng thể có tính chất nào đó sẽ bằng tỉ lệ có tính chất đó của tổng thể).

Lời giải

Phép thử: Lấy ngẫu nhiên một sản phẩm.
Gọi X là biến cố "Sản phẩm lấy ra là phế phẩm" thì $P(X) = \dfrac{500}{10000} = 0,05$.

Bài tập 1.3. Một túi đựng 10 quả cầu, trong đó có 6 quả màu xanh và 4 quả màu vàng. Lấy ngẫu nhiên từ túi ra 3 quả cầu. Tìm xác suất để trong 3 quả cầu lấy ra có 2 quả cầu xanh.

Lời giải

Phép thử: Lấu ngẫu nhiên 3 quả cầu từ túi có 10 quả cầu.
Gọi X là biến cố "Trong 3 quả cầu lấy ra có 2 quả cầu xanh" thì trong 3 quả lấy ra gồm 2 quả xanh, 1 quả vàng $\Rightarrow$ có $C_6^2 \cdot C_4^1$ cách. Suy ra

$$P(X) = \frac{C_6^2 \cdot C_4^1}{C_{10}^3} = \frac{1}{2}$$

Bài tập 1.4. Một người cần gọi điện thoại nhưng quên mất hai chữ số cuối của số điện thoại cần gọi và chỉ nhớ là hai chữ số đó khác nhau. Ông bấm số điện thoại với 2 chữ số cuối là ngẫu nhiên theo cách nhớ. Tính xác suất để ông gọi trúng ngay số điện thoại cần gọi.

Lời giải

Gọi 2 số cuối điện thoại cần tìm có dạng $\overline{ab}$ thì phép thử là chọn cặp số (a, b) với $a,\ b \in \{0; 1; \cdots; 9\}$ và $a \neq b$. Khi đó ta có 10 cách chọn a, 9 cách chọn b nên có tất cả 90 trường hợp có thể xảy ra.

Gọi X là biến cố "Ông gọi trúng số cần gọi" thì chỉ có 1 trường hợp duy nhất nên $P(X) = \dfrac{1}{90}$.

Bài tập 1.5. Một khách sạn có 6 phòng đơn. Có 10 khách đến thuê phòng, trong đó có 6 nam và 4 nữ. Người quản lý chọn ngẫu nhiên 6 người. Tìm xác suất để

a) Cả 6 người đều là nam.

b) Có 4 nam và 2 nữ.

c) Có ít nhất 2 nữ.

Lời giải

Phép thử: Chọn ngẫu nhiên 6 người từ 10 người.

a) Gọi A là biến cố "Cả 6 người đều là nam" thì

$$P(A) = \frac{C_6^6}{C_{10}^6} = \frac{1}{210}$$

b) Gọi B là biến cố "Có 4 nam và 2 nữ" thì

$$P(B) = \frac{C_6^4 \cdot C_4^2}{C_{10}^6} = \frac{3}{7}$$

c) Gọi C là biến cố "Có ít nhất 2 nữ" thì ta có các trường hợp:

• Có 4 nam, 2 nữ thì có $C_6^4 \cdot C_4^2$ cách.

• Có 3 nam, 3 nữ thì có $C_6^3 \cdot C_4^3$ cách.

• Có 2 nam, 4 nữ thì có $C_6^2 \cdot C_4^4$ cách.

Suy ra

$$P(C) = \frac{C_6^4 \cdot C_4^2 + C_6^3 \cdot C_4^3 + C_6^2 \cdot C_4^4}{C_{10}^6} = \frac{37}{42}$$

Bài tập 1.6. Một hộp có 10 sản phẩm, trong đó có 3 phế phẩm và 7 chính phẩm. Lấy ngẫu nhiên từ hộp ra 5 sản phẩm. Tính xác suất để

a) Không có phế phẩm nào.

b) Có không quá 1 phế phẩm.

c) Có ít nhất 1 phế phẩm.

Lời giải

Phép thử: Lấy ngẫu nhiên 5 sản phẩm từ hộp có 10 sản phẩm.

a) Gọi A là biến cố "Không có phế phẩm nào" thì

$$P(A) = \frac{C_7^5}{C_{10}^5} = \frac{1}{12}$$

c) Gọi B là biến cố "Có không quá 1 phế phẩm" thì ta có các trường hợp:

• Không có phế phẩm nào thì có C_7^5 cách.

• Có 1 phế phẩm, 4 chính phẩm thì có $C_7^4 \cdot C_3^1$ cách.

Suy ra

$$P(B) = \frac{C_7^4 \cdot C_3^1 + C_7^5}{C_{10}^5} = 0,5$$

c) Gọi C là biến cố "Có ít nhất 1 phế phẩm" thì ta có $C = \overline{A}$. Suy ra

$$P\left(C\right) = P\left(\overline{A}\right) = 1 - P\left(A\right) = \frac{11}{12}$$

Bài tập 1.7. Một công ty có 60 nhân viên, trong đó có 20 nam và 40 nữ. Tỉ lệ nhân viên nữ có thể nói tiếng Anh lưu loát là 15% và tỉ lệ này đối với nam là 20%.
a) Gặp ngẫu nhiên một nhân viên của công ty. Tìm xác suất để gặp được nhân viên nói tiếng Anh lưu loát.
b) Gặp ngẫu nhiên hai nhân viên của công ty. Tìm xác suất để có ít nhất một người nói tiếng Anh lưu loát trong số 2 người này.

Lời giải

Số người nói tiếng Anh lưu loát trong công ty là $20 \cdot 20\% + 40 \cdot 15\% = 10$ nhân viên. Suy ra 50 nhân viên còn lại không thể nói tiếng Anh lưu loát.
a) Phép thử: Gặp ngẫu nhiên 1 nhân viên.
Gọi A là biến cố "Nhân viên gặp được nói tiếng Anh lưu loát" thì ta có

$$P\left(A\right) = \frac{C_{10}^{1}}{C_{60}^{1}} = \frac{1}{6}$$

b) Phép thử: Gặp ngẫu nhiên 2 nhân viên.
Gọi B là biến cố "Ít nhất một trong hai người nói tiếng Anh lưu loát" thì $\overline{B}$ là biến cố "Cả hai đều không thể nói tiếng Anh lưu loát". Suy ra

$$P\left(\overline{B}\right) = 1 - P\left(B\right) = 1 - \frac{C_{50}^{2}}{C_{60}^{2}} = \frac{109}{354}$$

Bài tập 1.8. Theo khảo sát của tổ chức y tế WHO trong một vùng dân cư, tỉ lệ người mắc bệnh tim là 9%, bệnh huyết áp là 12% và mắc cả hai bệnh là 7%. Chọn ngẫu nhiên một người trong vùng. Tìm xác suất để người đó không mắc bệnh nào trong hai bệnh trên.

Lời giải

Phép thử: Chọn ngẫu nhiên một người trong vùng.
Gọi A là biến cố "Người được chọn mắc bệnh tim", B là biến cố "Người được chọn mắc bệnh huyết áp" thì

$$P\left(A\right) = 0,09; P\left(B\right) = 0,12; P\left(A \cdot B\right) = 0,07$$

Gọi X là biến cố "Người được chọn không mắc bệnh nào" thì $\overline{X}$ là biến cố "Người được chọn mắc ít nhất một trong hai bệnh", tức $\overline{X} = A \cup B$. Suy ra

$$\begin{aligned} P\left(X\right) &= 1 - P\left(\overline{X}\right) = 1 - P\left(A \cup B\right) = 1 - \left[P\left(A\right) + P\left(B\right) - P\left(A \cdot B\right)\right] = 1 - \left(0,09 + 0,12 - 0,07\right) \\ &= 0,86 \end{aligned}$$

Bài tập 1.9. Trong lớp gồm 100 học sinh, có 20 em sinh viên giỏi Toán, 25 em giỏi ngoại ngữ và 10 em giỏi cả Toán lẫn Ngoại ngữ. Quy định giỏi ít nhất một môn thì được thưởng. Chọn ngẫu nhiên một em trong lớp. Tính xác suất để em đó được thưởng. Suy ra tỉ lệ học sinh được thưởng của lớp.

Lời giải

Phép thử: Chọn ngẫu nhiên một em trong lớp.
Gọi A là biến cố "Học sinh được chọn giỏi Toán", B là biến cố "Học sinh được chọn giỏi Ngoại ngữ" thì

$$P\left(A\right) = \frac{20}{100} = 0,2; P\left(B\right) = \frac{25}{100} = 0,25; P\left(A \cdot B\right) = \frac{10}{100} = 0,1$$

Gọi X là biến cố "Học sinh được chọn có thưởng" thì ta có $X = A \cup B$. Suy ra

$$P\left(X\right) = P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cdot B\right) = 0,2 + 0,25 - 0,1 = 0,35$$

Vậy tỉ lệ học sinh được thưởng của lớp là 35%.

Bài tập 1.10. Một khách sạn có 3 thang máy 1, 2, 3 hoạt động độc lập với nhau. Xác suất để thang máy 1, 2, 3 bị hỏng lần lượt là $0,4$; $0,5$ và $0,6$. Tính xác suất để
a) Chỉ có duy nhất một thang máy bị hỏng.
b) Ít nhất một thang máy bị hỏng.
c) Biết chỉ có một thang máy bị hỏng, tìm xác suất để đó là thang máy 1.

Lời giải

Gọi A_i là biến cố "Thang máy thứ i bị hỏng" với $i = \overline{1,3}$ thì A_1, A_2, A_3 độc lập với nhau và $P(A_1) = 0,4$; $P(A_2) = 0,5$; $P(A_3) = 0,6$.
a) Gọi X là biến cố "Chỉ có duy nhất một thang máy bị hỏng" thì ta có

$$X = A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3$$

Do các biến cố $A_1 \cdot \overline{A_2} \cdot \overline{A_3}$, $\overline{A_1} \cdot A_2 \cdot \overline{A_3}$, $\overline{A_1} \cdot \overline{A_2} \cdot A_3$ đôi một xung khắc nên suy ra

$$\begin{aligned}
P(X) &= P\left(A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3\right) \\
&= P(A_1) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) + P\left(\overline{A_1}\right) \cdot P(A_2) \cdot P\left(\overline{A_3}\right) + P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P(A_3) \\
&= 0,4 \cdot (1 - 0,5) \cdot (1 - 0,6) + (1 - 0,4) \cdot 0,5 \cdot (1 - 0,6) + (1 - 0,4) \cdot (1 - 0,5) \cdot 0,6 \\
&= 0,38
\end{aligned}$$

b) Gọi Y là biến cố "Ít nhất một thang máy bị hỏng" thì ta có $\overline{Y}$ là biến cố "Không có thang máy nào bị hỏng", tức

$$\overline{Y} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$$

Suy ra

$$\begin{aligned}
P(Y) &= 1 - P(\overline{Y}) = 1 - P\left(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}\right) \\
&= 1 - P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) = 1 - (1 - 0,4) \cdot (1 - 0,5) \cdot (1 - 0,6) = 0,88
\end{aligned}$$

c) Ta có

$$P(A_1|X) = \frac{P(A_1 \cdot X)}{P(X)} = \frac{P\left(A_1 \cdot \overline{A_2} \cdot \overline{A_3}\right)}{P(X)} = \frac{0,4 \cdot (1 - 0,5) \cdot (1 - 0,6)}{0,38} = \frac{4}{19}$$

Bài tập 1.11. Một hãng vận tải cho 3 xe hoạt động độc lập trong năm. Xác suất để các xe bị hỏng tương ứng là $0,1$; $0,2$ và $0,15$. Tính xác suất để
a) Có một xe bị hỏng.
b) Có ít nhất một xe bị hỏng.
c) Biết chỉ có một xe bị hỏng, tìm xác suất để đó là xe 2.

Lời giải

Gọi A_i là biến cố "Xe thứ i bị hỏng" với $i = \overline{1,3}$ thì ta có A_1, A_2, A_3 độc lập với nhau và $P(A_1) = 0,1$; $P(A_2) = 0,2$; $P(A_3) = 0,15$.
a) Gọi X là biến cố "Có một xe bị hỏng" thì ta có

$$X = A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3$$

Do các biến cố $A_1 \cdot \overline{A_2} \cdot \overline{A_3}$, $\overline{A_1} \cdot A_2 \cdot \overline{A_3}$, $\overline{A_1} \cdot \overline{A_2} \cdot A_3$ đôi một xung khắc nên suy ra

$$\begin{aligned}
P(X) &= P\left(A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3\right) \\
&= P(A_1) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) + P\left(\overline{A_1}\right) \cdot P(A_2) \cdot P\left(\overline{A_3}\right) + P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P(A_3) \\
&= 0,1 \cdot (1 - 0,2) \cdot (1 - 0,15) + (1 - 0,1) \cdot 0,2 \cdot (1 - 0,15) + (1 - 0,1) \cdot (1 - 0,2) \cdot 0,15 \\
&= 0,329
\end{aligned}$$

b) Gọi Y là biến cố "Ít nhất một xe bị hỏng" thì ta có $\overline{Y}$ là biến cố "Không có xe nào bị hỏng", tức

$$\overline{Y} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$$

Suy ra

$$\begin{aligned} P(Y) &= 1 - P(\overline{Y}) = 1 - P\left(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}\right) \\ &= 1 - P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) = 1 - (1-0,1) \cdot (1-0,2) \cdot (1-0,15) = 0,388 \end{aligned}$$

c) Ta có

$$P(A_2|X) = \frac{P(A_2 \cdot X)}{P(X)} = \frac{P\left(\overline{A_1} \cdot A_2 \cdot \overline{A_3}\right)}{P(X)} = \frac{(1-0,1) \cdot 0,2 \cdot (1-0,15)}{0,329} = \frac{153}{329}$$

Bài tập 1.12. Chị Lan có một chùm chìa khóa gồm 9 chiếc chìa khóa bề ngoài rất giống nhau nhưng trong đó chỉ có 2 chiếc mở được cửa tủ. Chị Lan thử ngẫu nhiên từng chìa và chìa nào không đúng thì bỏ ra. Tìm xác suất để chị Lan mở được cửa ở lần thứ 4.

Lời giải

Phép thử: Lấy ngẫu nhiên không hoàn lại mỗi lần một chìa khóa.
Gọi A_i là biến cố "Chị Lan lấy đúng chìa khóa ở lần thứ i" với $i = \overline{1,3}$ và A là biến cố "Chị Lan mở được cửa ở lần thứ 4" thì ta có

$$A = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3} \cdot A_4$$

Ta có

$$P\left(\overline{A_1}\right) = \frac{C_7^1}{C_9^1} = \frac{7}{9}; P\left(\overline{A_2}|\overline{A_1}\right) = \frac{C_6^1}{C_8^1} = \frac{3}{4}; P\left(\overline{A_3}|\overline{A_1} \cdot \overline{A_2}\right) = \frac{C_5^1}{C_7^1} = \frac{5}{7}; P\left(A_4|\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}\right) = \frac{C_2^1}{C_6^1} = \frac{1}{3}$$

Suy ra

$$P(A) = P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}|\overline{A_1}\right) \cdot P\left(\overline{A_3}|\overline{A_1} \cdot \overline{A_2}\right) \cdot P\left(A_4|\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}\right) = \frac{7}{9} \cdot \frac{3}{4} \cdot \frac{5}{7} \cdot \frac{1}{3} = \frac{5}{36}$$

Bài tập 1.13. Một bộ đề thi vấn đáp gồm 10 đề, trong đó có 4 đề về câu hỏi lý thuyết và 6 đề bài tập tính toán. Có 3 sinh viên lần lượt vào thi, mỗi sinh viên chỉ được lấy 1 đề và không hoàn lại. Tìm xác suất để sinh viên 1 gặp đề bài tập, sinh viên 2 gặp đề lý thuyết và sinh viên 3 gặp đề bài tập.

Lời giải

Phép thử: Mỗi sinh viên lấy ngẫu nhiên không hoàn lại 1 đề.
Gọi A_i là biến cố "Sinh viên i gặp đề bài tập" với $i = \overline{1,3}$ thì $\overline{A_i}$ là biến cố "Sinh viên i gặp đề lý thuyết". Gọi A là biến cố "Sinh viên 1 gặp đề bài tập, sinh viên 2 gặp đề lý thuyết và sinh viên 3 gặp đề bài tập" thì ta có

$$A = A_1 \overline{A_2} A_3$$

Ta có

$$P(A_1) = \frac{C_6^1}{C_{10}^1} = \frac{3}{5}; P\left(\overline{A_2}|A_1\right) = \frac{C_4^1}{C_9^1} = \frac{4}{9}; P\left(A_3|A_1\overline{A_2}\right) = \frac{C_5^1}{C_8^1} = \frac{5}{8}$$

Suy ra

$$P(A) = P(A_1) \cdot P\left(\overline{A_2}|A_1\right) \cdot P\left(A_3|A_1\overline{A_2}\right) = \frac{3}{5} \cdot \frac{4}{9} \cdot \frac{5}{8} = \frac{1}{6}$$

Bài tập 1.14. Một cơ sở sản xuất mũ gồm có 3 tổ cùng sản xuất với tỉ lệ sản phẩm trong tổng số sản phẩm lần lượt là 20%, 30% và 50%. Tổ 1 có tỉ lệ phế phẩm là 5% sản phẩm tổ, tổ 2% và tổ 3 là 1%. Tất cả sản phẩm làm ra được xếp chung vào một kho. Lấy ngẫu nhiên một sản phẩm từ kho.
a) Tìm xác suất để sản phẩm đó là phế phẩm. Tỉ lệ phế phẩm của kho là bao nhiêu.
b) Biết sản phẩm là phế phẩm. Tìm xác suất để nó do tổ 2 sản xuất.

Lời giải

Phép thử: Lấy ngẫu nhiên một sản phẩm từ kho.
Gọi A_i là biến cố "Sản phẩm lấy ra được tổ i sản xuất" với $i = \overline{1,3}$ thì ta có $\{A_1, A_2, A_3\}$ tạo thành hệ biến cố đầy đủ với $P(A_1) = 0,2$; $P(A_2) = 0,3$ và $P(A_3) = 0,5$.
Gọi H là biến cố "Sản phẩm lấy ra là phế phẩm" thì theo đề bài, ta có $P(H|A_1) = 0,05$; $P(H|A_2) = 0,02$ và $P(H|A_3) = 0,01$.
a) Ta có

$$P(H) = P(A_1) \cdot P(H|A_1) + P(A_2) \cdot P(H|A_2) + P(A_3) \cdot P(H|A_3) = 0,2 \cdot 0,05 + 0,3 \cdot 0,02 + 0,5 \cdot 0,01 = 0,021$$

Vậy tỉ lệ phế phẩm của kho là $2,1\%$.
b) Ta có

$$P(A_2|H) = \frac{P(A_2) \cdot P(H|A_2)}{P(H)} = \frac{0,3 \cdot 0,02}{0,021} = \frac{2}{7}$$

Bài tập 1.15. Cho tỉ lệ người dân nghiện thuốc lá ở một vùng là 30%. Biết tỉ lệ người viêm họng trong số người nghiện thuốc lá là 60% và tỉ lệ viêm họng trong số người không hút thuốc lá là 40%. Chọn ngẫu nhiên một người trong vùng.
a) Giả sử người đó viêm họng. Tìm xác suất để người đó nghiện thuốc.
b) Giả sử người đó không viêm họng. Tìm xác suất để người đó nghiện thuốc.

Lời giải

Phép thử: Chọn ngẫu nhiên một người trong vùng.
Gọi N là biến cố "Người được chọn nghiện thuốc lá" thì ta có $P(N) = 0,3$. Suy raChọn ngẫu nhiên một người trong vùng. $\overline{N}$ là biến cố "Người được chọn không hút thuốc lá" và $P(\overline{N}) = 1 - P(N) = 0,7$.
Gọi H là biến cố "Người được chọn bị viêm họng" thì ta có $P(H|N) = 0,6$ và $P(H|\overline{N}) = 0,4$.
a) Ta có

$$P(H) = P(N) \cdot P(H|N) + P(\overline{N}) \cdot P(H|\overline{N}) = 0,3 \cdot 0,6 + 0,7 \cdot 0,4 = 0,46$$

Suy ra

$$P(N|H) = \frac{P(N) \cdot P(H|N)}{P(H)} = \frac{0,3 \cdot 0,6}{0,46} = \frac{9}{23}$$

b) Ta có

$$P(N|\overline{H}) = \frac{P(N) \cdot P(\overline{H}|N)}{P(\overline{H})} = \frac{P(N) \cdot [1 - P(H|N)]}{1 - P(H)} = \frac{0,3 \cdot (1 - 0,6)}{1 - 0,46} = \frac{2}{9}$$

2 Bài tập làm thêm

Bài tập 1.16. Điền các giá trị thích hợp vào ô trống

$P(A)$	$P(B)$	$P(A \cup B)$	$P(A \cap B)$	$P(A\|B)$	$P(B\|A)$
3/4		9/10	1/5		
			2/3	8/9	4/5
1/5	1/5				1/20
5/17	3/17		1/17		

Lời giải

$P(A)$	$P(B)$	$P(A \cup B)$	$P(A \cap B)$	$P(A\|B)$	$P(B\|A)$
3/4	7/20	9/10	1/5	4/7	4/15
5/6	3/4	11/12	2/3	8/9	4/5
1/5	1/5	39/100	1/100	1/20	1/20
5/17	3/17	7/17	1/17	1/3	1/5

Bài tập 1.17. Một nhân viên bán hàng mỗi năm đến bán hàng ở một công ty nọ. Xác suất để lần đầu bán được hàng là $0,8$. Nếu lần trước bán được hàng thì xác suất để lần sau bán được hàng là $0,9$; còn nếu lần trước không bán được hàng thì xác suất để lần sau bán được hàng là $0,4$.
a) Tìm xác suất để cả ba lần đều bán được hàng.
b) Tìm xác suất để có đúng hai lần bán được hàng.

Lời giải

Gọi A_i là biến cố "Bán được hàng ở lần thứ i" với $i = \overline{1,3}$.
a) Gọi A là biến cố "Cả ba lần đều bán được hàng" thì ta có

$$A = A_1A_2A_3$$

Vì lần bán sau phụ thuộc vào lần bán kề trước nên $P(A_3|A_1A_2) = P(A_3|A_2)$ nên ta có

$$P(A_1) = 0,8; P(A_2|A_1) = 0,9; P(A_3|A_1A_2) = P(A_3|A_2) = 0,9$$

Suy ra

$$P(A) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1A_2) = 0,8 \cdot 0,9 \cdot 0,9 = 0,648$$

b) Gọi B là biến cố "Có đúng hai lần bán được hàng" thì ta có

$$B = A_1A_2\overline{A_3} \cup A_1\overline{A_2}A_3 \cup \overline{A_1}A_2A_3$$

Vì lần bán sau phụ thuộc vào lần bán kề trước nên $P\left(\overline{A_3}|A_1A_2\right) = P\left(\overline{A_3}|A_2\right)$. Suy ra

$$\begin{aligned} P\left(A_1A_2\overline{A_3}\right) &= P(A_1) \cdot P(A_2|A_1) \cdot P\left(\overline{A_3}|A_1A_2\right) = P(A_1) \cdot P(A_2|A_1) \cdot P\left(\overline{A_3}|A_2\right) \\ &= 0,8 \cdot 0,9 \cdot (1 - 0,9) = 0,072 \end{aligned}$$

Tương tự với các biểu thức còn lại thì ta có

$$P\left(A_1\overline{A_2}A_3\right) = P(A_1) \cdot P\left(\overline{A_2}|A_1\right) \cdot P\left(A_3|\overline{A_2}\right) = 0,8 \cdot (1 - 0,9) \cdot 0,4 = 0,032$$

và

$$P\left(\overline{A_1}A_2A_3\right) = P\left(\overline{A_1}\right) \cdot P\left(A_2|\overline{A_1}\right) \cdot P(A_3|A_2) = (1 - 0,8) \cdot 0,4 \cdot 0,9 = 0,072$$

Do các biến cố $A_1 \cdot \overline{A_2} \cdot \overline{A_3}$, $\overline{A_1} \cdot A_2 \cdot \overline{A_3}$, $\overline{A_1} \cdot \overline{A_2} \cdot A_3$ đôi một xung khắc nên suy ra

$$\begin{aligned} P(B) &= P\left(A_1A_2\overline{A_3} \cup A_1\overline{A_2}A_3 \cup \overline{A_1}A_2A_3\right) = P\left(A_1A_2\overline{A_3}\right) + P\left(A_1\overline{A_2}A_3\right) + P\left(\overline{A_1}A_2A_3\right) \\ &= 0,072 + 0,032 + 0,072 = 0,176 \end{aligned}$$

Bài tập 1.18. Một test kiểm tra sự hiện diện của virus Covid19 cho kết quả dương tính nếu bệnh nhân thực sự nhiễm virus Covid19. Tuy nhiên test này cũng có sai sót, đôi khi cho kết quả dương tính đối với người không thực sự nhiễm virus, tỉ lệ sai sót là $\frac{1}{20.000}$. Giả sử có 10.000 người thì có 1 người bị nhiễm virus Covid19. Tìm tỉ lệ người đó có kết quả dương tính thực sự nhiễm Covid19.

Lời giải

Gọi A là biến cố "Người được chọn nhiễm Covid19" thì ta có $P(A) = \frac{1}{10000}$. Gọi B là biến cố "Người được chọn có kết quả kiểm tra dương tính" thì ta có $P(B|A) = 1$ và $P\left(B|\overline{A}\right) = \frac{1}{20000}$. Suy ra

$$\begin{aligned} P(A|B) &= \frac{P(A) \cdot P(B|A)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P\left(\overline{A}\right) \cdot P\left(B|\overline{A}\right)} = \frac{\frac{1}{10000} \cdot 1}{\frac{1}{10000} \cdot 1 + \left(1 - \frac{1}{10000}\right) \cdot \frac{1}{20000}} \\ &= 0,6667 \end{aligned}$$

Vậy tỉ lệ người đó có kết quả dương tính thực sự nhiễm Covid19 là $66,67\%$.

Bài tập 1.19. Người ta phỏng vấn ngẫu nhiên 500 khách hàng về một sản phẩm định đưa ra thị trường và thấy có 100 người trả lời "Sẽ mua", 150 người trả lời "Có thể sẽ mua" và 250 người trả lời "Không mua". Theo kinh nghiệm cho thấy tỉ lệ khách hàng thực sự mua sản phẩm tương ứng với những cách trả lời trên là 40%, 20% và 1%.
a) Hãy đánh giá thị trường tiềm năng của sản phẩm đó (theo nghĩa tỉ lệ người thực sự mua sản phẩm đó).
b) Trong số khách hàng thực sự mua sản phẩm đó có bao nhiêu phần trăm trả lời "Không mua".

Lời giải

Gọi A_1, A_2, A_3 lần lượt là các biến cố "Người trả lời sẽ mua", "Người trả lời có thể sẽ mua" và "Người trả lời không mua" thì $\{A_1, A_2, A_3\}$ tạo thành hệ biến cố đầy đủ với $P(A_1) = \frac{100}{500} = 0,2$; $P(A_2) = \frac{150}{500} = 0,3$ và $P(A_3) = \frac{250}{500} = 0,5$.
a) Gọi D là biến cố "Người thực sự mua sản phẩm" thì ta có $P(D|A_1) = 0,4$; $P(D|A_2) = 0,2$ và $P(D|A_3) = 0,01$. Suy ra

$$\begin{aligned} P(D) &= P(A_1) \cdot P(D|A_1) + P(A_2) \cdot P(D|A_2) + P(A_3) \cdot P(D|A_3) = 0,2 \cdot 0,4 + 0,3 \cdot 0,2 + 0,5 \cdot 0,01 \\ &= 0,145 \end{aligned}$$

Vậy tỉ lệ người thực sự mua sản phẩm đó là $14,5\%$.
b) Ta có

$$P(A_3|D) = \frac{P(A_3) \cdot P(D|A_3)}{P(D)} = \frac{0,5 \cdot 0,01}{0,145} = \frac{1}{29} = 0,0345$$

Vậy trong số những khách hàng thực sự mua sản phẩm thì có khoảng $3,45\%$ trả lời "Không mua".

Bài tập 1.20. Trước tình hình dịch bệnh Covid19 phức tạp, nhà trường tổ chức kỳ thi học kỳ II năm học 2021, môn Xác suất thống kê, hình thức thi online với hai bài thi là tự luận và vấn đáp trực tuyến. Giả sử, khả năng thi đạt bài thi tự luận của sinh viên An là $0,8$. Nếu sinh viên An thi đạt bài thi tự luận thì khả năng An thi đạt bài thi vấn đáp trực tuyến là $0,9$; còn nếu An không đạt bài thi tự luận thì khả năng An thi đạt bài thi vấn đáp trực tuyến là $0,3$. Tính xác suất cho các biến cố sau
a) A = "An thi đạt cả hai bài thi".
b) B = "An chỉ thi đạt một bài thi".

Lời giải

Gọi T_1, T_2 lần lượt là các biến cố "An đạt bài thi tự luận" và "An đạt bài thi vấn đáp trực tuyến" thì ta có $P(T_1) = 0,8$; $P(T_2|T_1) = 0,9$ và $P\left(T_2|\overline{T_1}\right) = 0,3$.
a) Gọi X là biến cố "An thi đạt cả hai bài thi" thì $X = T_1T_2$. Suy ra

$$P(X) = P(T_1T_2) = P(T_1) \cdot P(T_2|T_1) = 0,8 \cdot 0,9 = 0,72$$

b) Gọi Y là biến cố "An chỉ thi đạt một bài thi" thì $Y = T_1\overline{T_2} \cup \overline{T_1}T_2$. Do các biến cố $T_1\overline{T_2}$ và $\overline{T_1}T_2$ xung khắc nên suy ra

$$\begin{aligned} P(Y) &= P\left(T_1\overline{T_2} \cup \overline{T_1}T_2\right) = P\left(T_1\overline{T_2}\right) + P\left(\overline{T_1}T_2\right) = P(T_1) \cdot P\left(\overline{T_2}|T_1\right) + P\left(\overline{T_1}\right) \cdot P\left(T_2|\overline{T_1}\right) \\ &= 0,8 \cdot (1 - 0,9) + (1 - 0,8) \cdot 0,3 = 0,14 \end{aligned}$$

Chương II

Đại lượng ngẫu nhiên

1 Bài tập chương 2

Bài tập 2.1. Một dây chuyền gồm 3 bộ phận hoạt động độc lập trong thời gian 1 năm. Xác suất các bộ phận 1, 2, 3 bị hỏng trong thời gian hoạt động 1 năm tương ứng là $0,4$; $0,2$ và $0,3$. Gọi X là số bộ phận bị hỏng.
a) Lập bảng phân phối xác suất của đại lượng ngẫu nhiên X.
b) Tính xác suất có không quá 2 bộ phận bị hỏng.
c) Tính $E[X]$, $D[X]$, $\sigma(X)$ và giá trị tin chắc nhất của X.

Lời giải

a) Ta có

$$P(X=0) = (1-0,4)\cdot(1-0,2)\cdot(1-0,3) = 0,336$$
$$P(X=1) = 0,4\cdot(1-0,2)\cdot(1-0,3) + (1-0,4)\cdot 0,2\cdot(1-0,3) + (1-0,4)\cdot(1-0,2)\cdot 0,3 = 0,452$$
$$P(X=2) = 0,4\cdot 0,2\cdot(1-0,3) + (1-0,4)\cdot 0,2\cdot 0,3 + 0,4\cdot(1-0,2)\cdot 0,3 = 0,188$$
$$P(X=3) = 0,4\cdot 0,2\cdot 0,3 = 0,024$$

Từ đây ta có bảng phân phối xác suất

X_i	0	1	2	3	Σ
$P(X=X_i)$	$0,336$	$0,452$	$0,188$	$0,024$	1

b) Ta có

$$P(X\le 2) = 1 - P(X>2) = 1 - P(X=3) = 1 - 0,024 = 0,976$$

c) Ta có

$$E[X] = 0\cdot 0,336 + 1\cdot 0,452 + 2\cdot 0,188 + 3\cdot 0,024 = 0,9$$
$$E\left[X^2\right] = 0^2\cdot 0,336 + 1^2\cdot 0,452 + 2^2\cdot 0,188 + 3^2\cdot 0,024 = 1,42 \Rightarrow D[X] = E\left[X^2\right] - E^2[X] = 0,61$$
$$\sigma(X) = \sqrt{D[X]} = 0,781$$
$$\text{Mod}(X) = 1$$

Bài tập 2.2. Lãi suất thu được trong một năm (%) khi đầu tư vào công ty A, công ty B tương ứng là các đại lượng ngẫu nhiên độc lập X, Y. Cho biết quy luật phân phối xác suất của X và Y như sau

X	4	6	8	10	12
P	$0,05$	$0,1$	$0,3$	$0,4$	$0,15$

Y	-4	2	8	10	12	16
P	$0,1$	$0,2$	$0,2$	$0,25$	$0,15$	$0,1$

a) Đầu tư vào công ty nào có lãi suất trung bình (lãi suất kỳ vọng) cao hơn.
b) Đầu tư vào công ty nào có mức độ rủi ro ít hơn.
c) Đầu tư vào cả hai công ty theo tỉ lệ nào để ít rủi ro nhất.

Lời giải

a) Ta có

$$E[X] = 4 \cdot 0,05 + 6 \cdot 0,1 + 8 \cdot 0,3 + 10 \cdot 0,4 + 12 \cdot 0,15 = 9$$

và

$$E[Y] = (-4) \cdot 0,1 + 2 \cdot 0,2 + 8 \cdot 0,2 + 10 \cdot 0,25 + 12 \cdot 0,15 + 16 \cdot 0,1 = 7,5$$

Ta thấy $E[X] > E[Y]$ nên đầu tư vào công ty A có lãi suất trung bình cao hơn.

b) Ta có

$$D[X] = E\left[X^2\right] - E^2[X] = \left(4^2 \cdot 0,05 + 6^2 \cdot 0,1 + 8^2 \cdot 0,3 + 10^2 \cdot 0,4 + 12^2 \cdot 0,15\right) - 9^2 = 4,2$$

và

$$\begin{aligned} D[Y] &= E\left[Y^2\right] - E^2[Y] \\ &= \left[(-4)^2 \cdot 0,1 + 2^2 \cdot 0,2 + 8^2 \cdot 0,2 + 10^2 \cdot 0,25 + 12^2 \cdot 0,15 + 16^2 \cdot 0,1\right] - (7,5)^2 \\ &= 31,15 \end{aligned}$$

Ta thấy $D[X] < D[Y]$ nên đầu tư vào công ty A có mức độ rủi ro ít hơn.

c) Gọi p là tỉ lệ đầu tư vào công ty A với $p \in [0,1]$. Khi đó tỉ lệ đầu tư vào công ty B là $1-p$. Gọi Z là lãi suất thu được khi đầu tư theo phương án này thì ta có

$$Z = pX + (1-p)Y$$

Suy ra

$$D[Z] = D[pX + (1-p)Y] = p^2 D[X] + (1-p)^2 D[Y] = 81p^2 + 31,15(1-p)^2 = f(p)$$

Yêu cầu bài toán tương đương $D(Z)$ đạt giá trị nhỏ nhất. Ta có

$$f(p) = 112,15p^2 - 62,3p + 31,15 = \frac{2243}{20}\left(p - \frac{623}{2243}\right)^2 + \frac{50463}{2243} \geq \frac{50463}{2243}$$

Dấu bằng xảy ra khi $p = \dfrac{623}{2243} = 27,77\%$. Vậy nên đầu tư $27,77\%$ vào công ty A và $72,23\%$ vào công ty B để ít rủi ro nhất.

Bài tập 2.3. Số tiền lời trong năm tới (tính theo đơn vị: triệu đồng) thu được khi đầu tư 100 triệu đồng vào hai ngành A và B tùy thuộc vào tình hình kinh tế (THKT) trong nước và cho ở bảng sau

THKT / Số tiền lời	Kém phát triển	Khá ổn định	Phát triển
Ngành A	10	40	80
Ngành B	-30	70	100
Dự báo xác suất THKT	$0,25$	$0,45$	$0,3$

a) Số tiền lời trung bình (số tiền lời kì vọng) ngành nào là cao hơn.

b) Mức độ rủi ro tiền lời ngành nào là ít hơn. (Mức độ rủi ro của tiền lời có thể hiểu là độ phân tán của tiền lời xung quanh tiền lời trung bình và chỉ ra bởi độ lệch chuẩn của tiền lời hoặc phương sai của tiền lời).

Lời giải

a) Gọi X, Y lần lượt là tiền lời khi đầu tư vào ngành A và B. Ta có

$$E[X] = 10 \cdot 0,25 + 40 \cdot 0,45 + 80 \cdot 0,3 = 44,5$$

và

$$E[Y] = (-30) \cdot 0,25 + 70 \cdot 0,45 + 100 \cdot 0,3 = 54$$

Ta thấy $E[Y] > E[X]$ nên đầu tư vào ngành B thì tiền lời cao hơn.
b) Ta có

$$D[X] = E\left[X^2\right] - E^2[X]\left(10^2 \cdot 0,25 + 40^2 \cdot 0,45 + 80^2 \cdot 0,3\right) - 44,5^2 = 684,75$$

và

$$D[Y] = E\left[Y^2\right] - E^2[Y] = \left[(-30)^2 \cdot 0,25 + 70^2 \cdot 0,45 + 100^2 \cdot 0,3\right] - 54^2 = 2514$$

Ta thấy $D[X] < D[Y]$ nên đầu tư vào ngành A thì ít rủi ro hơn.

Bài tập 2.4. Cho X (nghìn sản phẩm) là nhu cầu mỗi năm về một loại hàng A ở nước ta với X là một đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất

$$f(x) = \begin{cases} k(30-x) & \text{với } x \in (0,30) \\ 0 & \text{với } x \notin (0,30) \end{cases}$$

a) Tìm k.
b) Tìm nhu cầu trung bình hàng năm của loại hàng A.
c) Tìm xác suất để nhu cầu về mặt hàng A không vượt quá 12 nghìn sản phẩm trong một năm.
d) Tìm phương sai và độ lệch (chuẩn) của X.

Lời giải

a) Vì $f(x)$ là hàm mật độ xác suất nên ta có

$$\begin{cases} k(30-x) \geq 0, \forall x \in (0,30) \\ \int\limits_{-\infty}^{+\infty} k(30-x)\mathrm{d}x = 1 \end{cases} \Rightarrow \begin{cases} k \geq 0 \\ \int\limits_{0}^{30} k(30-x)\mathrm{d}x = 1 \end{cases} \Rightarrow k = \frac{1}{450}$$

Từ đây ta có hàm phân phối xác suất của X là

$$F(x) = \int\limits_{-\infty}^{x} f(t)\mathrm{d}t = \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{15} - \dfrac{x^2}{90} & 0 < x < 30 \\ 1 & x \geq 30 \end{cases}$$

b) Ta có

$$E[X] = \int\limits_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \int\limits_{0}^{30} \frac{x(30-x)}{450}\mathrm{d}x = 10$$

Vậy nhu cầu trung bình hàng năm của loại hàng A là 10 nghìn sản phẩm.
c) Ta có

$$P(X \leq 12) = F(12) = 0,64$$

d) Ta có

$$D[X] = E\left[X^2\right] - E^2[X] = \int\limits_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x - \left[\int\limits_{-\infty}^{+\infty} xf(x)\mathrm{d}x\right]^2 = \int\limits_{0}^{30} \frac{x^2(30-x)}{450}\mathrm{d}x - 10^2 = 50$$

Suy ra $\sigma(X) = \sqrt{D[X]} = 5\sqrt{2}$.

Bài tập 2.5. Tỉ lệ sản phẩm tốt của một lô hàng lớn là 90%. Kiểm tra ngẫu nhiên 10 sản phẩm. Gọi X là số sản phẩm tốt trong 10 sản phẩm lấy ra.
a) Chỉ ra quy luật phân phối xác suất của X.
b) Số sản phẩm tốt trung bình.
c) Tính xác suất để có nhiều nhất là 2 sản phẩm tốt trong 10 sản phẩm lấy ra.

Lời giải

a) X có phân phối nhị thức với $X \sim \mathcal{B}(10; 0,9)$.
b) Ta có

$$E[X] = 10 \cdot 0,9 = 9$$

Vậy trung bình có 9 sản phẩm tốt trong 10 sản phẩm lấy ra.
c) Ta có

$$\begin{aligned} P(X \leq 2) &= P(X=2) + P(X=1) + P(X=0) \\ &= C_{10}^2 \cdot (0,9)^2 \cdot (1-0,9)^8 + C_{10}^1 \cdot (0,9)^1 \cdot (1-0,9)^9 + C_{10}^0 \cdot (0,9)^0 \cdot (1-0,9)^{10} \\ &= 3,736 \cdot 10^{-7} \end{aligned}$$

Bài tập 2.6. Bắn 5 viên đạn vào mục tiêu. Xác suất trúng đích của mỗi lần bắn là như nhau và bằng 0, 2. Muốn phá hủy mục tiêu thì phải có ít nhất 3 viên đạn trúng mục tiêu.
a) Chỉ ra phân phối xác suất của số viên đạn trúng mục tiêu.
b) Trung bình có bao nhiêu viên trúng mục tiêu.
c) Tìm xác suất mục tiêu bị phá hủy.

Lời giải

a) Gọi X là số viên đạn trúng mục tiêu thì X có phân phối nhị thức với $X \sim \mathcal{B}(5; 0,2)$.
b) Ta có

$$E[X] = 5 \cdot 0,2 = 1$$

Vậy trung bình có 1 viên trúng mục tiêu.
c) Xác suất mục tiêu bị phá hủy là

$$\begin{aligned} P(X \geq 3) &= P(X=3) + P(X=4) + P(X=5) \\ &= C_5^3 \cdot (0,2)^3 \cdot (1-0,2)^2 + C_5^4 \cdot (0,2)^4 \cdot (1-0,2)^1 + C_5^5 \cdot (0,2)^5 \cdot (1-0,2)^0 \\ &= 0,05792 \end{aligned}$$

Bài tập 2.7. Một đại lý điện thoại di động dự định sẽ áp dụng một trong 2 phương án kinh doanh. Gọi X_1, X_2 (triệu đồng/tháng) tương ứng là lợi nhuận thu được khi áp dụng phương án thứ nhất, phương án thứ hai. Giả sử $X_1 \sim \mathcal{N}(140; 2.500)$ và $X_2 \sim \mathcal{N}(200; 3.600)$. Nếu biết rằng để đại lý tồn tại và phát triển thì lợi nhuận thu được từ kinh doanh điện thoại phải đạt ít nhất 80 (triệu đồng/tháng). Theo bạn công ty nên áp dụng phương án nào để kinh doanh điện thoại di động. Vì sao.

Lời giải

Ta có

$$P(X_1 \geq 80) = \phi(+\infty) - \phi\left(\frac{80-140}{\sqrt{2500}}\right) = \phi(+\infty) - \phi(-1,2) = 0,5 + 0,3849 = 0,8849$$

và

$$P(X_2 \geq 80) = \phi(+\infty) - \phi\left(\frac{80-200}{\sqrt{3600}}\right) = \phi(+\infty) - \phi(-2) = 0,5 + 0,4772 = 0,9772$$

Vì $P(X_2 \geq 80) > P(X_1 \geq 80)$ nên công ty nên áp dụng phương án thứ hai.

Bài tập 2.8. Một người cân nhắc giữa việc mua nhà bây giờ hay gửi tiết kiệm với lãi suất một năm là 12% để chờ năm sau sẽ mua. Biết mức tăng giá nhà sau một năm $X(\%)$ là đại lượng ngẫu nhiên có phân phối chuẩn với $E[X] = 8(\%)$ và $\sigma(X) = 10(\%)$. Giả sử người đó quyết định gửi tiền vào tiết kiệm. Tìm khả năng để quyết định để quyết định đó là sai lầm.

Lời giải

Theo đề bài thì ta có $X \sim \mathcal{N}\left(8; 10^2\right)$. Để quyết định đó là sai lầm thì ta phải có $X > 12(\%)$. Suy ra

$$P(X > 12) = \phi(+\infty) - \phi\left(\frac{12-8}{10}\right) = \phi(+\infty) - \phi(0,4) = 0,5 - 0,1554 = 0,3446$$

Bài tập 2.9. Biết tuổi thọ X (năm) của mỗi sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn với tuổi thọ trung bình là 8 năm và độ lệch chuẩn là 2 năm. Nếu bán được một sản phẩm thì cửa hàng lãi 150 (ngàn). Còn nếu sản phẩm bị hỏng trong thời gian bảo hành thì cửa hàng phải chi trả lại 500 (ngàn) cho bên bảo hành. Thời gian bảo hành mỗi sản phẩm được quy định là 6 năm.
a) Tìm tỉ lệ sản phẩm bị bảo hành.
b) Tiền lãi trung bình cửa hàng thu được sau khi bán mỗi sản phẩm.

Lời giải

a) Từ đề bài thì ta có $X \sim \mathcal{N}\left(8; 2^2\right)$. Khi đó tỉ lệ sản phẩm bị bảo hành là

$$P(X \leq 6) = \phi\left(\frac{6-8}{2}\right) - \phi(-\infty) = \phi(-1) - \phi(-\infty) = -0,3413 + 0,5 = 0,1587$$

Vậy tỉ lệ sản phẩm bị bảo hành là $15,87\%$.
b) Gọi T (ngàn đồng) là tiền lãi thu được sau khi bán mỗi sản phẩm thì ta có bảng phân phối xác suất của T là

T_i	150	-350	Σ
$P(T = T_i)$	$1 - 0,1587$	$0,1587$	1

Suy ra

$$E[T] = 150 \cdot (1 - 0,1587) + (-350) \cdot 0,1587 = 70,65$$

Vậy tiền lãi trung bình cửa hàng thu được sau khi bán mỗi sản phẩm là $70,65$ (ngàn).

2 Bài tập làm thêm

Bài tập 2.10. Theo thống kê, tỉ lệ để một người độ tuổi 40 sống thêm ít nhất 1 năm nữa là $99,5\%$. Một công ty nhân thọ bán mỗi thẻ bảo hiểm 1 năm cho mỗi người độ tuổi đó với 10 (ngàn đồng) và trong trường hợp người mua bảo hiểm chết sẽ có tiền bồi thường là 1 triệu đồng. Tìm lợi nhuận kì vọng của công ty bảo hiểm khi bán mỗi thẻ bảo hiểm loại này.

Lời giải

Gọi X (ngàn đồng) là lợi nhuận của công ty khi bán mỗi thẻ bảo hiểm loại này thì ta có bảng phân phối xác suất của X như sau

X_i	10	-990	Σ
$P(X = X_i)$	$0,995$	$1 - 0,005$	1

Suy ra

$$E[X] = 10 \cdot 0,995 + (-350) \cdot (1 - 0,995) = 5$$

Vậy lợi nhuận kì vọng của công ty bảo hiểm khi bán mỗi thẻ bảo hiểm loại này là 5 (ngàn đồng).

Bài tập 2.11. Tỉ lệ bị cận thị của học sinh Việt Nam là 9%. Chọn ngẫu nhiên 100 học sinh. Gọi X là số học sinh bị cận thị.
a) Chỉ ra quy luật phân phối xác suất của X.
b) Số học sinh bị cận thị trung bình.
c) Tính xác suất có ít nhất 1 học sinh bị cận thị.
d) Tính xác suất có nhiều nhất 11 học sinh bị cận thị.

Lời giải

a) X có phân phối nhị thức với $X \sim \mathcal{B}(100; 0,09)$.
b) Ta có

$$E[X] = 100 \cdot 0,09 = 9$$

Vậy trung bình có 9 học sinh bị cận thị.
c) Xác suất để ít nhất 1 học sinh bị cận thị là

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - C_{100}^{0} \cdot (0,09)^{0} \cdot (1 - 0,09)^{100} = 0,99992$$

d) Ta thấy $np < 10$ nên ta xấp xỉ X thành phân phối Poisson với tham số

$$\lambda = 100 \cdot 0,09 = 9$$

Vậy $X \sim \mathcal{P}(9)$. Khi đó ta có

$$P(X \leq 11) = \sum_{i=0}^{11} P(X = i) = e^{-9} \sum_{k=0}^{11} \frac{9^k}{k!} = 0,803$$

Bài tập 2.12. Thời gian hoạt động tốt X (giờ) của một TV cùng loại là một đại lượng ngẫu nhiên phân phối chuẩn, với trung bình $\mu = 4.300$ (giờ) và độ lệch chuẩn $\sigma = 250$ (giờ). Giả thiết mỗi ngày trung bình người ta dùng TV là 10 (giờ) và thời hạn bảo hành miễn phí là 360 ngày. Tính tỉ lệ sản phẩm phải bảo hành.

Lời giải

Từ đề bài thì ta có $X \sim \mathcal{N}\left(4300; 250^2\right)$. Để sản phẩm được bảo hành miễn phí thì $X \leq 360 \cdot 10 = 3600$. Suy ra

$$P(X \leq 3600) = \phi\left(\frac{3600 - 4300}{250}\right) - \phi(-\infty) = \phi(-2,8) - \phi(-\infty) = -0,4974 + 0,5 = 0,0026$$

Vậy tỉ lệ sản phẩm được bảo hành miễn phí là $0,26\%$.

Bài tập 2.13. Gọi X (kWh) là lượng điện tiêu thụ mỗi tháng của mỗi hộ gia đình ở miền Trung. Biết X có phân phối chuẩn với trung bình 160 kWh, độ lệch chuẩn là 40 kWh. Giả sử trong 50 kWh đầu tiên phải trả 1.000 đồng cho mỗi kWh điện. Những kWh điện tiêu thụ tiếp theo phải trả 2.000 đồng cho mỗi kWh điện.
a) Tính tỉ lệ hộ gia đình tiêu thụ dưới 90 kWh trong 1 tháng.
b) Tính tỉ lệ hộ gia đình trả tiền điện trong 1 tháng nhiều hơn 300.000 đồng.

Lời giải

a) Từ đề bài thì ta có $X \sim \mathcal{N}\left(160; 40^2\right)$. Khi đó ta có

$$P(X \leq 90) = \phi\left(\frac{90 - 160}{40}\right) - \phi(-\infty) = \phi(-1,75) - \phi(-\infty) = -0,4599 + 0,5 = 0,0401$$

Vậy tỉ lệ hộ gia đình tiêu thụ dưới 90 kWh trong 1 tháng là $4,01\%$.
b) Gọi T là số tiền điện phải trả trong 1 tháng. Vì $300.000 > 50 \cdot 1.000$ nên ta có

$$T = 50 \cdot 1.000 + (X - 50) \cdot 2.000 = 2.000X - 50.000$$

Theo đề bài thì ta có

$$T > 300.000 \Leftrightarrow 2.000X - 50.000 > 300.000 \Leftrightarrow X > 175$$

Suy ra

$$\begin{aligned} P(T > 300.000) &= P(X > 175) \\ &= \phi(+\infty) - \phi\left(\frac{175 - 160}{40}\right) = \phi(+\infty) - \phi(0,375) = 0,5 - 0,1462 = 0,3538 \end{aligned}$$

Vậy tỉ lệ hộ gia đình trả tiền điện trong 1 tháng nhiều hơn 300.000 đồng là $35,38\%$.

Bài tập 2.14. Gọi $X(mm)$ là chiều dài mỗi sản phẩm do một phân xưởng sản xuất. Biết rằng X có phân phối chuẩn với độ lệch chuẩn là $0,5mm$. Sản phẩm gọi là đạt chất lượng cao nếu chiều dài sản phẩm sai lệch chiều dài trung bình không qua $0,1mm$.
a) Tìm tỉ lệ sản phẩm đạt chất lượng cao của phân xưởng.
b) Chọn ngẫu nhiên 10 sản phẩm. Tính xác suất có ít nhất 2 sản phẩm đạt chất lượng cao.
c) Trung bình số sản phẩm đạt chất lượng cao ở câu b là bao nhiêu.

Lời giải

a) Từ đề bài thì ta có $X \sim \mathcal{N}\left(\mu; (0,5)^2\right)$. Khi đó tỉ lệ sản phẩm đạt chất lượng cao của phân xưởng là

$$P\left(|X-\mu| \leq 0,1\right) = 2\phi\left(\frac{0,1}{0,5}\right) = 2\phi\left(0,2\right) = 2 \cdot 0,0793 = 0,1586$$

Vậy tỉ lệ sản phẩm đạt chất lượng cao của phân xưởng là $15,86\%$.
b) Gọi Y là số lượng sản phẩm đạt chất lượng cao thì ta có $Y \sim \mathcal{B}\left(10; 0,1586\right)$. Suy ra

$$\begin{aligned} P\left(Y \geq 2\right) &= 1 - P\left(Y < 2\right) = 1 - P\left(Y = 1\right) - P\left(Y = 0\right) \\ &= 1 - C_{10}^{1} \cdot (0,1586)^1 \cdot (1-0,1586)^9 - C_{10}^{0} \cdot (0,1586)^0 \cdot (1-0,1586)^{10} \\ &= 0,4869 \end{aligned}$$

c) Ta có

$$E\left[Y\right] = 10 \cdot 0,1568 = 1,568$$

Vậy trung bình có 2 sản phẩm đạt chất lượng cao.

Bài tập 2.15. Gọi $X(kg)$ là cân nặng mỗi con heo của các trang trại lớn ở một tỉnh. Biết rằng X có phân phối chuẩn với cân nặng trung bình là $80kg$ và độ lệch chuẩn là $10kg$. Heo được gọi là loại I nếu có cân nặng hơn $90kg$.
a) Tìm tỉ lệ heo đạt loại I ở các trang trại đó.
b) Chọn ngẫu nhiên 8 con heo. Tìm xác suất để có ít nhất 2 con đạt loại I.
c) Trung bình số heo đạt loại I ở câu b là bao nhiêu.

Lời giải

a) Từ đề bài thì ta có $X \sim \mathcal{N}\left(80; 10^2\right)$. Khi đó tỉ lệ heo đạt loại I ở trang trại là

$$P\left(X \geq 90\right) = \phi\left(+\infty\right) - \phi\left(\frac{90-80}{10}\right) = \phi\left(+\infty\right) - \phi\left(1\right) = 0,5 - 0,3413 = 0,1587$$

Vậy tỉ lệ heo đạt loại I ở các trang trại trên là $15,87\%$.
b) Gọi Y là số lượng heo đạt loại I thì ta có $Y \sim \mathcal{B}\left(8; 0,1587\right)$. Suy ra

$$\begin{aligned} P\left(Y \geq 2\right) &= 1 - P\left(Y < 2\right) = 1 - P\left(Y = 1\right) - P\left(Y = 0\right) \\ &= 1 - C_{8}^{1} \cdot (0,1587)^1 \cdot (1-0,1587)^7 - C_{8}^{0} \cdot (0,1587)^0 \cdot (1-0,1587)^{8} \\ &= 0,3703 \end{aligned}$$

c) Ta có

$$E\left[Y\right] = 8 \cdot 0,1567 = 1,2696$$

Vậy trung bình có 2 con heo đạt loại I.

Bài tập 2.16. Lãi suất $X(\%)$ đầu tư vào một dự án được xem như một đại lượng ngẫu nhiên phân phối theo quy luật chuẩn. Theo đánh giá của ủy ban đầu tư thì lãi suất cao hơn 20% có xác suất là $0,1587$ và lãi suất cao hơn 25% có xác suất là $0,0228$. Vậy khả năng đầu tư mà không bị thua lỗ là bao nhiêu.

Lời giải

Theo đề bài thì ta đặt $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$. Khi đó giả thiết $P(X > 20) = 0,1587$ tương đương với

$$0,1587 = \phi(+\infty) - \phi\left(\frac{20-\mu}{\sigma}\right) = 0,5 - \phi\left(\frac{20-\mu}{\sigma}\right) \Leftrightarrow \phi\left(\frac{20-\mu}{\sigma}\right) = 0,3413 \Leftrightarrow \frac{20-\mu}{\sigma} = 1$$

Tương tự, với $P(X > 25) = 0,0228$ thì ta có

$$0,0228 = \phi(+\infty) - \phi\left(\frac{25-\mu}{\sigma}\right) = 0,5 - \phi\left(\frac{25-\mu}{\sigma}\right) \Leftrightarrow \phi\left(\frac{25-\mu}{\sigma}\right) = 0,4772 \Leftrightarrow \frac{25-\mu}{\sigma} = 2$$

Vậy ta có hệ

$$\begin{cases} \dfrac{20-\mu}{\sigma} = 1 \\ \dfrac{25-\mu}{\sigma} = 2 \end{cases} \Leftrightarrow \begin{cases} \mu + \sigma = 20 \\ \mu + 2\sigma = 25 \end{cases} \Leftrightarrow \begin{cases} \mu = 15 \\ \sigma = 5 \end{cases} \Rightarrow X \sim \mathcal{N}\left(15;5^2\right)$$

Để đầu tư không bị lỗ thì ta có $X \geq 0$. Suy ra

$$P(X \geq 0) = \phi(+\infty) - \phi\left(\frac{0-15}{5}\right) = \phi(+\infty) - \phi(-3) = 0,5 + 0,4987 = 0,9987$$

Vậy khả năng đầu tư mà không bị thua lỗ là $99,87\%$.

Bài tập 2.17. Độ dài chi tiết $X(cm)$ do máy sản xuất là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn $10cm$. Biết tỉ lệ chi tiết có độ dài dưới $84cm$ là $84,13\%$.
a) Tìm độ dài trung bình mỗi chi tiết.
b) Tìm tỉ lệ chi tiết có độ dài hơn $80cm$.
c) Lấy ngẫu nhiên 4 chi tiết. Tìm xác suất có ít nhất 1 chi tiết có độ dài hơn $80cm$.

Lời giải

a) Theo đề bài thì ta có $X \sim \mathcal{N}\left(\mu;10^2\right)$. Khi đó giả thiết $P(X < 84) = 84,13\% = 0,8413$ viết lại thành

$$0,8413 = \phi\left(\frac{84-\mu}{10}\right) - \phi(-\infty) = \phi\left(\frac{84-\mu}{10}\right) + 0,5 \Leftrightarrow \phi\left(\frac{84-\mu}{10}\right) = 0,3413 \Leftrightarrow \frac{84-\mu}{10} = 1 \Leftrightarrow \mu = 74$$

Vậy ta có $X \sim \mathcal{N}\left(74;10^2\right)$ nên độ dài trung bình mỗi chi tiết là $E[X] = \mu = 74cm$.
b) Ta có

$$P(X > 80) = \phi(+\infty) - \phi\left(\frac{80-74}{10}\right) = \phi(+\infty) - \phi(0,6) = 0,5 - 0,2257 = 0,2743$$

Vậy tỉ lệ chi tiết có độ dài hơn $80cm$ là $27,43\%$.
c) Gọi Y là số chi tiết có độ dài hơn $80cm$ thì ta có $Y \sim \mathcal{B}(4;0,2743)$. Suy ra

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1 - C_4^0 \cdot (0,2743)^0 \cdot (1-0,2743)^4 = 0,7226$$

Bài tập 2.18. Chị A nuôi 160 con vịt đẻ cùng loại. Xác suất để 1 con vịt đẻ trứng trong ngày là $0,8$. Quy ước mỗi con vịt chỉ đẻ nhiều nhất 1 quả trứng.
a) Chỉ ra quy luật phân phối xác suất của số trứng vịt đẻ trong ngày.
b) Tìm xác suất để chị A có được ít nhất 130 trứng trong ngày.
c) Nếu mỗi quả trứng bán được 2.000 đồng, tiền thức ăn cho vịt trong ngày là 900 đồng thì số tiền lãi trung bình chị A thu được trong ngày là bao nhiêu.

Lời giải

a) Gọi X là số trứng vịt đẻ trong ngày thì X có phân phối nhị thức với $X \sim \mathcal{B}(160;0,8)$.
b) Ta thấy $\begin{cases} np = 160 \cdot 0,8 = 128 > 5 \\ n(1-p) = 160 \cdot 0,2 = 32 > 5 \end{cases}$ nên ta sẽ xấp xỉ $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$ với

$$\begin{cases} \mu = np = 160 \cdot 0,8 = 128 \\ \sigma^2 = np(1-p) = 160 \cdot 0,8 \cdot 0,2 = 25,6 \end{cases}$$

Vậy $X \sim \mathcal{N}(128; 25,6)$. Suy ra

$$P(X \geq 130) = \phi(+\infty) - \phi\left(\frac{130-128}{\sqrt{25,6}}\right) = \phi(+\infty) - \phi(0,39) = 0,5 - 0,1517 = 0,3483$$

c) Gọi T là tiền lãi chị A thu được trong ngày thì ta có

$$T = 2000X - 900 \cdot 160 = 2000X - 144000$$

Suy ra

$$E[T] = 2000 \cdot E[X] - 144000 = 2000 \cdot 128 - 144000 = 112000$$

Vậy tiền lãi trung bình chị A thu được trong ngày là 112 ngàn đồng.

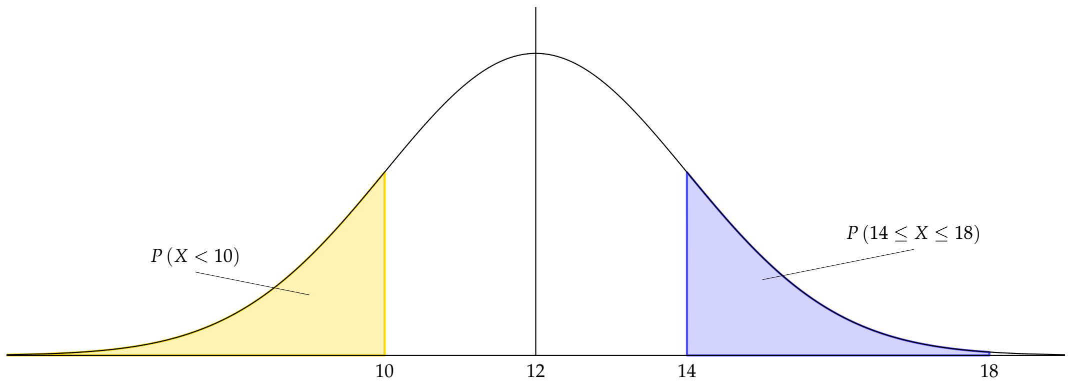
Bài tập 2.19. Gọi X (m^3/tháng) là lượng nước sử dụng của mỗi hộ gia đình ở thành phố Nha Trang là biến ngẫu nhiên có phân phối chuẩn với trung bình là $\mu = 12(m^3)$ và độ lệch chuẩn $\sigma = 2(m^3)$.
a) Tính $P(X < 10)$ và $P(4 \leq X \leq 18)$. Minh họa các kết quả trên bằng hình vẽ.
b) Từ $0(m^3)$ đến $11(m^3)$ có đơn giá là 8.000 đồng/m^3. Trên $11(m^3)$ có đơn giá là 10.000 đồng/m^3. Tìm tỉ lệ hộ gia đình ở Nha Trang phải trả từ 158.000 đồng/tháng trở lên.

Lời giải

a) Theo đề bài thì ta có $X \sim \mathcal{N}\left(12; 2^2\right)$. Suy ra

- $P(X < 10) = \phi\left(\frac{10-12}{2}\right) - \phi(-\infty) = \phi(-1) - \phi(-\infty) = -0,3413 + 0,5 = 0,1587.$
- $P(14 \leq X \leq 18) = \phi\left(\frac{18-12}{2}\right) - \phi\left(\frac{14-12}{2}\right) = \phi(3) - \phi(1) = 0,4987 - 0,3413 = 0,1574.$

Hình vẽ minh họa



b) Gọi T là số tiền nước phải trả trong 1 tháng. Vì $158.000 > 11 \cdot 8.000$ nên ta có

$$T = 11 \cdot 8.000 + (X - 11) \cdot 10.000 = 10.000X - 22.000$$

Theo đề bài thì ta có

$$T > 158.000 \Leftrightarrow 10.000X - 22.000 > 158.000 \Leftrightarrow X > 18$$

Suy ra

$$P(T > 158.000) = P(X > 18) = \phi(+\infty) - \phi\left(\frac{18-12}{2}\right) = \phi(+\infty) - \phi(3) = 0,5 - 0,4987 = 0,0013$$

Vậy tỉ lệ hộ gia đình ở Nha Trang phải trả từ 158.000 đồng/tháng trở lên là $0,13\%$.

Bài tập 2.20. Cho hai hộp. Hộp 1 có 1 tờ 500.000 đồng, 9 tờ 20.000 đồng; hộp 2 có 5 tờ 100.000 đồng, 2 tờ 50.000 đồng và 3 tờ 10.000 đồng. Anh Lâm được quyền chọn một trong hai hộp rồi lấy ngẫu nhiên ra một tờ tiền. Nếu lấy tời tiền nào thì được nhận tờ tiền đó. Hỏi rằng anh Lâm nên chọn hộp nào thì có lợi hơn (theo nghĩa kỳ vọng số tiền anh Lâm được nhận là cao hơn).

Lời giải

Gọi X, Y (ngàn đồng) lần lượt là số tiền mà anh Lâm nhận được khi chọn hộp 1 và hộp 2. Khi đó ta có bảng phân phối xác suất của X và Y như sau

X_i	500	20	Σ
$P(X = X_i)$	$0,1$	$0,9$	1

Y_i	100	50	10	Σ
$P(Y = Y_i)$	$0,5$	$0,2$	$0,3$	1

Suy ra

$$E[X] = 500 \cdot 0,1 + 20 \cdot 0,9 = 68$$

và

$$E[Y] = 100 \cdot 0,5 + 50 \cdot 0,2 + 10 \cdot 0,3 = 63$$

Vì $E[X] > E[Y]$ nên anh Lâm nên chọn hộp 1 thì có lợi hơn.

Chương III

Mẫu thống kê và ước lượng tham số

1 Bài tập chương 3

Bài tập 3.1. Quan sát ngẫu nhiên về thời gian cần thiết để sản xuất một sản phẩm trong một phân xưởng, ta thu được các số liệu cho ở bảng

Khoảng thời gian (giây)	Số quan sát	Khoảng thời gian (giây)	Số quan sát
20 – 22	2	26 – 28	32
22 – 24	10	28 – 30	14
24 – 26	34	30 – 32	8

Ước lượng thời gian trung bình để sản xuất 1 sản phẩm của phân xưởng với độ tin cậy 95%.

Lời giải

Gọi X (giây) là thời gian để sản xuất 1 sản phẩm của phân xưởng và μ (giây) là thời gian trung bình để sản xuất. Từ bảng dữ liệu trên đề bài, ta tính được $n = 100$, $\overline{x} = 26,4$ và $s = 2,283$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{2,283}{\sqrt{100}} = 0,4475$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (26,4 - 0,4475; 26,4 + 0,4475) = (25,9525; 26,8475)$$

Vậy thời gian trung bình để sản xuất 1 sản phẩm của phân xưởng với độ tin cậy 95% là khoảng từ 25,9525 đến 26,8475 (giây).

Bài tập 3.2. Quan sát ngẫu nhiên về thời gian cần thiết để sản xuất một chi tiết máy do một máy sản xuất, ta thu được các số liệu cho ở bảng

Thời gian (giây)	Số quan sát	Thời gian (giây)	Số quan sát
122 – 124	8	128 – 130	30
124 – 126	10	130 – 132	14
126 – 128	32	132 – 134	6

Ước lượng thời gian trung bình để sản xuất 1 chi tiết máy với độ tin cậy 95%.

Lời giải

Gọi X (giây) là thời gian để sản xuất 1 chi tiết máy và μ (giây) là thời gian trung bình. Từ bảng dữ liệu trên đề bài, ta tính được $n = 100$, $\overline{x} = 128$ và $s = 2,5186$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{2,5186}{\sqrt{100}} = 0,4936$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (128 - 0,4936; 128 + 0,4936) = (127,5064; 128,4936)$$

Vậy thời gian trung bình để sản xuất 1 chi tiết máy với độ tin cậy 95% là khoảng từ $127,5064$ đến $128,4936$ (giây).

Bài tập 3.3. Kiểm tra năng suất của 1 số hecta lúa chọn ngẫu nhiên ở một vùng, người ta thu được kết quả cho ở bảng

Năng suất (tấn/ha)	Diện tích (ha)	Năng suất (tấn/ha)	Diện tích (ha)
$3,0 - 3,5$	2	$5,0 - 5,5$	16
$3,5 - 4,0$	5	$5,5 - 6,0$	25
$4,0 - 4,5$	8	$6,0 - 6,5$	17
$4,5 - 5,0$	10	$6,5 - 7,0$	13

Những thửa ruộng có năng suất hơn $4,5$ (tấn/ha) gọi là ruộng tốt. Ước lượng năng suất trung bình của ruộng tốt vùng đó với độ tin cậy 95%.

Lời giải

Gọi X (tấn/ha) là năng suất của các ruộng tốt vùng đó và μ (giây) là năng suất trung bình. Từ bảng dữ liệu trên đề bài, ta tính được $n = 81$, $\overline{x} = 5,7932$ và $s = 0,6235$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{0,6235}{\sqrt{81}} = 0,1358$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (5,7932 - 0,1358; 5,7932 + 0,1358) = (5,6574; 5,929)$$

Vậy thời gian trung bình để sản xuất 1 chi tiết máy với độ tin cậy 95% là khoảng từ $5,6574$ đến $5,929$ (tấn/ha).

Bài tập 3.4. Để nghiên cứu nhu cầu của một loại hàng hóa ở một khu vực, người ta tiến hành khảo sát 800 hộ gia đình chọn ngẫu nhiên trong khu vực. Kết quả cho ở bảng dưới

Nhu cầu (kg/tháng)	Số gia đình (n_i)	Nhu cầu (kg/tháng)	Số gia đình (n_i)
$35 - 40$	10	$50 - 55$	230
$40 - 45$	50	$55 - 60$	226
$45 - 50$	140	$60 - 65$	144

Hộ có nhu cầu hơn 45 (kg/tháng) là hộ không nghèo.
a) Ước lượng tỉ lệ hộ không nghèo trong khu vực với độ tin cậy 95%.
b) Ước lượng nhu cầu trung bình mỗi hộ gia đình không nghèo trong khu vực với độ tin cậy 98%.

Lời giải

a) Gọi p là tỉ lệ hộ không nghèo trong khu vực. Từ bảng dữ liệu trên, ta tính được $n = 800$ và tỉ lệ mẫu $f = \dfrac{740}{800} = 0,925$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,925 \cdot 0,075}{800}} = 0,0183$$

Suy ra khoảng tin cậy 95% của p là

$$(f-\varepsilon; f+\varepsilon) = (0,925 - 0,0183; 0,925 + 0,0183) = (0,9067; 0,9433)$$

Vậy tỉ lệ hộ không nghèo trong khu vực với độ tin cậy 95% là khoảng từ 90,67% đến 94,33%.
b) Gọi X (kg/tháng) là nhu cầu về một loại hàng hóa của các hộ không nghèo trong khu vực và μ (kg/tháng) là nhu cầu trung bình. Từ bảng dữ liệu trên đề bài, ta tính được $n = 740$, $\overline{x} = 55,027$ và $s = 5,0471$. Độ tin cậy

$$1-\alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2,3263 \cdot \frac{5,0471}{\sqrt{740}} = 0,4316$$

Vậy khoảng tin cậy 98% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (55,027 - 0,4316; 52,027 + 0,4316) = (54,5954; 55,4586)$$

Vậy nhu cầu trung bình của mỗi hộ gia đình không nghèo trong khu vực với độ tin cậy 98% là khoảng từ 54,5956 đến 55,4584 (kg/tháng).

Bài tập 3.5. Chọn ngẫu nhiên các ngày bán hàng của công ty A, thống kê số lượng hàng hóa bán được mỗi ngày (SHB/N) và số ngày bán được lượng hàng tương ứng, ta có bảng số liệu sau

SHB/N (kg)	Số ngày (n_i)	SHB/N (kg)	Số ngày (n_i)
150 − 200	5	400 − 450	110
200 − 250	19	450 − 500	95
250 − 300	26	500 − 550	60
300 − 350	40	550 − 600	30
350 − 400	65		

Những ngày bán được nhiều hơn 300 (kg) là những ngày đạt doanh thu.
a) Ước lượng tỉ lệ ngày công ty A đạt doanh thu với độ tin cậy 90%.
b) Ước lượng lượng hàng bán trung bình mỗi ngày trong những ngày đạt doanh thu của công ty A với độ tin cậy 98%.

Lời giải

a) Gọi p là tỉ lệ ngày công ty A đạt doanh thu. Từ bảng dữ liệu trên đề bài, ta tính được $n = 450$ và $f = \frac{400}{450} = \frac{8}{9}$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1-\alpha = 90\% = 0,9 \Leftrightarrow \alpha = 0,1 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,05} = 1,6449$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,6449 \cdot \sqrt{\frac{\frac{8}{9}\cdot\frac{1}{9}}{450}} = 0,0244$$

Suy ra khoảng tin cậy 90% của p là

$$(f-\varepsilon; f+\varepsilon) = \left(\frac{8}{9} - 0,0244; \frac{8}{9} + 0,0244\right) = (0,8645; 0,9133)$$

Vậy tỉ lệ ngày công ty A đạt doanh thu với độ tin cậy 90% là khoảng từ 86,45% đến 91,33%.
b) Gọi X (kg) là lượng hàng bán mỗi ngày trong những ngày đạt doanh thu của công ty A và μ là lượng

hàng bán trung bình mỗi ngày. Từ bảng dữ liệu trên đề bài, ta tính được 400, $\overline{x} = 445$ và $s = 69,2784$. Độ tin cậy

$$1 - \alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2,3263 \cdot \frac{69,2784}{\sqrt{400}} = 8,0581$$

Vậy khoảng tin cậy 98% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (445 - 8,0581; 445 + 8,0581) = (436,9419; 453,0581)$$

Vậy lượng hàng bán trung bình mỗi ngày trong những ngày đạt doanh thu của công ty A với độ tin cậy 98% là khoảng từ 436,9464 đến 453,0536 (kg).

Bài tập 3.6. Để định mức thời gian gia công một chi tiết máy trong công ty, người ta theo dõi ngẫu nhiên thời gian gia công 25 chi tiết và thu được bảng số liệu sau đây

Thời gian (phút)	14	16	18	20	22
Số chi tiết	2	6	11	4	2

Với độ tin cậy 0,95, hãy ước lượng thời gian gia công trung bình loại chi tiết trên trong công ty. Cho biết thời gian X (phút) để gia công chi tiết đó trong công ty tuân theo quy luật chuẩn.

Lời giải

Gọi μ (phút) là thời gian gia công trung bình loại chi tiết trên trong công ty. Từ bảng dữ liệu trên đề bài, ta tính được $n = 25$, $\overline{x} = 17,84$ và $s = 2,0752$. Do $n < 30$ nên ta có độ tin cậy

$$1 - \alpha = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow t^{n-1}_{\frac{\alpha}{2}} = t^{24}_{0,025} = 2,0639$$

Suy ra độ chính xác là

$$\varepsilon = t^{n-1}_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2,0639 \cdot \frac{2,0752}{\sqrt{25}} = 0,8566$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (17,84 - 0,8566; 17,84 + 0,8566) = (16,9834; 18,6966)$$

Vậy thời gian gia công trung bình loại chi tiết trên trong công ty với độ tin cậy 0,95 là khoảng từ 16,9834 đến 18,6966 (phút).

Bài tập 3.7. Giả sử điểm trung bình môn Toán của 100 thí sinh thi vào trường Đại học Nha Trang được chọn ngẫu nhiên là 5,0 với độ lệch mẫu là 2,5.
a) Ước lượng điểm trung bình môn Toán của toàn thể thí sinh thi vào trường Đại học Nha Trang với độ tin cậy 95%.
b) Với độ chính xác là 0,25 điểm, hãy xác định độ tin cậy.

Lời giải

a) Gọi X (điểm) là điểm môn Toán của toàn thể thí sinh thi vào trường Đại học Nha Trang và μ là điểm trung bình môn Toán. Từ đề bài, ta có $n = 100$, $\overline{x} = 5,0$ và $s = 2,5$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{2,5}{\sqrt{100}} = 0,49$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (5,0 - 0,49; 5,0 + 0,49) = (4,51; 5,49)$$

Vậy điểm trung bình môn Toán của toàn thể thí sinh thi vào trường Đại học Nha Trang với độ tin cậy 95% là từ $4,51$ đến $5,49$.
b) Với độ chính xác là $0,25$ điểm thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \Leftrightarrow Z_{\frac{\alpha}{2}} = \frac{\varepsilon \cdot \sqrt{n}}{s} = \frac{0,25 \cdot \sqrt{100}}{2,5} = 1$$

Suy ra $\frac{\alpha}{2} = 0,1587$. Vậy độ tin cậy bằng $1 - \alpha = 1 - 2 \cdot 0,1587 = 0,6826 = 68,26\%$.

Bài tập 3.8. Biết tuổi thọ X (giờ) của một loại bóng đèn do xí nghiệp A sản xuất tuân theo quy luật chuẩn với độ lệch chuẩn $\sigma = 100$ (giờ).
a) Chọn ngẫu nhiên 100 bóng đèn để thử nghiệm, ta thấy trung bình tuổi thọ mỗi bóng là 1000 giờ. Hãy ước lượng tuổi thọ trung bình của bóng đèn do xí nghiệp A sản xuất với độ tin cậy 95%.
b) Với độ chính xác 15 giờ, hãy xác định độ tin cậy.
c) Với độ chính xác 25 giờ và độ tin cậy 99% thì cần thử nghiệm bao nhiêu bóng.

Lời giải

a) Gọi μ (giờ) là tuổi thọ trung bình của bóng đèn do xí nghiệp A sản xuất. Từ đề bài, ta có $n = 100$, $\overline{x} = 1000$ và $\sigma = 100$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{100}{\sqrt{100}} = 19,6$$

Suy ra khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (1000 - 19,6; 1000 + 19,6) = (980,4; 1019,6)$$

Vậy tuổi thọ trung bình của bóng đèn do xí nghiệp A sản xuất với độ tin cậy 95% là từ $980,4$ đến $1019,6$ (giờ).
b) Với độ chính xác 15 giờ thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Leftrightarrow Z_{\frac{\alpha}{2}} = \frac{\varepsilon \cdot \sqrt{n}}{\sigma} = \frac{15 \cdot \sqrt{100}}{100} = 1,5$$

Suy ra $\frac{\alpha}{2} = 0,0668$. Vậy độ tin cậy bằng $1 - \alpha = 1 - 2 \cdot 0,0668 = 0,8664 = 86,64\%$.
c) Ta có độ tin cậy

$$1 - \alpha = 99\% = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Với độ chính xác 25 giờ thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Leftrightarrow n = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2 = \left(\frac{2,5758 \cdot 100}{25}\right)^2 = 106,1559$$

Vậy cần thử nghiệm khoảng 107 bóng đèn.

Bài tập 3.9. Chiều dài của một loại sản phẩm A do một máy tự động sản xuất là một biến số ngẫu nhiên tuân theo quy luật phân phối chuẩn vơi độ lệch chuẩn là $3(cm)$. Phải chọn ít nhất bao nhiêu chi tiết để đo, nếu muốn độ dài khoảng tin cậy không vượt quá $0,6$ và độ tin cậy của ước lượng là $0,99$.

Lời giải

Gọi n là số chi tiết cần chọn để đo. Ta có $\sigma = 3$ và độ tin cậy

$$1 - \alpha = 99\% = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 2,5758 \cdot \frac{3}{\sqrt{n}} = \frac{7,7274}{\sqrt{n}}$$

Ta có khoảng tin cậy của ước lượng là $(\overline{x} - \varepsilon; \overline{x} + \varepsilon)$. Theo đề bài thì ta có

$$(\overline{x} + \varepsilon) - (\overline{x} - \varepsilon) \leq 0,6 \Leftrightarrow 2\varepsilon \leq 0,6 \Leftrightarrow \varepsilon = \frac{7,7274}{\sqrt{n}} \leq 0,3 \Leftrightarrow n \geq \left(\frac{7,7274}{0,3}\right)^2 = 663,4746$$

Vậy cần chọn ít nhất 664 chi tiết.

Bài tập 3.10. Để ước lượng trọng lượng trung bình sản phẩm của nhà máy, người ta điều tra 100 sản phẩm được phương sai mẫu cụ thể $s^2 = 16$ và trung bình mẫu cụ thể là 98 kg. Ước lượng với độ chính xác là 1 kg thì độ tin cậy bằng bao nhiêu.

Lời giải

Với độ chính xác 1 kg thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \Leftrightarrow Z_{\frac{\alpha}{2}} = \frac{\varepsilon \cdot \sqrt{n}}{s} = \frac{1 \cdot \sqrt{100}}{\sqrt{16}} = 2,5$$

Suy ra $\frac{\alpha}{2} = 0,0062$. Vậy độ tin cậy bằng $1 - \alpha = 1 - 2 \cdot 0,0062 = 0,9876 = 98,76\%$.

Bài tập 3.11. Nghiên cứu nhu cầu tiêu dùng của một mặt hàng ở một thành phố, người ta điều tra trên 1000 người được chọn ngẫu nhiên và thấy có 400 người có nhu cầu.
a) Hãy ước lượng tỉ lệ người có nhu cầu về mặt hàng đó trong toàn thành phố với độ tin cậy 95%.
b) Muốn độ tin cậy lên đến 98% thì phải điều tra thêm bao nhiêu người.

Lời giải

a) Gọi p là tỉ lệ người có nhu cầu về mặt hàng trên trong toàn thành phố. Từ đề bài, ta có $n = 1000$ và $f = \frac{400}{1000} = 0,4$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,4 \cdot 0,6}{1000}} = 0,0304$$

Suy ra khoảng tin cậy 95% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,4 - 0,0304; 0,4 + 0,0304) = (0,3696; 0,4304)$$

Vậy tỉ lệ người có nhu cầu về mặt hàng trên trong toàn thành phố với độ tin cậy 95% là khoảng từ 36,96% đến 43,04%.
b) Với độ tin cậy 98% thì ta có

$$1 - \alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Do giữ nguyên độ chính xác và tỉ lệ mẫu nên

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n_1}} \Leftrightarrow n_1 = Z^2_{\frac{\alpha}{2}} \cdot \frac{f(1-f)}{\varepsilon^2} = 2,3263^2 \cdot \frac{0,4 \cdot 0,6}{0,0304^2} = 1403,3856$$

Vậy phải điều tra thêm $n_1 - n = 1406 - 1000 = 406$ người.

Bài tập 3.12. Để điều tra số cá trong một hồ, người ta đánh bắt 1000 con cá, đánh dấu rồi thả xuống hồ. Lần sau bắt lại 400 con thì được 40 con có đánh dấu. Với độ tin cậy 95%, hãy
a) Ước lượng tỉ lệ cá được đánh dấu trong hồ.
b) Ước lượng số cá trong hồ.

Lời giải

a) Gọi p là tỉ lệ cá được đánh dấu trong hồ. Từ đề bài, ta tính được $n = 400$ và $f = \dfrac{40}{400} = 0,1$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,1 \cdot 0,9}{400}} = 0,0294$$

Suy ra khoảng tin cậy 95% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,1 - 0,0294; 0,1 + 0,0294) = (0,0706; 0,1294)$$

Vậy tỉ lệ cá được đánh dấu trong hồ với độ tin cậy 95% là khoảng từ $7,06\%$ đến $12,94\%$.

b) Gọi X là số cá trong hồ thì tỉ lệ cá được đánh dấu là $\dfrac{1000}{X}$. Mặt khác, với độ tin cậy 95% thì tỉ lệ cá được đánh dấu trong hồ dao động trong khoảng từ $0,0706$ đến $0,1294$, tức ta có

$$0,0706 \leq \frac{1000}{X} \leq 0,1294 \Leftrightarrow \frac{1000}{0,1294} \leq X \leq \frac{1000}{0,0706} \Leftrightarrow 7727,9753 \leq X \leq 14164,3059$$

Vậy trong hồ có khoảng từ 7728 đến 14164 con cá với độ tin cậy 95%.

Bài tập 3.13. Cơ quan cảnh sát giao thông kiểm tra hệ thống phanh của 500 chiếc xe tải chạy trên đường quốc lộ được chọn ngẫu nhiên. Họ phát hiện ra có 40 chiếc có phanh chưa đảm bảo an toàn. Tìm khoảng tin cậy 98% cho tỉ lệ xe tải chưa có phanh an toàn.

Lời giải

Gọi p là tỉ lệ xe tải chưa có phanh an toàn. Từ đề bài, ta có $n = 500$ và $f = \dfrac{40}{500} = 0,08$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,3263 \cdot \sqrt{\frac{0,08 \cdot 0,92}{500}} = 0,0282$$

Suy ra khoảng tin cậy 98% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,08 - 0,0282; 0,08 + 0,0282) = (0,0518; 0,1082)$$

Vậy tỉ lệ xe tải chưa có phanh an toàn với độ tin cậy 98% là khoảng từ $5,08\%$ đến $10,82\%$.

Bài tập 3.14. Trong đợt vận động bầu cử tổng thống, người ta phỏng vấn ngẫu nhiên 1600 cử tri thì được biết 960 người trong số đó bỏ phiếu cho ứng cử viên A. Với độ tin cậy $0,99$, hãy ước lượng tỉ lệ cử tri bỏ phiếu cho ứng cử viên A.

Lời giải

Gọi p là tỉ lệ cử tri bỏ phiếu cho ứng cử viên A. Từ đề bài, ta có $n = 1600$ và $f = \dfrac{960}{1600} = 0,6$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,5758 \cdot \sqrt{\frac{0,6 \cdot 0,4}{1600}} = 0,0315$$

Suy ra khoảng tin cậy 99% của p là

$$(f-\varepsilon; f+\varepsilon) = (0,6-0,0315; 0,6+0,0315) = (0,5685; 0,6315)$$

Vậy tỉ lệ cử tri bỏ phiếu cho ứng cử viên A với độ tin cậy $0,99$ là khoảng từ $56,85\%$ đến $63,15\%$.

2 Bài tập làm thêm

Bài tập 3.15. Tỉ lệ nảy nầm của một loại hạt giống trong mẫu 400 hạt là 90%.
a) Ước lượng tỉ lệ nảy mầm của loại hạt giống đó với độ tin cậy $0,95$.
b) Muốn độ dài khoảng tin cậy không vượt quá $0,02$ với độ tin cậy như cũ thì phải gieo bao nhiêu hạt.

Lời giải

a) Gọi p là tỉ lệ nảy mầm của loại hạt giống trên. Từ đề bài, ta có $n = 400$ và $f = 0,9$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1-\alpha = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,9 \cdot 0,1}{400}} = 0,0294$$

Suy ra khoảng tin cậy $0,95$ của p là

$$(f-\varepsilon; f+\varepsilon) = (0,9-0,0294; 0,9+0,0294) = (0,8706; 0,9294)$$

Vậy tỉ lệ cá được đánh dấu trong hồ với độ tin cậy $0,95$ là khoảng từ $87,06\%$ đến $92,94\%$.
b) Do giữ nguyên độ tin cậy nên ta có độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n_1}} = 1,96 \cdot \sqrt{\frac{0,9 \cdot 0,1}{n_1}} = \frac{0,588}{\sqrt{n_1}}$$

Ta có khoảng tin cậy của ước lượng là $(f-\varepsilon; f+\varepsilon)$. Theo đề bài thì ta có

$$(f+\varepsilon)-(f-\varepsilon) \leq 0,02 \Leftrightarrow 2\varepsilon \leq 0,02 \Leftrightarrow \varepsilon = \frac{0,588}{\sqrt{n_1}} \leq 0,01 \Leftrightarrow n_1 \geq \left(\frac{0,588}{0,01}\right)^2 = 3457,44$$

Vậy cần gieo ít nhất 3458 hạt.

Bài tập 3.16. Để ước lượng tỉ lệ sản phẩm xấu của một kho đồ hộp, người ta kiểm tra ngẫu nhiên 400 hộp thấy có 80 hộp xấu.
a) Ước lượng tỉ lệ sản phẩm xấu của kho đồ hộp với độ tin cậy 95%.
b) Với mẫu trên, biết độ chính xác 3%, hãy xác định độ tin cậy.

Lời giải

a) Gọi p là tỉ lệ sản phẩm xấu của kho đồ hộp. Từ đề bài, ta có $n = 400$ và $f = \dfrac{80}{400} = 0,2$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1-\alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,2 \cdot 0,8}{400}} = 0,0392$$

Suy ra khoảng tin cậy 95% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,2 - 0,0392; 0,2 + 0,0392) = (0,1608; 0,2392)$$

Vậy tỉ lệ sản phẩm xấu của kho đồ hộp với độ tin cậy 95% là khoảng từ 16,08% đến 23,92%.
b) Với độ chính xác 3% = 0,03 và cỡ mẫu giữ nguyên thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} \Leftrightarrow Z_{\frac{\alpha}{2}} = \frac{\varepsilon \cdot \sqrt{n}}{\sqrt{f(1-f)}} = \frac{0,03 \cdot \sqrt{400}}{\sqrt{0,2 \cdot 0,8}} = 1,5$$

Suy ra $\frac{\alpha}{2} = 0,0668$. Vậy độ tin cậy bằng $1 - \alpha = 1 - 2 \cdot 0,0668 = 0,8664 = 86,64\%$.

Bài tập 3.17. Nghiên cứu nhu cầu tiêu dùng một mặt hàng ở một thành phố, người ta điều tra trên 2400 người của thành phố và biết tỉ lệ người trong thành phố có nhu cầu về mặt hàng này là từ 48% đến 52% với độ tin cậy là 95%.
a) Với độ dài khoảng tin cậy như cũ, muốn tăng độ tin cậy lên 96% thì số người phải điều tra thêm là bao nhiêu.
b) Với độ chính xác như cũ và số người điều tra là 3000 người thì độ tin cậy là bao nhiêu.

Lời giải

a) Gọi f là tỉ lệ mẫu và ε là độ chính xác thì từ giả thiết, ta có

$$\begin{cases} f - \varepsilon = 48\% = 0,48 \\ f + \varepsilon = 52\% = 0,52 \end{cases} \Leftrightarrow \begin{cases} f = 0,5 \\ \varepsilon = 0,02 \end{cases}$$

Độ tin cậy

$$1 - \alpha = 96\% = 0,96 \Leftrightarrow \alpha = 0,04 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,02} = 2,0537$$

Suy ra

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n_1}} \Leftrightarrow n_1 = Z^2_{\frac{\alpha}{2}} \cdot \frac{f(1-f)}{\varepsilon^2} = 2,0537^2 \cdot \frac{0,5 \cdot 0,5}{0,02^2} = 2636,0523$$

Vậy phải điều tra thêm $n_1 - n = 2637 - 2400 = 237$ người.
b) Với độ chính xác như cũ và $n = 3000$ thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} \Leftrightarrow Z_{\frac{\alpha}{2}} = \frac{\varepsilon \cdot \sqrt{n}}{\sqrt{f(1-f)}} = \frac{0,02 \cdot \sqrt{3000}}{\sqrt{0,5 \cdot 0,5}} = 2,19$$

Suy ra $\frac{\alpha}{2} = 0,0143$. Vậy độ tin cậy bằng $1 - \alpha = 1 - 2 \cdot 0,0143 = 0,9714 = 97,14\%$.

Bài tập 3.18. Để xác định tỉ lệ người mắc chứng bướu cổ do thiếu hụt i-ốt ở một khu vực dân cư, cần khám bao nhiêu người nếu muốn cho khoảng tin cậy có
a) Độ chính xác không vượt quá 0,04 với độ tin cậy 95%.
b) Độ chính xác không vượt quá 0,02 với độ tin cậy 98%.

Lời giải

Vì chưa biết tỉ lệ mẫu nên ta đặt $f = 0,5$.
a) Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,5 \cdot (1 - 0,5)}{n}} = \frac{0,98}{\sqrt{n}}$$

Theo đề bài thì ta có

$$\varepsilon \leq 0,04 \Leftrightarrow \frac{0,98}{\sqrt{n}} \leq 0,04 \Leftrightarrow n \geq \left(\frac{0,98}{0,04}\right)^2 = 600,25$$

Vậy cần khám 601 người.
a) Độ tin cậy

$$1-\alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,3263 \cdot \sqrt{\frac{0,5\cdot(1-0,5)}{n}} = \frac{1,16315}{\sqrt{n}}$$

Theo đề bài thì ta có

$$\varepsilon \leq 0,02 \Leftrightarrow \frac{1,16315}{\sqrt{n}} \leq 0,02 \Leftrightarrow n \geq \left(\frac{1,16315}{0,02}\right)^2 = 3382,2948$$

Vậy cần khám 3383 người.

Chương IV

Kiểm định giả thuyết thống kê

1 Bài tập chương 4

Bài tập 4.1. Lấy ngẫu nhiên 25 sản phẩm do một công ty sản xuất ra, ta tính được trọng lượng trung bình của chúng là $995,8(g)$ và phương sai mẫu là $0,144(g^2)$. Giả thiết trọng lượng các sản phẩm là đại lượng ngẫu nhiên $X(g)$ có phân phối chuẩn.
a) Ước lượng trọng lượng trung bình các sản phẩm do công ty sản xuất với độ tin cậy 95%.
b) Có ý kiến cho rằng trọng lượng trung bình sản phẩm công ty là khác $996(g)$. Hãy kết luận ý kiến trên với mức ý nghĩa $\alpha = 5\%$.

Lời giải

a) Gọi $\mu(g)$ là trọng lượng trung bình các sản phẩm do công ty sản xuất. Theo đề bài, ta có $n = 25$, $\overline{x} = 995,8$ và $s^2 = 0,144$. Do $n < 30$ nên ta có độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow \mathrm{t}_{\frac{\alpha}{2}}^{n-1} = \mathrm{t}_{0,025}^{24} = 2,0639$$

Suy ra độ chính xác là

$$\varepsilon = \mathrm{t}_{\frac{\alpha}{2}}^{n-1} \cdot \frac{s}{\sqrt{n}} = 2,0639 \cdot \frac{\sqrt{0,144}}{\sqrt{25}} = 0,1566$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (995,8 - 0,1566; 995,8 + 0,1566) = (995,6436; 995,9566)$$

Vậy trọng lượng trung bình các sản phẩm do công ty sản xuất với độ tin cậy 95% là khoảng từ $995,6436(g)$ đến $995,9566(g)$.
b) Với mức ý nghĩa $\alpha = 5\% = 0,05$, ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \mu = 996 \\ H_1 : \mu \neq 996 \end{cases}$$

Tiêu chuẩn kiểm định

$$\mathrm{T} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{995,8 - 996}{\sqrt{0,144}/\sqrt{25}} = -2,6352$$

Ta sẽ bác bỏ H_0 khi

$$|\mathrm{T}| > \mathrm{t}_{\frac{\alpha}{2}}^{n-1} = t_{0,025}^{24} = 2,0639$$

So sánh thì ta có $|\mathrm{T}| > \mathrm{t}_{\frac{\alpha}{2}}^{n-1}$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1.Vậy đủ cơ sở để khẳng định trọng lượng trung bình các sản phẩm do công ty sản xuất là khác $996(g)$ với mức ý nghĩa 5%.

Bài tập 4.2. Để định mức thời gian gia công một chi tiết máy trong công ty, người ta theo dõi ngẫu nhiên thời gian gia công 225 chi tiết và thu được bảng số liệu sau đây

Thời gian (phút)	14	16	18	20	22
Số chi tiết	12	56	91	54	12

a) Với độ tin cậy 99%, hãy ước lượng thời gian gia công trung bình loại chi tiết trên.
b) Có ý kiến cho rằng thời gian gia công trung bình loại chi tiết trên là dưới 18,5 (phút). Hãy kết luận ý kiến trên với mức ý nghĩa 1%.

Lời giải

a) Gọi X (phút) là thời gian gia công loại chi tiết trên trong công ty và μ (phút) là thời gian gia công trung bình. Từ bảng dữ liệu trên đề bài, ta tính được $n = 225$, $\overline{x} = 17,982$ và $s = 1,9179$. Độ tin cậy

$$1 - \alpha = 99\% = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2,5758 \cdot \frac{1,9179}{\sqrt{225}} = 0,3293$$

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (17,982 - 0,3293; 17,982 + 0,3293) = (17,6527; 18,3113)$$

Vậy thời gian gia công trung bình loại chi tiết trên trong công ty với độ tin cậy 99% là khoảng từ 17,6527 đến 18,3113 (phút).
b) Với mức ý nghĩa $\alpha = 1\% = 0,01$, ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \mu = 18,5 \\ H_1 : \mu < 18,5 \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{17,982 - 18,5}{1,9179/\sqrt{225}} = -4,0513$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_{\alpha} = -Z_{0,01} = -2,3263$$

So sánh thì ta có $Z < -Z_{\alpha}$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1.Vậy đủ cơ sở để khẳng định thời gian gia công trung bình loại chi tiết trên là dưới 18,5 (phút) với mức ý nghĩa 1%.

Bài tập 4.3. Năng suất ngô của một vùng A được báo cáo lên qua 100 điểm thu hoạch được chọn ngẫu nhiên là

Năng suất (tạ/ha)	7	9	11	13	15
Số điểm thu hoạch	2	27	40	30	1

Cho biết năng suất ngô tuân theo quy luật chuẩn.
a) Với độ tin cậy là 95%, hãy ước lượng năng suất ngô trung bình của vùng này.
b) Có ý kiến cho rằng năng suất ngô trong vùng là hơn 10,4 (tạ/ha). Với mức ý nghĩa 2%, hãy kết luận ý kiến trên.

Lời giải

a) Gọi X (tạ/ha) là năng suất ngô của vùng A và μ (tạ/ha) là năng suất ngô trung bình. Từ bảng dữ liệu trên đề bài, ta tính được $n = 100$, $\overline{x} = 11,02$ và $s = 1,6966$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{1,6696}{\sqrt{100}} = 0,3272$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x}-\varepsilon;\overline{x}+\varepsilon)=(11,02-0,3272;11,02+0,3272)=(10,6928;11,3472)$$

Vậy năng suất ngô trung bình của vùng với độ tin cậy 95% là khoảng từ $10,6928$ đến $11,3472$ (tạ/ha).
b) Với mức ý nghĩa $\alpha=2\%=0,02$, ta đặt giả thuyết kiểm định như sau

$$\begin{cases}H_0:\mu=10,4\\H_1:\mu>10,4\end{cases}$$

Tiêu chuẩn kiểm định

$$Z=\frac{\overline{x}-\mu_0}{s/\sqrt{n}}=\frac{11,02-10,4}{1,6696/\sqrt{100}}=3,7135$$

Ta sẽ bác bỏ H_0 khi

$$Z>Z_\alpha=Z_{0,02}=2,0537$$

So sánh thì ta có $Z>Z_\alpha$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1.Vây đủ cơ sở để khẳng định năng suất ngô trung bình của vùng là trên $10,4$ (tạ/ha) với mức ý nghĩa 2%.

Bài tập 4.4. Chọn ngẫu nhiên 100 công nhân của một xí nghiệp thì thấy lương tháng trung bình là $5,8$ (triệu đồng). Giả sử lương X (triệu đồng) của mỗi công nhân tuân theo quy luật chuẩn, với độ lệch chuẩn $\sigma=1,2$ (triệu đồng).
a) Với độ tin cậy 98%, hãy ước lượng mức lương trung bình của công nhân trong toàn xí nghiệp.
b) Có ý kiến cho rằng mức lương trung bình của công nhân là trên $5,5$ (triệu đồng). Với mức ý nghĩa 1%, hãy kết luận ý kiến trên.

Lời giải

a) Gọi $\mu(g)$ là mức lương trung bình của công nhân trong toàn xí nghiệp. Theo đề bài, ta có $n=100$, $\overline{x}=5,8$ và $\sigma=1,2$. Độ tin cậy

$$1-\alpha=98\%=0,98\Leftrightarrow\alpha=0,02\Leftrightarrow Z_{\frac{\alpha}{2}}=Z_{0,01}=2,3263$$

Suy ra độ chính xác là

$$\varepsilon=Z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}=2,3263\cdot\frac{\sqrt{1,2}}{\sqrt{25}}=0,5096$$

Vậy khoảng tin cậy 98% của μ là

$$(\overline{x}-\varepsilon;\overline{x}+\varepsilon)=(5,8-0,5096;5,8+0,5096)=(5,2904;6,3096)$$

Vậy mức lương trung bình của công nhân trong toàn xí nghiệp với độ tin cậy 98% là khoảng từ $5,2904$ đến $6,3096$ (triệu đồng).
b) Với mức ý nghĩa $\alpha=1\%=0,01$, ta đặt giả thuyết kiểm định như sau

$$\begin{cases}H_0:\mu=5,5\\H_1:\mu>5,5\end{cases}$$

Tiêu chuẩn kiểm định

$$Z=\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}=\frac{5,8-5,5}{1,2/\sqrt{100}}=2,5$$

Ta sẽ bác bỏ H_0 khi

$$Z>Z_\alpha=Z_{0,01}=2,3263$$

So sánh thì ta có $Z>Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vậy đủ cơ sở để khẳng định mức lương trung bình của công nhân trong toàn xí nghiệp là trên $5,5$ (triệu đồng) với mức ý nghĩa 1%.

Bài tập 4.5. Trọng lượng trung bình quy định cho các bao hàng do một máy đóng bao sản xuất là 50 (kg). Sau một thời gian hoạt động, người ta nghi ngờ máy có trục trặc. Cho trọng lượng X (kg) mỗi bao hàng là đại lượng ngẫu nhiên phân phối chuẩn.
a) Cân thử 100 bao hàng và tính được $\overline{x} = 49,88$ (kg) và $s = 0,4$ (kg). Với mức ý nghĩa 1%, hãy kết luận trọng lượng trung bình các bao hàng do máy đóng có sai khác quy định không.
b) Cân thử 400 bao hàng và tính được $\overline{x} = 49,93$ (kg) và $s = 0,5$ (kg). Với mức ý nghĩa 1%, hãy kết luận trọng lượng trung bình các bao hàng do máy đóng có thấp hơn quy định không.

Lời giải

Gọi μ (kg) là trọng lượng trung bình các bao hàng do máy đóng bao trên sản xuất.
a) Với mức ý nghĩa $\alpha = 1\% = 0,01$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{49,88 - 50}{0,4/\sqrt{100}} = -3$$

Ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

So sánh thì ta có $|Z| > Z_{\frac{\alpha}{2}}$ nên đủ cơ sở để bác bỏ H_0.Vây đủ cơ sở để khẳng định trọng lượng trung bình các bao hàng do máy đóng sai khác quy định với mức ý nghĩa 1%.
b) Với mức ý nghĩa $\alpha = 1\% = 0,01$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu < 50 \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{49,93 - 50}{0,5/\sqrt{400}} = -2,8$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_{\alpha} = -Z_{0,01} = -2,3263$$

So sánh thì ta có $Z < -Z_{\alpha}$ nên đủ cơ sở để bác bỏ H_0.Vây đủ cơ sở để khẳng định trọng lượng trung bình các bao hàng do máy đóng thấp hơn quy định với mức ý nghĩa 1%.

Bài tập 4.6. Tỉ lệ phế phẩm của một dây chuyền sản xuất là 5%. Sau khi tiến hành một cải tiến kĩ thuật, người ta kiểm tra ngẫu nhiên 2000 sản phẩm thì thấy có 95 phế phẩm. Với mức ý nghĩa 1%, hãy kết luận việc cải tiến kĩ thuật có làm giảm tỉ lệ phế phẩm không.

Lời giải

Gọi p là tỉ lệ phế phẩm mới của dây chuyền sau khi cải tiến. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p_1 = 0,05 \\ H_1 : p_1 < 0,05 \end{cases}$$

Từ đề bài, ta có cỡ mẫu $n = 2000$, tỉ lệ mẫu $f = \dfrac{95}{2000} = 0,0475$ và $p_0 = 0,05$ thỏa $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} = \frac{0,0475 - 0,05}{\sqrt{0,05 \cdot 0,95}/\sqrt{2000}} = -0,513$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_{\alpha} = -Z_{0,01} = -2,3263$$

So sánh thì ta có $Z > -Z_\alpha$ nên không đủ cơ sở để bác bỏ H_0.Vậy không đủ cơ sở để khẳng định việc cải tiến kĩ thuật làm giảm tỉ lệ phế phẩm với mức ý nghĩa 1%.

Bài tập 4.7. Nghiên cứu trọng lượng các trẻ sơ sinh của hai nhóm mẹ nghiện thuốc lá (Nhóm 1) và nhóm mẹ không hút thuốc lá (Nhóm 2), ta có kết quả (kg):

Nhóm 1	Nhóm 2	Nhóm 1	Nhóm 2
3,99	3,18	3,61	2,76
3,79	2,84	3,83	3,60
3,60	2,90	3,31	3,75
3,73	3,27	4,13	3,59
3,21	3,85	3,26	3,63
3,60	3,52	3,54	2,38
4,08	3,23		

Với mức ý nghĩa 5%, có thể nói rằng trẻ sơ sinh của nhóm mẹ nghiện thuốc lá nhẹ cân hơn trẻ sơ sinh của nhóm mẹ không hút thuốc lá được không. Cho biết trọng lượng trẻ sơ sinh của hai nhóm trên tuân theo quy luật chuẩn và có cùng phương sai.

Lời giải

Gọi X, Y (kg) lần lượt là trọng lượng của trẻ sơ sinh ở Nhóm 1 và Nhóm 2. Với mức ý nghĩa $\alpha = 5\% = 0,05$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : E(X) = E(Y) \\ H_1 : E(X) < E(Y) \end{cases}$$

Từ bảng dữ liệu trên đề bài, ta tính được $\overline{x} = 3,6677$; $s_X = 0,2979$; $\overline{y} = 3,2692$; $s_Y = 0,4429$ và $n_1 = n_2 = 13$. Suy ra

$$s^2 = \frac{(n_1-1)s_X^2 + (n_2-1)s_Y^2}{n_1+n_2-2} = \frac{12 \cdot 0,2979^2 + 12 \cdot 0,4429^2}{24} = 0,1424$$

Tiêu chuẩn kiểm định

$$T = \frac{\overline{x}-\overline{y}}{\sqrt{s^2\left(1/n_1 + 1/n_2\right)}} = \frac{3,6677 - 3,2692}{\sqrt{0,1424\left(1/13+1/13\right)}} = 2,6923$$

Ta sẽ bác bỏ H_0 khi

$$T < -t_\alpha^{n_1+n_2-2} = -t_{0,05}^{24} = -1,7109$$

So sánh thì ta có $T > -t_\alpha^{n_1+n_2-2}$ nên không đủ cơ sở để bác bỏ H_0.Vậy không thể khẳng định trẻ sơ sinh của nhóm mẹ nghiện thuốc lá nhẹ cân hơn trẻ sơ sinh của nhóm mẹ không hút thuốc lá với mức ý nghĩa 5%.

Bài tập 4.8. Trọng lượng một loại sản phẩm do hai nhà máy sản xuất ra là đại lượng ngẫu nhiên X, $Y(kg)$ có phân phối chuẩn và có cùng độ lệch chuẩn là $\sigma = 0,5(kg)$. Cân thử 29 sản phẩm của nhà máy thứ nhất ta có $\overline{x} = 50(kg)$ và cân thử 30 sản phẩm của nhà máy thứ hai, ta có $\overline{y} = 50,58(kg)$. Với mức ý nghĩa 5%, hãy kết luận xem trọng lượng trung bình của sản phẩm do hai nhà máy sản xuất ra có như nhau không.

Lời giải

Với mức ý nghĩa $\alpha = 5\% = 0,05$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : E(X) = E(Y) \\ H_1 : E(X) \neq E(Y) \end{cases}$$

Từ đề, ta có $n_1 = 29$, $\overline{x} = 50$, $n_2 = 30$, $\overline{y} = 50,58$ và $D(X) = D(Y) = \sigma^2 = 0,25$. Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x}-\overline{y}}{\sqrt{D(X)/n_1 + D(Y)/n_2}} = \frac{50-50,58}{\sqrt{0,25/29+0,25/30}} = -4,4544$$

Ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

So sánh thì ta có $|Z| > Z_{\frac{\alpha}{2}}$ nên đủ cơ sở để bác bỏ H_0.Vây đủ cơ sở để khẳng định trọng lượng trung bình của sản phẩm do hai nhà máy sản xuất ra là khác nhau với mức ý nghĩa 5%.

Bài tập 4.9. Đường kính một loại chi tiết do hai nhà máy sản xuất ra là hai đại lượng ngẫu nhiên $X, Y(mm)$. Kiểm tra ngẫu nhiên 800 chi tiết do nhà máy thứ nhất sản xuất, ta được $\overline{x} = 100,1(mm)$ và $s_X^2 = 0,001(mm^2)$. Kiểm tra ngẫu nhiên 750 chi tiết do nhà máy thứ hai sản xuất, ta được $\overline{y} = 100,05(mm)$ và $s_Y^2 = 0,0012(mm^2)$. Người ta nói đường kính trung bình của chi tiết nhà máy thứ nhất lớn hơn của chi tiết nhà máy thứ hai. Với mức ý nghĩa 2%, bạn hãy cho ý kiến của mình.

Lời giải

Với mức ý nghĩa $\alpha = 2\% = 0,02$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : E(X) = E(Y) \\ H_1 : E(X) > E(Y) \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \overline{y}}{\sqrt{s_X^2/n_1 + s_Y^2/n_2}} = \frac{100,1 - 100,05}{\sqrt{0,001/800 + 0,0012/750}} = 29,6714$$

Ta sẽ bác bỏ H_0 khi

$$Z > Z_\alpha = Z_{0,02} = 2,0537$$

So sánh thì ta có $Z > Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vây đủ cơ sở để khẳng định đường kính trung bình của chi tiết nhà máy thứ nhất lớn hơn của chi tiết nhà máy thứ hai với mức ý nghĩa 2%.

Bài tập 4.10. Trọng lượng một loại sản phẩm do hai nhà máy sản xuất ra là đại lượng ngẫu nhiên $X, Y(kg)$ có phân phối chuẩn và có cùng độ lệch chuẩn. Cân thử 35 sản phẩm của nhà máy thứ nhất ta có $\overline{x} = 200(kg)$ và $s_X^2 = 1(kg^2)$. Cân thử 36 sản phẩm của nhà máy thứ hai, ta có $\overline{y} = 200,6(kg)$ và $s_Y^2 = 1,01(kg^2)$. Với mức ý nghĩa 2%, hãy kết luận xem trọng lượng trung bình của sản phẩm do nhà máy thứ nhất sản xuất có nhỏ hon của nhà máy thứ hai không.

Lời giải

Với mức ý nghĩa $\alpha = 2\% = 0,02$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : E(X) = E(Y) \\ H_1 : E(X) < E(Y) \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \overline{y}}{\sqrt{s_X^2/n_1 + s_Y^2/n_2}} = \frac{200 - 200,6}{\sqrt{1/35 + 1,01/36}} = -2,5214$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,02} = -2,0537$$

So sánh thì ta có $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vây đủ cơ sở để khẳng định trọng lượng trung bình của sản phẩm do nhà máy thứ nhất sản xuất nhỏ hon của nhà máy thứ hai với mức ý nghĩa 2%.

Bài tập 4.11. Để tìm hiểu hiệu quả của việc giảng dạy một vấn đề nào đó theo phương pháp (PP) cũ và mới, người ta làm một bảng kiểm tra trên 15 sinh viên và có kết quả tính bằng điểm số với điểm tối đa là 100 cho bởi bảng sau đây

Sinh viên	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Điểm (PP cũ)	54	79	91	75	68	43	33	85	22	56	73	63	29	75	87
Điểm (PP mới)	66	85	83	88	93	40	38	91	44	82	59	81	64	83	81

Với mức ý nghĩa 5%, có thể xem việc học theo phương pháp mới có hiệu quả hơn hay không, nghĩa là điểm trung bình của sinh viên học theo phương pháp mới cao hơn điểm trung bình của sinh viên khi học theo phương pháp cũ. Cho biết điểm số của sinh viên tuân theo quy luật chuẩn.

Lời giải

Gọi X, Y lần lượt là điểm số của sinh viên khi học theo phương pháp cũ và mới và $U = Y - X$ là chênh lệch điểm giữa học theo phương pháp mới và cũ. Với mức ý nghĩa $\alpha = 5\% = 0,05$, ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : E(X) = E(Y) \\ H_1 : E(X) < E(Y) \end{cases} \Leftrightarrow \begin{cases} H_0 : \mu = 0 \\ H_1 : \mu > 0 \end{cases} \text{với } \mu = E(U)$$

Khi đó ta có bảng sau

Sinh viên	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Chênh lệch điểm	12	6	−8	13	25	−3	5	6	22	26	−14	18	35	8	−6

Từ đây ta tính được $n = 15$, $\overline{u} = 9,6667$ và $s_U = 13,9164$. Tiêu chuẩn kiểm định

$$\mathrm{T} = \frac{\overline{u} - \mu_0}{s_U/\sqrt{n}} = \frac{9,6667 - 0}{13,9164/\sqrt{15}} = 2,6903$$

Ta sẽ bác bỏ H_0 khi

$$\mathrm{T} > \mathrm{t}_\alpha^{n-1} = \mathrm{t}_{0,05}^{14} = 1,7613$$

So sánh thì ta có $\mathrm{T} > \mathrm{t}_\alpha^{n-1}$ nên đủ cơ sở để bác bỏ H_0.Vậy đủ cơ sở để khẳng định việc học theo phương pháp mới có hiệu quả hơn theo phương pháp cũ với mức ý nghĩa 5%.

Bài tập 4.12. Kiểm tra các sản phẩm chọn ngẫu nhiên do hai nhà máy sản xuất, ta có bảng sau

Nhà máy	Số sản phẩm được kiểm tra	Số phế phẩm
A	1000	20
B	1000	25

Với mức ý nghĩa 5%, có thể kết luận tỉ lệ phế phẩm nhà máy B là lớn hơn không.

Lời giải

Gọi p_1, p_2 lần lượt là tỉ lệ phế phẩm của nhà máy A và nhà máy B. Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases}$$

Theo đề bài, ta có $n_1 = n_2 = 1000$, $f_1 = \dfrac{20}{1000} = 0,02$ và $f_2 = \dfrac{25}{1000} = 0,025$. Suy ra

$$p_0 = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \frac{20 + 25}{1000 + 1000} = 0,0225$$

Tiêu chuẩn kiểm định

$$\mathrm{Z} = \frac{f_1 - f_2}{\sqrt{p_0(1 - p_0)(1/n_1 + 1/n_2)}} = \frac{0,02 - 0,025}{\sqrt{0,0225 \cdot (1 - 0,0225) \cdot (1/1000 + 1/1000)}} = -0,7539$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,05} = -1,6449$$

So sánh thì ta có $Z > -Z_\alpha$ nên không đủ cơ sở để bác bỏ H_0. Vậy ta không đủ cơ sở để khẳng định tỉ lệ phế phẩm nhà máy B lớn hơn nhà máy A với mức ý nghĩa 5%.

Bài tập 4.13. Tỉ lệ phế phẩm trước kia của một dây chuyền sản xuất là 5%. Sau khi tiến hành một thay đổi kỹ thuật, kiểm tra ngẫu nhiên 4000 sản phẩm thì thấy có 164 phế phẩm.
a) Với $\alpha = 1\%$, hãy kết luận việc thay đổi kỹ thuật có giảm tỉ lệ phế phẩm không.
b) Ước lượng tỉ lệ phế phẩm của dây chuyền sau thay đổi kỹ thuật với độ tin cậy 99%.

Lời giải

a) Gọi p là tỉ lệ phế phẩm mới của dây chuyền sau khi thay đổi kỹ thuật. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p = 0,05 \\ H_1 : p < 0,05 \end{cases}$$

Từ đề bài, ta có cỡ mẫu $n = 4000$, tỉ lệ mẫu $f = \dfrac{164}{4000} = 0,041$ và $p_0 = 0,05$ thỏa $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} = \frac{0,041 - 0,05}{\sqrt{0,05 \cdot 0,95}/\sqrt{4000}} = -2,6117$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,01} = -2,3263$$

So sánh thì ta có $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vây ta có thể khẳng định việc thay đổi kĩ thuật làm giảm tỉ lệ phế phẩm với mức ý nghĩa 1%.

b) Ta có cỡ mẫu $n = 4000$ và tỉ lệ mẫu $f = 0,041$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 99\% = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,5758 \cdot \sqrt{\frac{0,041 \cdot 0,959}{4000}} = 0,0081$$

Suy ra khoảng tin cậy 95% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,041 - 0,0081; 0,041 + 0,0081) = (0,0329; 0,0491)$$

Vậy tỉ lệ phế phẩm của dây chuyền sau thay đổi kỹ thuật với độ tin cậy 99% là khoảng từ $3,29\%$ đến $4,91\%$ (phù hợp với câu a).

Bài tập 4.14. Một đảng chính trị dự đoán rằng trong cuộc bầu cử tổng thống sắp tới, ứng viên đảng mình sẽ giành được 45% số phiếu bầu. Chọn ngẫu nhiên 800 cử tri để thăm dò ý kiến cho thấy 320 người nói rằng họ sẽ bỏ phiếu cho ứng viên của đảng đó.
a) Với mức ý nghĩa 1%, nhận định thế nào về dự đoán của đảng đó.
b) Với mức ý nghĩa 1%, tỉ lệ cử tri bỏ phiếu cho ứng viên đó có ít hơn 45% không.
c) Tìm khoảng tin cậy 99% của tỉ lệ cử tri bỏ phiếu cho ứng viên đó.

Lời giải

a) Gọi p là tỉ lệ cử tri bỏ phiếu cho ứng viên đó. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p = 0,45 \\ H_1 : p \neq 0,45 \end{cases}$$

Từ đề bài, ta có cỡ mẫu $n = 800$, tỉ lệ mẫu $f = \dfrac{320}{800} = 0,4$ và $p_0 = 0,45$ thỏa $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} = \frac{0,4 - 0,45}{\sqrt{0,45 \cdot 0,55}/\sqrt{800}} = -2,8427$$

Ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

So sánh thì ta có $|Z| > Z_{\frac{\alpha}{2}}$ nên đủ cơ sở để bác bỏ H_0, nhận H_1.Vậy ta có thể khẳng định tỉ lệ cử tri bỏ phiếu cho ứng viên đó khác 45% với mức ý nghĩa 1%.
b) Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p = 0,45 \\ H_1 : p < 0,45 \end{cases}$$

Từ câu a, ta có tiêu chuẩn kiểm định $Z = -2,8427$. Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,01} = -2,3263$$

So sánh thì ta thấy $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0, nhận H_1.Vậy ta có thể khẳng định tỉ lệ cử tri bỏ phiếu cho ứng viên đó nhỏ hơn 45% với mức ý nghĩa 1%.
c) Ta có cỡ mẫu $n = 800$ và tỉ lệ mẫu $f = 0,4$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 99\% = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,5758 \cdot \sqrt{\frac{0,4 \cdot 0,6}{800}} = 0,0446$$

Suy ra khoảng tin cậy 95% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,4 - 0,0446; 0,4 + 0,0446) = (0,3554; 0,4446)$$

Vậy tỉ lệ cử tri bỏ phiếu cho ứng viên đó với độ tin cậy 99% là khoảng từ $35,54\%$ đến $44,46\%$ (phù hợp với câu a và câu b).

Bài tập 4.15. Tỉ lệ học sinh tốt nghiệp phổ thông năm ngoái của tỉnh A là 88%. Trong kì thi năm nay, trong 100 em được chọn ngẫu nhiên thì có 82 em thi đỗ. Với mức ý nghĩa 5%, có thể kết luận rằng tỉ lệ học sinh thi đỗ năm nay thấp hơn năm ngoái hay không.

Lời giải

Gọi p là tỉ lệ học sinh thi đỗ năm nay. Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p = 0,88 \\ H_1 : p < 0,88 \end{cases}$$

Từ đề bài, ta có cỡ mẫu $n = 100$, tỉ lệ mẫu $f = \dfrac{82}{100} = 0,82$ và $p_0 = 0,88$ thỏa $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} = \frac{0,82 - 0,88}{\sqrt{0,88 \cdot 0,12}/\sqrt{100}} = -1,8464$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,05} = -1,6449$$

So sánh thì ta có $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vậy ta có thể khẳng định tỉ lệ học sinh thi đỗ năm nay thấp hơn năm ngoái với mức ý nghĩa 5%.

Bài tập 4.16. Điều tra năng suất lúa vụ hè thu năm 2017 trên một số hecta lúa chọn ngẫu nhiên ở một vùng, người ta thu được bảng số liệu sau

Năng suất (tấn/ha)	Diện tích (ha)
$3,5-4$	2
$4-4,5$	10
$4,5-5$	10
$5-5,5$	16
$5,5-6$	28
$6-6,5$	32
$6,5-7$	26
$7-7,5$	12

a) Những thửa ruộng có năng suất trên 5 (tấn/ha) là những thửa ruộng tốt. Ước lượng tỉ lệ những thửa ruộng tốt của các hecta lúa trong vùng với độ tin cậy 95%.
b) Ước lượng năng suất vụ hè thu 2017 của các thửa ruộng tốt trong vùng với độ tin cậy 95%.
c) Điều tra vụ hè thu 2016 thì thấy năng suất lúa tốt ở vùng này là 6 (tấn/ha). Đến vụ hè thu năm 2017 thì người ta áp dụng một biện pháp kỹ thuật mới. Biện pháp kỹ thuật mới này có làm tăng năng suất đối với các diện tích lúa tốt, với mức ý nghĩa 1%.

Lời giải

a) Gọi p là tỉ lệ thửa ruộng tốt của các hecta lúa trong vùng. Từ bảng dữ liệu trên đề bài, ta tính được $n = 136$ và $f = \dfrac{114}{136} = 0,8382$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1-\alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,8382 \cdot 0,1618}{136}} = 0,0619$$

Suy ra khoảng tin cậy 90% của p là

$$(f-\varepsilon; f+\varepsilon) = (0,8382 - 0,0619; 0,8382 + 0,0619) = (0,7763; 0,9001)$$

Vậy tỉ lệ tỉ lệ những thửa ruộng tốt của các hecta lúa trong vùng với độ tin cậy 95% là khoảng từ $77,63\%$ đến $90,01\%$.

b) Gọi X (tấn/ha) là năng suất vụ hè thu 2017 của các thửa ruộng tốt trong vùng và μ là năng suất trung bình. Từ bảng dữ liệu trên đề bài, ta tính được 114, $\overline{x} = 6,2061$ và $s = 0,6044$. Độ tin cậy

$$1-\alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{0,6044}{\sqrt{114}} = 0,1109$$

Vậy khoảng tin cậy 95% của μ là

$$(\overline{x}-\varepsilon; \overline{x}+\varepsilon) = (6,2061 - 0,1109; 6,2061 + 0,1109) = (6,0952; 6,317)$$

Vậy năng suất vụ hè thu 2017 của các thửa ruộng tốt trong vùng với độ tin cậy 95% là khoảng từ $6,0952$ đến $6,317$ (tấn/ha).

c) Với mức ý nghĩa $\alpha = 1\% = 0,01$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : \mu = 6 \\ H_1 : \mu > 6 \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{6,2061 - 6}{0,6044/\sqrt{114}} = 2,8305$$

Ta sẽ bác bỏ H_0 khi

$$Z > Z_\alpha = Z_{0,01} = 2,3263$$

So sánh thì ta có $Z > Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vây đủ cơ sở để khẳng định biện pháp kỹ thuật mới này làm tăng năng suất đối với các diện tích lúa tốt với mức ý nghĩa 1%.

Bài tập 4.17. Kiểm tra chất lượng của hai lô sản phẩm, người ta thấy trong lô thứ nhất có 50 phế phẩm trên tổng số 500 sản phẩm kiểm tra và lô thứ hai có 60 phế phẩm trên tổng số 400 sản phẩm kiểm tra. Với mức ý nghĩa 5%, có thể xem lô hàng thứ nhất chất lượng tốt hơn lô hàng thứ hai không.

Lời giải

Gọi p_1, p_2 lần lượt là tỉ lệ phế phẩm của lô thứ nhất và lô thứ hai. Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 < p_2 \end{cases}$$

Theo đề bài, ta có $n_1 = 500$, $n_2 = 400$ và tỉ lệ phế phẩm ở lô thứ nhất và lô thứ hai lần lượt là

$$f_1 = \frac{50}{500} = 0,1; f_2 = \frac{60}{400} = 0,15$$

Suy ra

$$p_0 = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \frac{50 + 60}{500 + 400} = \frac{11}{90}$$

Tiêu chuẩn kiểm định

$$Z = \frac{f_1 - f_2}{\sqrt{p_0\,(1 - p_0)\,(1/n_1 + 1/n_2)}} = \frac{0,1 - 0,15}{\sqrt{\frac{11}{90} \cdot \left(1 - \frac{11}{90}\right) \cdot (1/500 + 1/400)}} = -2,2756$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,05} = -1,6449$$

So sánh thì ta có $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1. Vậy ta có thể xem lô hàng thứ nhất chất lượng tốt hơn lô hàng thứ hai với mức ý nghĩa 5%.

Bài tập 4.18. Nghiên cứu tình trạng hôn nhân trước ngày cưới của 800 cặp vợ chồng được chọn ngẫu nhiên ở nước ta được bảng số liệu

Vợ / Chồng	Chưa kết hôn lần nào	Ly hôn	Góa
Chưa kết hôn lần nào	150	124	66
Ly hôn	138	108	54
Góa	52	48	60

Với mức ý nghĩa 1%, có thể coi tình trạng hôn nhân trước ngày cưới của vợ chồng là độc lập không.

Lời giải

Gọi A, B lần lượt là tình trạng hôn nhân trước ngày cưới của vợ và chồng. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij} = \dfrac{n_i \cdot m_j}{n}$ và viết vào bảng trên như sau

Chồng \ Vợ	Chưa kết hôn lần nào	Ly hôn	Góa	n_i
Chưa kết hôn lần nào	150 (144, 5)	124 (119)	66 (76, 5)	340
Ly hôn	138 (127, 5)	108 (105)	54 (67, 5)	300
Góa	52 (68)	48 (56)	60 (36)	160
m_j	340	280	180	$n = 800$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,3}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=3}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{150^2}{144,5} + \frac{124^2}{119} + \cdots + \frac{48^2}{56} + \frac{60^2}{36}\right) - 800 = 26,4186$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,01}\left((3-1)(3-1)\right) = \mathcal{X}^2_{0,01}(4) = 13,2767$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 1%, ta có thể khẳng định tình trạng hôn nhân trước ngày cưới của vợ chồng có liên quan với nhau.

Bài tập 4.19. Nghiên cứu về màu tóc và giới tính của 500 người được chọn ngẫu nhiên ở châu Âu, ta có số liệu sau

Màu tóc \ Giới tính	Nam	Nữ
Đen	70	30
Hung	75	40
Nâu	80	55
Vàng	95	55

Với mức ý nghĩa 5%, có thể coi giữa màu tóc và giới tính có độc lập nhau không.

Lời giải

Gọi A là màu tóc, B là giới tính. Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Màu tóc \ Giới tính	Nam	Nữ	n_i
Đen	70 (64)	30 (36)	100
Hung	75 (73, 6)	40 (41, 4)	115
Nâu	80 (86, 4)	55 (48, 6)	135
Vàng	95 (96)	55 (54)	150
m_j	320	180	$n = 500$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,4}$ và $\forall j = \overline{1,2}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=4}\sum_{j=1}^{k=2} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{70^2}{64} + \frac{30^2}{36} + \frac{75^2}{73,6} + \frac{40^2}{41,4} + \frac{80^2}{86,4} + \frac{55^2}{48,6} + \frac{95^2}{96} + \frac{55^2}{54}\right) - 500 = 2,9823$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,05}\left((4-1)(2-1)\right) = \mathcal{X}^2_{0,05}(3) = 7,8147$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha((h-1)(k-1))$ nên không đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 5%, ta có thể khẳng định màu tóc và giới tính độc lập với nhau.

Bài tập 4.20. Phỏng vấn ngẫu nhiên 200 người thuộc các vùng địa lý ở nước ta về tiêu dùng một loại sản phẩm nào đó, ta thu được kết quả sau

Tiêu dùng \ Vùng địa lý	Thành thị	Nông thôn	Miền núi
Có tiêu dùng	26	50	24
Không tiêu dùng	47	45	8

Với mức ý nghĩa 1%, có thể coi yếu tố địa lý và việc tiêu dùng loại sản phẩm nói trên độc lập với nhau. (Với mức ý nghĩa 1%, có thể nói tỉ lệ tiêu dùng sản phẩm đó ở thành thị, nông thôn và miền núi là như nhau không.)

Lời giải

Gọi A là yếu tố địa lý, B là việc tiêu dùng loại sản phẩm trên. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Tiêu dùng \ Vùng địa lý	Thành thị	Nông thôn	Miền núi	n_i
Có tiêu dùng	26 (36,5)	50 (47,5)	24 (16)	100
Không tiêu dùng	47 (36,5)	45 (47,5)	8 (16)	100
m_j	73	95	32	$n = 200$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{26^2}{36,5} + \frac{50^2}{47,5} + \frac{24^2}{16} + \frac{47^2}{36,5} + \frac{45^2}{47,5} + \frac{8^2}{16}\right) - 200 = 14,3042$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha((h-1)(k-1)) = \mathcal{X}^2_{0,01}((2-1)(3-1)) = \mathcal{X}^2_{0,01}(2) = 9,2103$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}^2_\alpha((h-1)(k-1))$ nên đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 1%, ta có thể khẳng định yếu tố địa lý và việc tiêu dùng loại sản phẩm nói trên có liên quan với nhau.

Bài tập 4.21. Một công ty xuất khẩu gạo nói gạo của họ ở các kho 1, 2, 3 là cùng chất lượng hạt, tức là chất lượng hạt và các kho là độc lập với nhau. Lấy mẫu cụ thể các hạt gạo ở các kho, ta có số liệu

Chất lượng hạt \ Kho	Kho 1	Kho 2	Kho 3
Còn nguyên hạt	600	460	500
Còn hơn 2/3 hạt	170	100	70
Còn dưới 2/3 hạt	30	40	30

Với mức ý nghĩa 1%, bạn hãy cho ý kiến.

Lời giải

Gọi A là chất lượng hạt, B là kho hàng. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Chất lượng hạt \ Kho	Kho 1	Kho 2	Kho 3	n_i
Còn nguyên hạt	600 (624)	460 (468)	500 (468)	1560
Còn hơn 2/3 hạt	170 (136)	100 (102)	70 (102)	340
Còn dưới 2/3 hạt	30 (40)	40 (30)	30 (30)	100
m_j	800	600	600	$n = 2000$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,3}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=3}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{600^2}{624} + \frac{460^2}{468} + \frac{500^2}{468} + \frac{170^2}{136} + \frac{100^2}{102} + \frac{70^2}{102} + \frac{30^2}{40} + \frac{40^2}{30} + \frac{30^2}{30}\right) - 2000$$
$$= 27,6596$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}_{\alpha}^2\left((h-1)(k-1)\right) = \mathcal{X}_{0,01}^2\left((3-1)(3-1)\right) = \mathcal{X}_{0,01}^2\left(4\right) = 13,2767$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}_{\alpha}^2\left((h-1)(k-1)\right)$ nên đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 1%, ta có thể khẳng định chất lượng hạt và các kho là có liên quan với nhau.

Bài tập 4.22. Một công ty có bốn loại kho chứa các sản phẩm cùng loại. Giám đốc công ty đó nói chất lượng sản phẩm và kho hàng là độc lập (tỉ lệ thành phần sản phẩm loại 1, 2 và 3 trong các kho hàng là như nhau). Kiểm tra ngẫu nhiên 2000 sản phẩm ở các kho 1, 2, 3 và 4 thì ta có bảng số liệu sau

Chất lượng \ Kho	Kho 1	Kho 2	Kho 3	Kho 4
Loại 1	110	120	100	115
Loại 2	210	210	220	225
Loại 3	160	190	180	160

Với mức ý nghĩa 5%, hãy kết luận về ý kiến trên.

Lời giải

Ta gọi A là loại hàng và B là kho hàng. Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Chất lượng \ Kho	Kho 1 (B_1)	Kho 2 ((B_2))	Kho 3 (B_3)	Kho 4 (B_4)	n_i
Loại 1 (A_1)	110 (106,8)	120 (115,7)	100 (111,25)	115 (111,25)	445
Loại 2 (A_2)	210 (207,6)	210 (224,9)	220 (216,25)	225 (216,25)	865
Loại 3 (A_3)	160 (165,6)	190 (179,4)	180 (172,5)	160 (172,5)	690
m_j	480	520	500	500	$n = 2000$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,3}$ và $\forall j = \overline{1,4}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=3}\sum_{j=1}^{k=4} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{110^2}{106,8} + \frac{120^2}{115,7} + \cdots + \frac{180^2}{172,5} + \frac{160^2}{172,5}\right) - 2000 = 5,0013$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,05}\left((3-1)(4-1)\right) = \mathcal{X}^2_{0,05}(6) = 12,5916$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên không đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 5%, ta có thể khẳng định chất lượng sản phẩm và kho hàng là độc lập với nhau.

Chương V

Tương quan và hồi quy

1 Bài tập chương 5

Bài tập 5.1. Trên một sàn giao dịch chứng khoán, có hai loại cổ phiếu KHB và ACB được bán với lãi suất tương ứng của chúng là hai đại lượng ngẫu nhiên X, Y. Giả sử (X, Y) có bảng phân phối sản xuất như sau

$X \backslash Y$	-2	0	5	10
0	0	$0,05$	$0,05$	$0,1$
4	$0,05$	$0,1$	$0,25$	$0,15$
6	$0,1$	$0,05$	$0,1$	0

a) Để đạt lãi suất kỳ vọng cao nhất thì nên đầu tư vào cả hai loại cổ phiếu theo tỉ lệ nào.
b) Để hạn chế rủi ro lãi suất đến mức thấp nhất thì đầu tư hai loại cổ phiếu theo tỉ lệ nào.

Lời giải

a) Ta có
$$E[X] = 4 \cdot 0,55 + 6 \cdot 0,25 = 3,7(\%)$$
và
$$E[Y] = (-2) \cdot 0,15 + 5 \cdot 0,4 + 10 \cdot 0,25 = 4,2(\%)$$
Giả sử đầu tư vào cổ phiếu KHB và ACB theo tỉ lệ p và $1-p$ với $0 \le p \le 1$. Gọi $Z(\%)$ là lãi suất thu được khi đầu tư vào cổ phiếu KHB và ACB theo các tỉ lệ trên thì ta có
$$Z = pX + (1-p)Y$$
Lãi suất kỳ vọng của Z là
$$E[Z] = pE[X] + (1-p)E[Y] = 3,7p + (1-p)4,2 = 4,2 - 0,5p \le 4,2, \forall p \in [0,1]$$
Dấu bằng xảy ra khi $p = 0$. Vậy nên đầu tư hoàn toàn vào cổ phiếu ACB thì đạt lãi suất kỳ vọng cao nhất.
b) Ta có
$$D[X] = E\left[X^2\right] - E^2[X] = \left(4^2 \cdot 0,55 + 6^2 \cdot 0,25\right) - (3,7)^2 = 4,11$$
và
$$D[Y] = E\left[Y^2\right] - E^2[Y] = \left[(-2)^2 \cdot 0,15 + 5^2 \cdot 0,4 + 10^2 \cdot 0,25\right] - (4,2)^2 = 17,96$$
Ta lại có
$$\begin{aligned}\operatorname{Cov}(X,Y) &= E[XY] - E[X]E[Y] \\ &= [(-2) \cdot 4 \cdot 0,05 + (-2) \cdot 6 \cdot 0,1 + 5 \cdot 4 \cdot 0,25 + 5 \cdot 6 \cdot 0,1 + 10 \cdot 4 \cdot 0,15] - 3,7 \cdot 4,2 \\ &= -3,14\end{aligned}$$

Độ rủi ro của Z là

$$D(Z) = p^2 D[X] + (1-p)^2 D[Y] + 2p(1-p)\,\mathrm{Cov}(X,Y) = 28,35p^2 - 42,2p + 17,96$$

$D(Z)$ đạt giá trị nhỏ nhất khi

$$p = \frac{422}{567} \approx 0,7443$$

Vậy đầu tư vào cổ phiếu KHB với tỉ lệ 74,43% và đầu tư vào cổ phiếu ACB với tỉ lệ 25,57% thì rủi ro về lãi suất đạt mức thấp nhất.

Bài tập 5.2. Điều tra thu nhập hằng năm (triệu đồng/năm) của các cặp vợ chồng đang làm việc tại một nhà máy, thu được kết quả sau

X \ Y	10	20	30	40
10	0,2	0,04	0,01	0
20	0,1	0,36	0,09	0
30	0	0,05	0,1	0
40	0	0	0	0,05

a) Tìm phân phối xác suất thu nhập của chồng (X) và thu nhập của vợ (Y).
b) Tìm phân phối thu nhập của những người vợ có chồng thu nhập 20 triệu/năm.
c) Tính thu nhập trung bình của các bà vợ có chồng thu nhập ở mức 20 triệu/năm.
d) Thu nhập của chồng và vợ có phụ thuộc nhau không và nếu có thì phụ thuộc như thế nào.

Lời giải

a) Bảng phân phối xác suất của X và Y

X_i	10	20	30	40
$P(X = X_i)$	0,25	0,55	0,15	0,05

Y_i	10	20	30	40
$P(Y = Y_i)$	0,3	0,45	0,2	0,05

b) Ta có bảng phân phối xác suất của Y với điều kiện $X = 20$

Y_i	10	20	30	40
$P(Y = Y_i \mid X = 20)$	2/11	36/55	9/55	0

c) Từ câu b, ta có

$$E(Y \mid X = 20) = 10 \cdot \frac{2}{11} + 20 \cdot \frac{36}{55} + 30 \cdot \frac{9}{55} = \frac{218}{11} \approx 19,8182$$

Vậy thu nhập trung bình của các bà vợ có chồng thu nhập 20 triệu/năm là khoảng 19,8182 triệu/năm.
d) Ta có

$$\begin{cases} E[X] = 20 \\ E[Y] = 20 \\ E[XY] = 449 \end{cases} \Rightarrow \mathrm{Cov}(X,Y) = E[XY] - E[X] \cdot E[Y] = 49 \neq 0$$

Vậy X và Y phụ thuộc nhau.

Bài tập 5.3. Một kiện hàng có 10 sản phẩm, trong đó có 4 sản phẩm loại A. Một máy sản xuất sản phẩm với xác suất sản xuất ra sản phẩm loại A là 0,2. Lấy ngẫu nhiên không hoàn lại từ kiện hàng ra 2 sản phẩm và cho máy sản xuất ra 2 sản phẩm. Gọi X là số sản phẩm loại A trong 4 sản phẩm đó. Lập bảng phân phối xác suất của X và tính $E[X]$, $D[X]$.

Lời giải

Gọi Y là số sản phẩm loại A trong 2 sản phẩm lấy ra từ kiện hàng và Z là số sản phẩm loại A trong 2 sản phẩm được sản xuất từ máy thì ta có bảng phân phối xác suất của Y và Z như sau

Y_i	0	1	2
$P(Y = Y_i)$	1/3	8/15	2/15

Z_i	0	1	2
$P(Z = Z_i)$	0,64	0,32	0,04

Mặt khác, từ cách đặt thì ta có $X = Y + Z$. Suy ra bảng phân phối xác suất của X là

X	0	1	2	3	4
$P(X = X_i)$	$16/75$	$0,448$	$101/375$	$0,064$	$2/375$

Từ đây ta tính được

$$E[X] = 1 \cdot 0,448 + 2 \cdot \frac{101}{375} + 3 \cdot 0,064 + 4 \cdot \frac{2}{375} = 1,2$$

và

$$D[X] = E\left[X^2\right] - E^2[X] = \left(1^2 \cdot 0,448 + 2^2 \cdot \frac{101}{375} + 3^2 \cdot 0,064 + 4^2 \cdot \frac{2}{375}\right) - (1,2)^2 = \frac{56}{75}$$

Bài tập 5.4. Tiến hành quan sát về hai chỉ tiêu X và Y trên cùng một tổng thể, ta thu được mẫu số liệu

X	0,25	0,37	0,44	0,55	0,6	0,62	0,68	0,7	0,73
Y	2,57	2,31	2,12	1,92	1,75	1,71	1,6	1,51	1,5
X	0,75	0,92	0,84	0,87	0,88	0,9	0,95	1	
Y	1,41	1,33	1,31	1,25	1,2	1,19	1,15	1	

a) Tính $\overline{x}, \overline{y}, \overline{xy}, s_X^2, s_Y^2, r$.
b) Viết phương trình đường hồi quy tuyến tính của Y theo X.

Lời giải

a) Từ bảng dữ liệu, ta tính được

$$n = 17; \sum_{i=1}^{n} x_i = 12,05; \sum_{i=1}^{n} x_i^2 = 9,2835; \sum_{i=1}^{n} y_i = 26,83; \sum_{i=1}^{n} y_i^2 = 45,4127; \sum_{i=1}^{n} x_i y_i = 17,5247$$

Suy ra

$$\begin{cases} \overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = 0,7088 \\ \overline{y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i = 1,5782 \\ \overline{xy} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i y_i = 1,0309 \end{cases} \Rightarrow \begin{cases} s_X = \sqrt{\dfrac{\sum_{i=1}^{n} x_i^2 - n \cdot (\overline{x})^2}{n-1}} = 0,2155 \\ s_Y = \sqrt{\dfrac{\sum_{i=1}^{n} y_i^2 - n \cdot (\overline{y})^2}{n-1}} = 0,4381 \end{cases}$$

Từ đây ta tính được hệ số tương quan mẫu là

$$r = \frac{n}{n-1}\left(\frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_X \cdot s_Y}\right) = \frac{17}{16} \cdot \frac{1,0309 - 0,7088 \cdot 1,5782}{0,2155 \cdot 0,4381} \approx -0,9873$$

b) Phương trình đường hồi quy tuyến tính của Y theo X là

$$Y - \overline{y} = r\frac{s_Y}{s_X}(X - \overline{x}) \Leftrightarrow Y - 1,5782 = (-0,9873) \cdot \frac{0,4381}{0,2155}(X - 0,7088)$$

$$\Leftrightarrow Y = -2,0071X + 3,0009$$

Bài tập 5.5. Điều tra thu nhập của 10 cặp vợ chồng (triệu đồng/năm) thu được kết quả sau

X (thu nhập của chồng)	20	30	30	20	20	30	40	30	40	40
Y (thu nhập của vợ)	15	35	25	25	25	15	25	25	35	25

a) Lập bảng phân phối xác suất đồng thời của (X, Y).
b) Lập bảng phân phối xác suất của biến X. Tính $E[X]$ và $D[X]$.
c) Lập bảng phân phối xác suất của biến Y. Tính $E[Y]$ và $D[Y]$.
d) Giả sử thu nhập sau thuế W của các cặp vợ chồng được xác định bởi biểu thức $W = 0,6X + 0,8Y$. Tính $E[W]$ và $D[W]$.

Lời giải

a) Bảng phân phối xác suất đồng thời của (X, Y)

Y \ X	20	30	40
15	$0,1$	$0,1$	0
25	$0,2$	$0,2$	$0,2$
35	0	$0,1$	$0,1$

b) Bảng phân phối xác suất của X

X	20	30	40
P	$0,3$	$0,4$	$0,3$

Từ đây ta có $E[X] = 30$ và $D[X] = 60$.
c) Bảng phân phối xác suất của Y

Y	15	25	35
P	$0,2$	$0,6$	$0,2$

Từ đây ta có $E[Y] = 25$ và $D[X] = 40$.
d) Từ cách đặt thì ta có

$$E[W] = 0,6E[X] + 0,8E[Y] = 0,6 \cdot 30 + 0,8 \cdot 25 = 38$$

và

$$D[W] = (0,6)^2 D[X] + (0,8)^2 D[Y] = 0,36 \cdot 60 + 0,64 \cdot 40 = 47,2$$

Bài tập 5.6. Nghiên cứu về thu nhập và tỉ lệ thu nhập chi cho giáo dục của các hộ gia đình ở một vùng, điều tra 400 hộ thì ta thu được số liệu sau

Y \ X	10	20	30	40	50
$150-250$	40	20	10		
$250-350$		60	40	10	
$350-450$		20	70	60	
$450-550$			10	20	40

Trong đó $X(\%)$ là tỉ lệ thu nhập chi cho giáo dục, Y (usd/tháng) là thu nhập bình quân của một người trong gia đình.
a) Tìm hệ số tương quan mẫu giữa X và Y.
b) Lập phương trình đường hồi quy tuyến tính của Y theo X.

Lời giải

a) Ta viết lại bảng trên như sau

Y \ X	10	20	30	40	50	n_i
$150-250$	40	20	10			70
$250-350$		60	40	10		110
$350-450$		20	70	60		150
$450-550$			10	20	40	70
m_j	40	100	130	90	40	$n = 400$

Từ đây ta tính được

$$\begin{cases} \sum\limits_{j=1}^{5} x_j m_j = 11.900; \sum\limits_{j=1}^{5} x_j^2 m_j = 405.000 \\ \sum\limits_{i=1}^{4} y_i n_i = 142.000; \sum\limits_{i=1}^{4} y_i^2 n_i = 54.200.000 \\ \sum\limits_{i=1}^{4}\sum\limits_{j=1}^{5} x_j y_i n_{ij} = 4.570.000 \end{cases} \Rightarrow \begin{cases} \overline{x} = \dfrac{1}{n}\sum\limits_{j=1}^{5} x_j m_j = 29,75 \\ \overline{y} = \dfrac{1}{n}\sum\limits_{i=1}^{4} y_i n_i = 355 \\ \overline{xy} = \dfrac{1}{n}\sum\limits_{i=1}^{4}\sum\limits_{j=1}^{5} x_j y_i n_{ij} = 11.425 \end{cases}$$

$$\Rightarrow \begin{cases} s_X = \sqrt{\dfrac{\sum\limits_{j=1}^{5} x_j^2 m_j - n \cdot (\overline{x})^2}{n-1}} = 11,303 \\ s_Y = \sqrt{\dfrac{\sum\limits_{i=1}^{4} y_i^2 n_i - n \cdot (\overline{y})^2}{n-1}} = 97,4615 \end{cases}$$

Suy ra hệ số tương quan mẫu là

$$r = \frac{n}{n-1}\left(\frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_X \cdot s_Y}\right) = \frac{400}{399} \cdot \frac{11.425 - 29,75 \cdot 355}{11,303 \cdot 97,4615} \approx 0,786$$

b) Phương trình đường hồi quy tuyến tính của Y theo X là

$$Y - \overline{y} = r\frac{s_Y}{s_X}(X - \overline{x}) \Leftrightarrow Y - 355 = 0,786 \cdot \frac{97,4615}{11,303}(X - 29,75)$$

$$\Leftrightarrow Y = 6,7774X + 153,373$$

Bài tập 5.7. Nghiên cứu về X (ngàn đồng) là thu nhập bình quân tháng mỗi người của hộ gia đình và $Y(\%)$ là tỉ lệ thu nhập chi cho ăn uống của hộ gia đình trong một vùng, điều tra 400 hộ gia đình ở vùng đó thì ta có bảng số liệu thực nghiệm sau đây

X \ Y	10	20	30	40	50
50 − 150			30	30	10
150 − 250		20	80	40	
250 − 350		40	60	20	
350 − 450	10	40	20		

a) Tìm khoảng tin cậy cho tỉ lệ thu nhập trung bình chi cho ăn uống của một gia đình với mức tin cậy 95%.
b) Những hộ gia đình có *thu nhập cao* trong vùng là những hộ gia đình có thu nhập bình quân mỗi người/tháng là trên 350.000 đồng. Nếu nói rằng tỉ lệ hộ gia đình có thu nhập cao trong toàn vùng là 20% với mức ý nghĩa 5% thì bạn có chấp nhận được không.
c) Tìm hệ số tương quan mẫu giữa X và Y. Lập phương trình đường hồi quy tuyến tính của Y theo X.

Lời giải

a) Gọi μ_Y là tỉ lệ thu nhập trung bình chi cho ăn uống của một gia đình. Từ bảng dữ liệu trên, ta tính được $n = 400$, $\overline{y} = 29,75$ và $s_Y = 8,2223$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s_Y}{\sqrt{n}} = 1,96 \cdot \frac{8,2223}{\sqrt{400}} = 0,8058$$

Vậy khoảng tin cậy 95% của μ_Y là

$$(\overline{y} - \varepsilon; \overline{y} + \varepsilon) = (29,75 - 0,8058; 29,75 + 0,8058) = (28,9442; 30,5558)$$

Vậy tỉ lệ thu nhập trung bình chi cho ăn uống của một gia đình với độ tin cậy 95% là khoảng từ 28,94% đến 30,56%.
b) Gọi p là tỉ lệ hộ gia đình có thu nhập cao trong toàn vùng. Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p = 0,2 \\ H_1 : p \neq 0,2 \end{cases}$$

Từ đề bài, ta có $n = 400$, $f = \dfrac{70}{400} = 0,175$ và $p_0 = 0,2$ thỏa $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0,175 - 0,2}{\sqrt{0,2 \cdot 0,8}}\sqrt{400} = -1,25$$

Ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

So sánh thì ta có $|Z| < Z_{\frac{\alpha}{2}}$ nên không đủ cơ sở để bác bỏ H_0.Vây ta không đủ cơ sở để khẳng định tỉ lệ hộ gia đình có thu nhập cao trong toàn vùng là 20% với mức ý nghĩa 5%.

c) Từ bảng dữ liệu, ta tính được $\overline{x} = 247,5$; $s_X = 97,5579$ và $\overline{xy} = 6900$. Suy ra hệ số tương quan mẫu là

$$r = \frac{n}{n-1}\left(\frac{\overline{xy} - \overline{x}\cdot\overline{y}}{s_X \cdot s_Y}\right) = \frac{400}{399} \cdot \frac{6900 - 29,75 \cdot 247,5}{97,5579 \cdot 8,2223} \approx -0,5788$$

Phương trình đường hồi quy tuyến tính của Y theo X là

$$Y - \overline{y} = r\frac{s_Y}{s_X}(X - \overline{x}) \Leftrightarrow Y - 29,75 = (-0,5788) \cdot \frac{8,2223}{97,5579}(X - 247,5)$$
$$\Leftrightarrow Y = -0,0488X + 41,8235$$

Bài tập 5.8. Một loại sản phẩm nào đó được đánh giá chất lượng qua hai chỉ tiêu X, Y nào đó. Kiểm tra một số sản phẩm về hai chỉ tiêu ở trên thì ta có kết quả sau đây

X \ Y	0 − 4	4 − 8	8 − 12	12 − 16	16 − 20
115 − 125	4				
125 − 135	6	8	10		
135 − 145		12	15	2	
145 − 155	10	10	14	7	5
155 − 165				4	3

a) Yêu cầu đạt tiêu chuẩn của chỉ tiêu X là 145. Có người cho rằng chỉ tiêu X trung bình là nhỏ hơn yêu cầu. Với mức ý nghĩa 5%, bạn hãy cho biết ý kiến của mình.

b) Sản phẩm có chỉ tiêu Y lớn hơn 12 là sản phẩm loại I. Hãy ước lượng trung bình chỉ tiêu Y của sản phẩm loại I với mức tin cậy 90%. Cho biết chỉ tiêu Y của sản phẩm loại I tuân theo quy luật chuẩn.

c) Tìm hệ số tương quan mẫu giữa X và Y. Lập phương trình đường hồi quy tuyến tính của Y theo X.

Lời giải

a) Gọi μ_X là chỉ tiêu X trung bình. Với mức ý nghĩa $\alpha = 5\% = 0,05$, ta đặt giả thiết kiểm định như sau

$$\begin{cases} H_0 : \mu_x = 145 \\ H_1 : \mu_x < 145 \end{cases}$$

Từ bảng dữ liệu, ta có $n = 110$, $\overline{x} = 142,5454$; $s_X = 9,9019$ và $\mu_0 = 145$. Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s_X}\sqrt{n} = \frac{142,5454 - 145}{9,9019}\sqrt{110} = -2,5999$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = Z_{0,05} = -1,6449$$

So sánh thì ta có $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0.Vậy đủ cơ sở để khẳng định chỉ tiêu X trung bình là nhỏ hơn yêu cầu với mức ý nghĩa 5%.

b) Gọi Z là chỉ tiêu Y của sản phẩm loại I và μ_Z là chỉ tiêu Y trung bình. Từ bảng dữ liệu trên đề bài, ta tính được $n_Z = 21$, $\overline{z} = 15,5238$ và $s_Z = 1,9905$. Do $n < 30$ nên ta có độ tin cậy

$$1 - \alpha = 90\% = 0,9 \Leftrightarrow \alpha = 0,1 \Leftrightarrow t^{n-1}_{\frac{\alpha}{2}} = t^{20}_{0,05} = 1,7247$$

Suy ra độ chính xác là

$$\varepsilon = t_{\frac{\alpha}{2}}^{n-1} \cdot \frac{s}{\sqrt{n}} = 1,7247 \cdot \frac{1,9905}{\sqrt{21}} = 0,7491$$

Vậy khoảng tin cậy 90% của μ_Z là

$$(\overline{z} - \varepsilon; \overline{z} + \varepsilon) = (15,5238 - 0,7491; 15,5238 + 0,7491) = (14,7747; 16,2729)$$

Vậy trung bình chỉ tiêu Y của sản phẩm loại I với độ tin cậy 90% là khoảng từ $14,7747$ đến $16,2729$.
c) Từ bảng dữ liệu, ta tính được $\overline{y} = 8,0591$; $s_Y = 5,2946$ và $\overline{xy} = 1232,3636$. Suy ra hệ số tương quan mẫu là

$$r = \frac{n}{n-1}\left(\frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_X \cdot s_Y}\right) = \frac{110}{109} \cdot \frac{1232,3636 - 142,5454 \cdot 8,0591}{9,9019 \cdot 5,2946} \approx 1,6076$$

Phương trình đường hồi quy tuyến tính của Y theo X là

$$Y - \overline{y} = r\frac{s_Y}{s_X}(X - \overline{x}) \Leftrightarrow Y - 8,0591 = 1,6076 \cdot \frac{5,2946}{9,9019}(X - 142,5454)$$

$$\Leftrightarrow Y = 0,8596X - 114,472$$

Chương VI

Hướng dẫn giải một số đề thi

1 Đề thi kết thúc học phần học kỳ I, năm học 2019 - 2020 của các lớp học phần 60.CNOT-2, 60.CNOT-3, 60.NTTS-2, 60.TTQL, 60.BHTS, 60.CNTT-3 (ngày 17/01/2020 - mã đề 8)

Câu 1. (2 điểm) Một phân xưởng có ba máy. Xác suất các máy I, II, III bị hỏng trong ngày tương ứng là $0,23$; $0,36$ và $0,14$. Tính xác suất của các biến cố sau
a) Có một máy bị hỏng trong ngày.
b. Có ít nhất một máy bị hỏng trong ngày.

Lời giải

Gọi A_i với $i = \overline{1,3}$ là biến cố "Máy thứ i bị hỏng trong ngày" thì ta có A_1, A_2, A_3 độc lập với nhau và $P(A_1) = 0,23$; $P(A_2) = 0,36$; $P(A_3) = 0,14$.
a) Gọi X là biến cố "Có một máy bị hỏng trong ngày" thì ta có

$$X = A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3$$

Do các biến cố $A_1 \cdot \overline{A_2} \cdot \overline{A_3}$, $\overline{A_1} \cdot A_2 \cdot \overline{A_3}$, $\overline{A_1} \cdot \overline{A_2} \cdot A_3$ đôi một xung khắc nên suy ra

$$\begin{aligned}
P(X) &= P\left(A_1 \cdot \overline{A_2} \cdot \overline{A_3} \cup \overline{A_1} \cdot A_2 \cdot \overline{A_3} \cup \overline{A_1} \cdot \overline{A_2} \cdot A_3\right) \\
&= P(A_1) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) + P\left(\overline{A_1}\right) \cdot P(A_2) \cdot P\left(\overline{A_3}\right) + P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P(A_3) \\
&= 0,23 \cdot (1-0,36) \cdot (1-0,14) + (1-0,23) \cdot 0,36 \cdot (1-0,14) + (1-0,23) \cdot (1-0,36) \cdot 0,14 \\
&= 0,434
\end{aligned}$$

b) Gọi Y là biến cố "Có ít nhất một máy bị hỏng trong ngày" thì ta có $\overline{Y}$ là biến cố "Không có máy nào bị hỏng trong ngày", tức

$$\overline{Y} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$$

Suy ra

$$\begin{aligned}
P(Y) &= 1 - P(\overline{Y}) = 1 - P\left(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}\right) \\
&= 1 - P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) = 1 - (1-0,23) \cdot (1-0,36) \cdot (1-0,14) \\
&= 0,5762
\end{aligned}$$

Câu 2. (2,5 điểm) Chiều dài mỗi sản phẩm do một phân xưởng sản xuất là một biến ngẫu nhiên tuân theo phân phối chuẩn, với độ lệch chuẩn bằng $0,6(mm)$. Sản phẩm được gọi là đạt chất lượng cao nếu chiều dài sản phẩm sai lệch với chiều dài trung bình không quá $0,083(mm)$.
a) Tìm tỉ lệ sản phẩm đạt chất lượng cao của phân xưởng.
b) Lấy ngẫu nhiên 11 sản phẩm. Tính xác suất để có ít nhất 2 sản phẩm đạt chất lượng cao.

Lời giải

a) Gọi $X(mm)$ là chiều dài sản phẩm do phân xưởng sản xuất thì ta có $X \sim \mathcal{N}\left(\mu;(0,6)^2\right)$. Khi đó tỉ lệ sản phẩm đạt chất lượng cao của phân xưởng là

$$P\left(|X-\mu| \leq 0,083\right) = 2\phi\left(\frac{0,083}{0,6}\right) = 2\phi\left(0,1383\right) = 0,11$$

Vậy tỉ lệ sản phẩm đạt chất lượng cao của phân xưởng là 11%.
b) Gọi Y là số lượng sản phẩm đạt chất lượng cao thì ta có $Y \sim \mathcal{B}\left(11;0,11\right)$. Suy ra

$$\begin{aligned} P\left(Y \geq 2\right) &= 1 - P\left(Y < 2\right) = 1 - P\left(Y = 1\right) - P\left(Y = 0\right) \\ &= 1 - C_8^1 \cdot (0,11)^1 \cdot (1-0,11)^7 - C_8^0 \cdot (0,11)^0 \cdot (1-0,11)^8 \\ &= 0,2171 \end{aligned}$$

Câu 3. (3,5 điểm) Quan sát về thời gian cần thiết để sản xuất một chi tiết máy, người ta thu được các số liệu cho ở bảng

Khoảng thời gian (phút)	25 − 30	30 − 35	35 − 40	40 − 45	45 − 50
Số quan sát	14	27	32	15	9

a) Hãy ước lượng thời gian trung bình để sản xuất một chi tiết máy với độ tin cậy 96%.
b) Nếu muốn ước lượng trên có độ chính xác là 0,96 (phút) thì độ tin cậy bằng bao nhiêu.
c) Nếu muốn ước lượng có độ chính xác như trên và có độ tin cậy 99% thì cần quan sát thêm bao nhiêu chi tiết.

Lời giải

a) Gọi X (phút) là thời gian để sản xuất chi tiết trên và μ là thời gian trung bình thì ta cần ước lượng μ với độ tin cậy 96%. Từ bảng dữ liệu trên đề bài, ta tính được $n = 97$, $\overline{x} = 36,366$ và $s = 5,7962$. Độ tin cậy

$$1-\alpha = 96\% = 0,96 \Leftrightarrow \alpha = 0,04 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,02} = 2,0537$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2,0537 \cdot \frac{5,7962}{\sqrt{97}} = 1,2086$$

Khoảng tin cậy 96% của μ là

$$\left(\overline{x}-\varepsilon;\overline{x}+\varepsilon\right) = \left(36,366-1,2086;36,366+1,2086\right) = \left(35,1574;37,5746\right)$$

Vậy thời gian trung bình để sản xuất chi tiết máy trên với độ tin cậy 96% là khoảng từ 35,1574 đến 37,5746 (phút).
b) Với độ chính xác là 0,96 (phút) thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \Leftrightarrow Z_{\frac{\alpha}{2}} = \frac{\varepsilon \cdot \sqrt{n}}{s} = \frac{0,96 \cdot \sqrt{97}}{5,7962} = 1,6312$$

Suy ra $\frac{\alpha}{2} = 0,0516$. Vậy độ tin cậy là $1-\alpha = 1 - 2 \cdot 0,0516 = 0,8968 = 89,68\%$.
c) Ta có

$$1-\alpha = 99\% = 0,99 \Leftrightarrow \alpha = 0,01 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,005} = 2,5758$$

Với độ chính xác như trên thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n_1}} \Leftrightarrow n_1 = Z_{\frac{\alpha}{2}}^2 \cdot \frac{s^2}{\varepsilon^2} = 2,5758^2 \cdot \frac{5,7962^2}{0,96^2} = 241,8625$$

Vậy cần quan sát thêm $n_1 - n = 242 - 97 = 145$ chi tiết nữa.

Câu 4. (2 điểm) Trong giai đoạn nền kinh tế đang tăng trưởng, các nhà kinh tế cho rằng các công ty nhỏ thường thuê nhiều nhân công hơn và sa thải ít hơn so với các công ty lớn. Để kiểm tra điều này, người ta lấy một mẫu ngẫu nhiên gồm một số công ty. Bảng sau đây cho biết số công nhân tuyển thêm và số công nhân bị sa thải của các công ty được chia theo quy mô của các công ty

B / A	Công ty lớn	Công ty vừa	Công ty nhỏ
Số lượng sa thải	82	96	30
Số lượng tuyển thêm	324	291	213

Với mức ý nghĩa 2%, dựa vào mẫu này có thể khẳng định việc thuê thêm hay sa thải công nhân độc lập với quy mô của các công ty hay không.

Lời giải

Với mức ý nghĩa $\alpha = 2\% = 0,02$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

B / A	Công ty lớn	Công ty vừa	Công ty nhỏ	n_i
Số lượng sa thải	82 (81,5135)	96 (77,6988)	30 (48,7876)	208
Số lượng tuyển thêm	324 (324,4865)	291 (309,3012)	213 (194,2123)	828
m_j	406	387	243	$n = 1036$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{82^2}{81,5135} + \frac{96^2}{77,6988} + \frac{30^2}{48,7876} + \frac{324^2}{324,4865} + \frac{291^2}{309,3012} + \frac{213^2}{194,2123}\right) - 1036$$
$$= 14,4497$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,02}\left((2-1)(3-1)\right) = \mathcal{X}^2_{0,02}(2) = 7,824$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên đủ cơ sở để bác bỏ H_0, nhận H_1. Vậy với mức ý nghĩa 2%, ta có thể khẳng định việc thuê thêm hay sa thải công nhân có liên quan với quy mô của các công ty.

2 Đề thi kết thúc học phần học kỳ II, năm học 2022 - 2023 của các lớp học phần 64.MARKT-1, 64.MARKT-3 (ngày 04/06/2023 - mã đề 2)

Câu 1 (2 điểm). Gọi $X(g)$ là trọng lượng của một loại sản phẩm do máy tự động sản xuất. Giả sử X là biến ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình là $\mu = 120(g)$ và độ lệch chuẩn $\sigma = 10(g)$. Những sản phẩm có trọng lượng từ $110(g)$ đến $130(g)$ là sản phẩm loại I.
a) Tính $P(X < 100)$, $P(110 \le X \le 130)$. Minh họa các kết quả bằng hình vẽ.
b) Một đại lý mua đồng giá mỗi sản phẩm của nhà máy với giá 10.000 đồng. Giá bán mỗi sản phẩm loại I là 15.000 đồng, mỗi sản phẩm thường giá 12.000 đồng. Kỳ vọng mỗi sản phẩm đại lý lời bao nhiêu tiền.

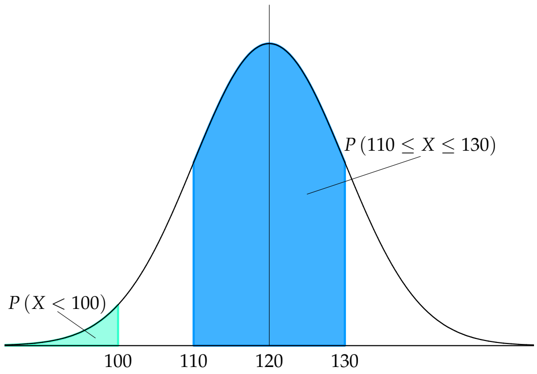
Lời giải

a) Từ đề bài, ta có $X \sim \mathcal{N}\left(120; 10^2\right)$. Suy ra

• $P(X < 100) = \phi\left(\dfrac{100-120}{10}\right) - \phi(-\infty) = \phi(-2) - \phi(-\infty) = -0,4772 + 0,5 = 0,0228$. • $P(110 \le X \le 130) =$

$\phi\left(\dfrac{130-120}{10}\right)-\phi\left(\dfrac{110-120}{10}\right)=\phi(1)-\phi(-1)=2\phi(1)=0,6826.$

Hình vẽ minh họa



b) Gọi Y là số tiền lời khi bán mỗi sản phẩm của công ty trên ta có Y là biến ngẫu nhiên rời rạc có bảng phân phối xác suất như sau

Y_i	2.000	5.000	Σ
$P(Y=Y_i)$	0,3174	0,6826	1

Suy ra

$$E[Y]=2.000\cdot 0,3174+5.000\cdot 0,6826=4.047,8$$

Vậy kỳ vọng mỗi sản phẩm đại lý lời 4.047, 8 (đồng).

Câu 2 (5 điểm). Để khảo sát nhu cầu tiêu thụ lương thực trong một tháng của người dân thành phố Nha Trang năm 2022, tiến hành khảo sát ngẫu nhiên 400 người cho kết quả $\overline{x}=12,5(kg)$, $s=2,5(kg)$ và có hơn 160 người có nhu cầu sử dụng hơn 15 (kg/tháng).
a) Ước lượng nhu cầu tiêu thụ lương thực trung bình trong một tháng của một người dân Nha Trang với độ tin cậy 95%.
b) Ước lượng tỉ lệ người dân Nha Trang có nhu cầu sử dụng hơn 15 (kg/tháng) với độ tin cậy 95%.
c) Với khoảng ước lượng ở câu b, để độ tin cậy là 98% thì cần khảo sát thêm bao nhiêu người nữa.
d) Nêu tổng thể, mẫu của bài toán trên, ý nghĩa của bài toán ước lượng.

Lời giải

a) Gọi $X(kg)$ là nhu cầu tiêu thụ lương thực trong một tháng của người dân thành phố Nha Trang năm 2022 và μ là nhu cầu trung bình. Từ đề bài, ta có $n=100$, $\overline{x}=12,5$ và $s=2,5$. Độ tin cậy

$$1-\alpha=95\%=0,95\Leftrightarrow \alpha=0,05\Leftrightarrow Z_{\frac{\alpha}{2}}=Z_{0,025}=1,96$$

Suy ra độ chính xác là

$$\varepsilon=Z_{\frac{\alpha}{2}}\cdot\frac{s}{\sqrt{n}}=1,96\cdot\frac{2,5}{\sqrt{200}}=0,3465$$

Khoảng tin cậy 95% của μ là

$$(\overline{x}-\varepsilon;\overline{x}+\varepsilon)=(12,5-0,3465;12,5+0,3465)=(12,1535;12,8465)$$

Vậy nhu cầu tiêu thụ lương thực trung bình trong một tháng của một người dân Nha Trang với độ tin cậy 95% là khoảng từ $12,1535(kg)$ đến $12,8465(kg)$.

b) Gọi p là tỉ lệ người dân Nha Trang có nhu cầu sử dụng hơn 15 (kg/tháng). Từ đề bài, ta có $n=200$ và tỉ lệ mẫu $f=\dfrac{160}{200}=0,8$ thỏa điều kiện $\begin{cases} nf>10 \\ n(1-f)>10 \end{cases}$. Độ tin cậy

$$1-\alpha=95\%=0,95\Leftrightarrow \alpha=0,05\Leftrightarrow Z_{\frac{\alpha}{2}}=Z_{0,025}=1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon=Z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{f(1-f)}{n}}=1,96\cdot\sqrt{\frac{0,8\cdot 0,2}{200}}=0,0554$$

Suy ra khoảng tin cậy 95% của p là

$$(f-\varepsilon; f+\varepsilon) = (0,8-0,0554; 0,8+0,0554) = (0,7446; 0,8455)$$

Vậy tỉ lệ người dân Nha Trang có nhu cầu sử dụng hơn 15 (kg/tháng) với độ tin cậy 95% là khoảng từ $74,46\%$ đến $85,54\%$.
c) Độ tin cậy

$$1-\alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Suy ra

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n_1}} \Leftrightarrow n_1 = Z^2_{\frac{\alpha}{2}} \cdot \frac{f(1-f)}{\varepsilon^2} = 2,3263^2 \cdot \frac{0,8\cdot 0,2}{0,0554^2} = 282,1187$$

Vậy cần khảo sát thêm $n_1 - n = 283 - 200 = 83$ người.
d) Tổng thể là nhu cầu tiêu thụ lương thực trong một tháng của người dân thành phố Nha Trang năm 2022.
Mẫu là nhu cầu tiêu thụ lương thực trong một tháng của 400 được khảo sát ngẫu nhiên.
Ý nghĩa của bài toán ước lượng là để xác định được khoảng chứa giá trị cần ước lượng với xác suất cao.

Câu 3 (3 điểm). Điều tra ngẫu nhiên thu nhập (triệu đồng/năm) của 400 công nhân Điện lực làm việc ở Hà Nội và TPHCM, ta thu được kết quả sau

Thu nhập / Thành phố	Dưới 130	130 − 180	Trên 180
Hà Nội	35	50	40
TPHCM	70	105	100

Với mức ý nghĩa 2%, hãy kết luận xem thu nhập của công nhân có phụ thuộc vào thành phố mà họ làm việc hay không.

Lời giải

Gọi A là thu nhập của công nhân, B là thành phố mà họ làm việc. Giả thuyết kiểm định

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Thu nhập / Thành phố	Dưới 130	130 − 180	Trên 180	n_i
Hà Nội	35 (32, 8125)	50 (48, 4375)	40 (43, 75)	125
TPHCM	70 (72, 1875)	105 (106, 5625)	100 (96, 25)	275
m_j	105	155	140	$n = 400$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{35^2}{32,8125} + \frac{50^2}{48,4375} + \frac{40^2}{43,75} + \frac{70^2}{72,1875} + \frac{105^2}{106,5625} + \frac{100^2}{96,25}\right) - 400$$
$$= 0,753$$

Với mức ý nghĩa $\alpha = 2\% = 0,02$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha((h-1)(k-1)) = \mathcal{X}^2_{0,02}((2-1)(3-1)) = \mathcal{X}^2_{0,02}(2) = 7,824$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha((h-1)(k-1))$ nên không đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 2%, ta có thể khẳng định thu nhập của công nhân không có phụ thuộc vào thành phố mà họ làm việc.

3 Đề thi kết thúc học phần học kỳ II, năm học 2022 - 2023 của các lớp học phần 63.KDTM-1, 63.KDTM-2 (ngày 08/06/2023 - mã đề 2)

Câu 1 (3 điểm). Đường kính của một loại trục máy được sản xuất tại một nhà máy là một đại lượng ngẫu nhiên có phân phối chuẩn với trung bình là $180mm$ và độ lệch chuẩn là $1,5mm$. Trục máy được gọi là hợp quy cách nếu có đường kính từ $177mm$ đến $183mm$.
a) Tính tỉ lệ trục máy sản xuất hợp quy cách tại nhà máy.
b) Kiểm tra ngẫu nhiên 8 trục được hoàn thành. Tính xác suất có ít nhất 7 trục hợp quy cách.
c) Nhà máy cho sản xuất 200 trục mỗi ngày. Tính số trục máy hợp quy cách trung bình mỗi ngày.

Lời giải

a) Gọi $X(mm)$ là đường kính của trục máy sản xuất tại nhà máy thì ta có $X \sim \mathcal{N}\left(180; (1,5)^2\right)$. Khi đó tỉ lệ trục máy sản xuất hợp quy cách tại nhà máy là

$$P(177 \leq X \leq 183) = \phi\left(\frac{183-180}{1,5}\right) - \phi\left(\frac{177-180}{1,5}\right) = \phi(2) - \phi(-2) = 2\phi(2) = 0,9544$$

Vậy tỉ lệ trục máy sản xuất hợp quy cách tại nhà máy là $95,44\%$.
b) Gọi Y là số lượng trục máy hợp quy cách thì ta có $Y \sim \mathcal{B}(8; 0,9544)$. Suy ra

$$\begin{aligned} P(Y \geq 7) &= P(Y=7) + P(Y=8) \\ &= C_8^7 \cdot (0,9544)^7 \cdot (1-0,9544)^1 + C_8^8 \cdot (0,9544)^8 \cdot (1-0,9544)^0 \\ &= 0,9515 \end{aligned}$$

c) Gọi Z là số trục máy được sản xuất hợp quy cách mỗi ngày thì ta có $Z \sim \mathcal{B}(200; 0,9544)$. Suy ra

$$E[Z] = 200 \cdot 0,9544 = 190,88$$

Vậy trung bình có 191 trục máy hợp quy cách được sản xuất mỗi ngày.

Câu 2 (4 điểm). Tiến hành khảo sát lượng gạo bán hàng ngày tại một đại lý, ta có số liệu sau

Lượng gạo (kg)	Số ngày	Lượng gạo (kg)	Số ngày
$110-120$	6	$150-160$	25
$120-130$	17	$160-170$	17
$130-140$	22	$170-180$	8
$140-150$	30		

a) Những ngày bán hơn 160 (kg) là những ngày cao điểm. Hãy ước lượng tỉ lệ những ngày cao điểm với độ tin cậy 96%.
b) Nếu muốn độ chính xác ở câu a) là $0,05$ thì cần khảo sát thêm bao nhiêu ngày nữa.
c) Đại lý sẽ ngưng bán vì lỗ nếu trung bình mỗi ngày bán dưới 150 (kg). Từ số liệu khảo sát, cửa hàng quyết định thế nào với mức ý nghĩa $0,05$.

Lời giải

a) Gọi p là tỉ lệ những ngày cao điểm. Từ bảng dữ liệu trên, ta tính được $n = 125$ và $f = \dfrac{25}{125} = 0,2$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Ta có

$$1-\alpha = 96\% = 0,96 \Leftrightarrow \alpha = 0,04 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,02} = 2,0537$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,0537 \cdot \sqrt{\frac{0,2 \cdot 0,8}{125}} = 0,0735$$

Suy ra khoảng tin cậy 96% của p là

$$(f-\varepsilon; f+\varepsilon) = (0,2-0,0735; 0,2+0,0735) = (0,1265; 0,2735)$$

Vậy tỉ lệ những ngày cao điểm với độ tin cậy 96% là khoảng từ 12,65% đến 27,35%.
b) Với độ chính xác là 0,05 thì ta có

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n_1}} \Leftrightarrow n_1 = Z^2_{\frac{\alpha}{2}} \cdot \frac{f(1-f)}{\varepsilon^2} = 2,0537^2 \cdot \frac{0,2\cdot 0,8}{0,05^2} = 269,9318$$

Vậy cần khảo sát thêm $n_1 - n = 270 - 125 = 145$ ngày nữa.
c) Gọi X (kg) là lượng gạo bán mỗi ngày và μ là lượng gạo trung bình. Với mức ý nghĩa $\alpha = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \mu = 150 \\ H_1 : \mu < 150 \end{cases}$$

Từ bảng số liệu, ta tính được $n = 125$, $\overline{x} = 145,72$ và $s = 15,7693$. Tiêu chuẩn kiểm định là

$$Z = \frac{\overline{x}-\mu_0}{s}\sqrt{n} = \frac{145,72-150}{15,7693}\sqrt{125} = -3,0345$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,05} = -1,6449$$

So sánh thì ta có $Z < -Z_\alpha$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1.Vậy với mức ý nghĩa 1%, ta đủ cơ sở để khẳng định đại lý sẽ ngưng bán vì lỗ do trung bình mỗi ngày bán dưới 150 (kg).

Câu 3 (3 điểm). Một công ty mỹ phẩm nghiên cứu thị trường về mức thu nhập của khách hàng và loại sản phẩm mà họ chọn mua. Số liệu cho ở bảng sau

Loại sản phẩm \ Mức thu nhập	Thấp	Trung bình	Cao
Loại A	65	95	46
Loại B	53	103	108

Với mức ý nghĩa 0,01, hãy cho biết mức thu nhập của khách hàng có ảnh hưởng gì đến việc chọn mua loại sản phẩm hay không. Hãy giải thích cụ thể mức ý nghĩa trong bài này.

Lời giải

Ta gọi X là loại hàng và Y là mức thu nhập. Giả thuyết kiểm định như sau

$$\begin{cases} H_0 : X \text{ và } Y \text{ độc lập} \\ H_1 : X \text{ và } Y \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Loại sản phẩm \ Mức thu nhập	Thấp	Trung bình	Cao	n_i
Loại A	65 (51,7191)	95 (86,783)	46 (67,4979)	206
Loại B	53 (66,2808)	103 (111,217)	108 (86,5021)	264
m_j	118	198	154	$n = 470$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{65^2}{51,7191} + \frac{95^2}{86,783} + \frac{46^2}{67,4979} + \frac{53^2}{66,2808} + \frac{103^2}{111,217} + \frac{108^2}{86,5021}\right) - 470$$
$$= 19,6465$$

Với mức ý nghĩa $\alpha = 0,01$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,01}\left((2-1)(3-1)\right) = \mathcal{X}^2_{0,01}\left(2\right) = 9,2103$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên đủ cơ sở để bác bỏ H_0, nhận H_1. Vậy với mức ý nghĩa $0,01$, ta có thể khẳng định mức thu nhập của khách hàng có ảnh hưởng đến việc chọn mua loại sản phẩm. Mức ý nghĩa $0,01$ cho biết rằng xác suất tối đa của việc bác bỏ H_0 trong khi trên thực tế nó đúng là 1%.

4 Đề thi kết thúc học phần học kỳ II, năm học 2022 - 2023 của lớp học phần 63.CDT (ngày 08/06/2023)

Câu 1 (2,0 điểm). Các kết quả của bài kiểm tra chỉ số thông minh (IQ) cho các học sinh của một trường trung học cho thấy điểm IQ của các học sinh này tuân theo phân phối chuẩn với các tham số $\mu = 100$ và $\sigma^2 = 225$.
a) Tính tỷ lệ học sinh có điểm IQ nhỏ hơn 85.
b) Tính tỷ lệ học sinh có điểm IQ từ 115 đến 130.

Lời giải

Gọi X là điểm IQ của các học sinh thì ta có $X \sim \mathcal{N}\left(100; 15^2\right)$.
a) Ta có

$$P\left(X < 85\right) = 0,5 + \phi\left(\frac{85-100}{15}\right) = 0,5 + \phi\left(-1\right) = 0,5 - \phi\left(1\right) = 0,5 - 0,3413 = 0,1587$$

Vậy tỷ lệ học sinh có điểm IQ nhỏ hơn 85 là $15,87\%$.
b) Ta có

$$P\left(115 \le X \le 130\right) = \phi\left(\frac{130-100}{15}\right) - \phi\left(\frac{115-100}{15}\right) = \phi\left(2\right) - \phi\left(1\right) = 0,4772 - 0,3413 = 0,1359$$

Vậy tỷ lệ học sinh có điểm IQ từ 115 đến 130 là $13,59\%$.

Câu 2 (6,0 điểm). Cân ngẫu nhiên 100 quả cam ta được trọng lượng trung bình là $35,89(g)$, độ lệch mẫu là $1,792(g)$.
a) Hãy ước lượng trọng lượng trung bình các quả cam với độ tin cậy 95%.
b) Nếu muốn ước lượng ở câu a) có độ chính xác $0,25(g)$ thì phải quan sát thêm bao nhiêu quả.
c) Cam có khối lượng dưới $35(g)$ được coi là cam loại 2, số lượng loại cam này trong 100 quả được cân ngẫu nhiên là 20 quả. Tìm khoảng ước lượng cho tỷ lệ cam loại 2 với độ tin cậy 90%.

Lời giải

Gọi $X(g)$ là trọng lượng các quả cam và μ là trọng lượng trung bình thì ta có cỡ mẫu $n = 100$, trung bình mẫu $\overline{x} = 35,89(g)$ và độ lệch mẫu $s = 1,792(g)$.
a) Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{1,792}{\sqrt{100}} = 0,3512$$

Khoảng tin cậy 95% của μ là

$$\left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) = \left(35,89 - 0,3512; 35,89 + 0,3512\right) = \left(35,5388; 36,2412\right)$$

Vậy trọng lượng trung bình các quả cam với độ tin cậy 95% là khoảng từ $35,5388$ đến $36,2412(g)$.
b) Với độ chính xác là $0,25(g)$ và độ tin cậy giữ nguyên thì

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \Leftrightarrow n = \left(Z_{\frac{\alpha}{2}} \cdot \frac{s}{\varepsilon}\right)^2 = \left(1,96 \cdot \frac{1,792}{0,25}\right)^2 = 197,3823$$

Vậy cần quan sát thêm $198 - 100 = 98$ quả cam nữa.

c) Gọi p là tỷ lệ cam loại 2 thì từ đề bài, ta tính được cỡ mẫu $n = 100$ và tỷ lệ mẫu $f = \dfrac{20}{100} = 0,2$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 90\% = 0,9 \Leftrightarrow \alpha = 0,1 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,05} = 1,6449$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,6449 \cdot \sqrt{\frac{0,2 \cdot 0,8}{100}} = 0,0784$$

Suy ra khoảng tin cậy 90% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,2 - 0,0784; 0,2 + 0,0784) = (0,1216; 0,2784)$$

Vậy tỷ lệ cam loại 2 với độ tin cậy 90% là khoảng từ $12,16\%$ đến $27,84\%$.

Câu 3 (2,0 điểm). Số liệu cho bởi bảng bên dưới liên quan đến tuổi các bà mẹ và trọng lượng (kg) của những đứa trẻ sơ sinh (của các bà mẹ).

Tuổi của mẹ	**Trọng lượng của trẻ sơ sinh**	
	$< 2,5$ kg	$\geq 2,5$ kg
≤ 20 tuổi	10	40
> 20 tuổi	15	135

Với mức ý nghĩa 5%, có thể cho rằng trọng lượng trẻ sơ sinh có phụ thuộc vào độ tuổi của mẹ không.

Lời giải

Giả thuyết kiểm định

$$\begin{cases} H_0: & \text{trọng lượng trẻ sơ sinh không phụ thuộc vào độ tuổi của mẹ} \\ H_1: & \text{trọng lượng trẻ sơ sinh có phụ thuộc vào độ tuổi của mẹ} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Trọng lượng trẻ sơ sinh / Tuổi của mẹ	$< 2,5$ kg	$\geq 2,5$ kg	Tổng
≤ 20 tuổi	10 $(6,25)$	40 $(43,75)$	50
> 20 tuổi	15 $(18,75)$	135 $(131,25)$	150
Tổng	25	175	$n = 200$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,2}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=2} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{10^2}{6,25} + \frac{40^2}{43,75} + \frac{15^2}{18,75} + \frac{135^2}{131,25}\right) - 200 = 3,4286$$

Với mức ý nghĩa $0,05$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_{\alpha}((h-1)(k-1)) = \mathcal{X}^2_{0,05}((2-1)(2-1)) = \mathcal{X}^2_{0,05}(1) = 3,814$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_{\alpha}((h-1)(k-1))$ nên chấp nhận H_0. Vậy có thể cho rằng trọng lượng trẻ sơ sinh không phụ thuộc vào độ tuổi của mẹ với mức ý nghĩa 5%.

5 Đề thi kết thúc học phần học kỳ I, năm học 2023 - 2024 của các lớp học phần 64.CNXD-1, 64.CNXD-2 (ngày 29/12/2023 - mã đề 1)

Câu 1 (2,5 điểm). Đường kính của một loại trục máy được sản xuất tại một nhà máy là một đại lượng ngẫu nhiên có phân phối chuẩn với trung bình là $170mm$ và độ lệch chuẩn là $1,5mm$. Trục máy được gọi là hợp quy cách nếu có đường kính từ $167mm$ đến $173mm$.
a) Tính tỉ lệ trục máy sản xuất hợp quy cách tại nhà máy.
b) Nhà máy cho sản xuất 100 trục mỗi ngày. Tính số trục máy hợp quy cách trung bình mỗi ngày.

Lời giải

a) Gọi $X(mm)$ là đường kính của trục máy sản xuất tại nhà máy thì ta có $X \sim \mathcal{N}\left(170; (1,5)^2\right)$. Khi đó tỉ lệ trục máy sản xuất hợp quy cách tại nhà máy là

$$P(167 \leq X \leq 173) = \phi\left(\frac{173-170}{1,5}\right) - \phi\left(\frac{167-170}{1,5}\right) = \phi(2) - \phi(-2) = 2\phi(2) = 0,9544$$

Vậy tỉ lệ trục máy sản xuất hợp quy cách tại nhà máy là $95,44\%$.
b) Gọi Y là số trục máy được sản xuất hợp quy cách mỗi ngày thì ta có $Y \sim \mathcal{B}(100; 0,9544)$. Suy ra

$$E[Y] = 100 \cdot 0,9544 = 95,44$$

Vậy trung bình có 96 trục máy hợp quy cách được sản xuất mỗi ngày.

Câu 2 (4,5 điểm). Khảo sát thời gian lắp ráp chi tiết máy của một loại thiết bị điện tử ở một nhà máy sau khi áp dụng kỹ thuật mới, người ta chọn ngẫu nhiên 100 sản phẩm đã hoàn tất và thu được số liệu sau

Thời gian (phút)	$20-22$	$22-24$	$24-26$	$26-28$	$28-30$	$30-32$	$32-34$
Số chi tiết	6	10	17	31	21	11	4

a) Hãy ước lượng thời gian lắp ráp trung bình một thiết bị với độ tin cậy 95%.
b) Nếu muốn tăng độ tin cậy của khoảng ước lượng ở câu a lên 98% và độ chính xác $0,4$ phút thì cần phải điều tra thêm bao nhiêu sản phẩm nữa.
c) Trước khi áp dụng kỹ thuật mới, tỉ lệ sản phẩm có thời gian lắp ráp chậm (vượt 30 phút) là 20%. Với mức ý nghĩa $0,05$, có thể kết luận việc áp dụng kỹ thuật mới đã làm giảm tỉ lệ sản phẩm lắp ráp chậm không.

Lời giải

a) Gọi X (phút) là thời gian lắp ráp thiết bị và μ là thời gian lắp ráp trung bình. Từ bảng dữ liệu, ta tính được $n = 100$, $\overline{x} = 27$ và $s = 2,9267$. Độ tin cậy

$$1-\alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{2,9267}{\sqrt{100}} = 0,5736$$

Khoảng tin cậy 95% của μ là

$$(\overline{x}-\varepsilon; \overline{x}+\varepsilon) = (27-0,5736; 27+0,5736) = (26,4264; 27,5736)$$

Vậy thời gian lắp ráp trung bình một thiết bị với độ tin cậy 95% là khoảng từ $26,4264$ đến $27,5736$ (phút).
b) Độ tin cậy

$$1-\alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Với độ chính xác $0,4$ phút thì

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n_1}} \Leftrightarrow n_1 = Z_{\frac{\alpha}{2}}^2 \cdot \frac{s^2}{\varepsilon^2} = 2,3263^2 \cdot \frac{2,9267^2}{0,4^2} = 289,7129$$

Vậy cần khảo sát thêm $n_1 - n = 290 - 100 = 190$ sản phẩm.
c) Gọi p là tỉ lệ tỉ lệ sản phẩm có thời gian lắp ráp chậm sau khi áp dụng kỹ thuật mới. Với mức ý nghĩa $\alpha = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : p = 0,2 \\ H_1 : p < 0,2 \end{cases}$$

Theo câu a, ta có cỡ mẫu $n = 100$, tỉ lệ mẫu $f = \dfrac{15}{100} = 0,15$ và $p_0 = 0,2$ thỏa $\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0,15 - 0,2}{\sqrt{0,2 \cdot 0,8}}\sqrt{100} = -1,25$$

Ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,05} = -1,6449$$

So sánh thì ta có $Z > -Z_\alpha$ nên không đủ cơ sở để bác bỏ H_0.Vây ta không đủ cơ sở để khẳng định việc áp dụng kỹ thuật mới làm giảm tỉ lệ sản phẩm lắp ráp chậm với mức ý nghĩa $0,05$.

Câu 3 (3 điểm). Để nghiên cứu mối liên hệ giữa việc nghiện thuốc lá và bệnh huyết áp, người ta tiến hành điều tra 200 người và thu được kết quả sau

Nghiện thuốc / Huyết áp	Không nghiện	Nghiện nhẹ	Nghiện nặng
Huyết áp bình thường	50	25	28
Huyết áp cao	30	35	32

Với mức ý nghĩa $0,05$, có thể kết luận có sự liên quan giữa việc nghiện thuốc lá và bệnh huyết áp không. Kết luận này có thay đổi không nếu mức ý nghĩa là $0,01$.

Lời giải

Gọi X là việc nghiện thuốc, Y là bệnh huyết áp. Giả thuyết kiểm định

$$\begin{cases} H_0 : X \text{ và } Y \text{ độc lập} \\ H_1 : X \text{ và } Y \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Nghiện thuốc / Huyết áp	Không nghiện	Nghiện nhẹ	Nghiện nặng	n_i
Huyết áp bình thường	50 $(41,2)$	25 $(30,9)$	28 $(30,9)$	103
Huyết áp cao	30 $(38,8)$	35 $(29,1)$	32 $(29,1)$	97
m_j	80	60	60	$n = 200$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{50^2}{41,2} + \frac{25^2}{30,9} + \frac{28^2}{30,9} + \frac{30^2}{38,8} + \frac{35^2}{29,1} + \frac{32^2}{29,1}\right) - 200 = 6,7594$$

• Với mức ý nghĩa $0,05$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,05}\left((2-1)(3-1)\right) = \mathcal{X}^2_{0,05}(2) = 5,9915$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên đủ cơ sở để bác bỏ H_0, nhận H_1. Vậy ta có thể khẳng định có sự liên quan giữa việc nghiện thuốc lá và bệnh huyết áp với mức ý nghĩa $0,05$.
• Với mức ý nghĩa $0,01$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,01}\left((2-1)(3-1)\right) = \mathcal{X}^2_{0,01}(2) = 9,2103$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên không đủ cơ sở để bác bỏ H_0. Vậy ta có thể khẳng định không có sự liên quan giữa việc nghiện thuốc lá và bệnh huyết áp với mức ý nghĩa $0,01$.

6 Đề thi kết thúc học phần học kỳ I, năm học 2023 - 2024 của các lớp học phần 64.CBTS-MP, 64.CBTS, 64.CNTP (ngày 04/01/2024 - mã đề 2)

Câu 1 (2,5 điểm). Tại một trại nuôi tôm hùm xuất khẩu, trọng lượng các con tôm được cho là đại lượng ngẫu nhiên có phân phối chuẩn, với trung bình là $1,3$ (kg) và độ lệch chuẩn là $0,1$ (kg). Con tôm đạt tiêu chuẩn để xuất khẩu phải đạt từ $1,2$ kg trở lên.
a) Tính tỉ lệ tôm đạt tiêu chuẩn xuất khẩu ở trại này.
b) Trại xuất 200 con cho một công ty chế biến. Tính số con đạt tiêu chuẩn xuất khẩu trung bình trong số này.

Lời giải

a) Gọi X (kg) là trọng lượng mỗi con tôm thì ta có $X \sim \mathcal{N}\left(1,3; (0,1)^2\right)$. Khi đó tỉ lệ tôm đạt tiêu chuẩn xuất khẩu ở trại này là

$$P(X \geq 1,2) = \phi(+\infty) - \phi\left(\frac{1,2-1,3}{0,1}\right) = \phi(+\infty) - \phi(-1) = 0,5 + 0,3413 = 0,8413$$

Vậy tỉ lệ tôm đạt tiêu chuẩn xuất khẩu ở trại này là $84,13\%$.
b) Gọi Y là số con tôm đạt tiêu chuẩn xuất khẩu trong 200 con thì ta có $Y \sim \mathcal{B}(200; 0,8413)$. Suy ra

$$E[Y] = 200 \cdot 0,8413 = 168,26$$

Vậy trung bình có 169 số con tôm đạt tiêu chuẩn xuất khẩu trong 200 con tôm.

Câu 2 (4,5 điểm). Trên một đoạn đường giao thông, mức vận tốc tối đa cho phép đối với các xe cơ giới là 60 km/h. Để kiểm tra tình hình vi phạm tốc độ trên đoạn đường này, cảnh sát giao thông đã đo tốc độ ngẫu nhiên một số xe và có bảng số liệu sau

Tốc độ (km/h)	$35-40$	$40-45$	$45-50$	$50-55$	$55-60$	$60-65$	$65-70$
Số xe	5	16	25	30	18	12	4

a) Hãy ước lượng tỉ lệ xe vi phạm tốc độ cho phép trên đoạn đường này với độ tin cậy 98%.
b) Nếu muốn độ chính xác trong ước lượng ở câu a không vượt quá $0,05$ với độ tin cậy 95% thì cần điều tra thêm bao nhiêu xe nữa.
c) Một báo cáo cho biết tốc độ trung bình lưu thông trên đoạn đường này là trên 50 (km/h). Với mức ý nghĩa $0,05$, hãy cho nhận xét về báo cáo này.

Lời giải

a) Gọi p là tỉ lệ xe vi phạm tốc độ cho phép trên đoạn đường này. Từ đề bài, ta có $n = 110$ và tỉ lệ mẫu $f = \dfrac{16}{110} = 0,1454$ thỏa $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1-\alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,3263 \cdot \sqrt{\frac{0,1454 \cdot 0,8546}{110}} = 0,0782$$

Khoảng tin cậy 98% của p là

$$(f-\varepsilon; f+\varepsilon) = (0,1454 - 0,0782; 0,1454 + 0,0782) = (0,0672; 0,2236)$$

Vậy tỉ lệ xe vi phạm tốc độ trên đoạn đường này với độ tin cậy 98% là khoảng từ $6,72\%$ đến $22,36\%$.
b) Độ tin cậy

$$1-\alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Độ chính xác

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n_1}} = 1,96 \cdot \sqrt{\frac{0,1454 \cdot 0,8546}{n_1}} = \frac{0,6909}{\sqrt{n_1}}$$

Theo đề bài thì ta có

$$\varepsilon \leq 0,05 \Leftrightarrow \frac{0,6909}{\sqrt{n_1}} \leq 0,05 \Leftrightarrow n_1 \geq \left(\frac{0,6909}{0,05}\right)^2 = 763,7485$$

Vậy cần điều tra thêm $n_1 - n = 764 - 110 = 654$ xe.

c) Gọi X (km/h) là tốc độ lưu thông của xe cơ giới trên đoạn đường này và μ là tốc độ trung bình. Với mức ý nghĩa $\alpha = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu > 50 \end{cases}$$

Từ bảng số liệu, ta có $n = 110$, $\overline{x} = 51,6818$ và $s = 7,3111$. Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{51,6818 - 50}{7,3111/\sqrt{110}} = 2,4126$$

Ta sẽ bác bỏ H_0 khi

$$Z > Z_\alpha = Z_{0,05} = 1,6449$$

So sánh thì ta có $Z > Z_\alpha$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1.Vậy với mức ý nghĩa 0,05, ta đủ cơ sở để khẳng định tốc độ trung bình lưu thông trên đoạn đường này là trên 50 (km/h).

Câu 3 (3 điểm). Một trung tâm nghiên cứu đặc trị bệnh trắng đuôi cho cá thử nghiệm ba loại thuốc trên ba ao cá có điều kiện chăm sóc như nhau. Kết quả như sau

Tình trạng \ Loại thuốc	Loại I	Loại II	Loại III
Còn bệnh sau 30 ngày	58	102	65
Hết bệnh sau 30 ngày	102	118	75

Với mức ý nghĩa 5%, hãy cho biết các loại thuốc có ảnh hưởng gì đến hiệu quả chữa trị bệnh không. Hãy giải thích mức ý nghĩa 5% trong trường hợp này.

Lời giải

Gọi X là loại thuốc, Y là tình trạng bệnh. Giả thuyết kiểm định

$$\begin{cases} H_0 : X \text{ và } Y \text{ độc lập} \\ H_1 : X \text{ và } Y \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Tình trạng \ Loại thuốc	Loại I	Loại II	Loại III	n_i
Còn bệnh sau 30 ngày	58 (69,2308)	102 (95,1923)	65 (60,8769)	225
Hết bệnh sau 30 ngày	102 (90,7692)	118 (124,8677)	75 (79,4231)	295
m_j	160	220	140	$n = 520$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{58^2}{69,2308} + \frac{102^2}{95,1923} + \frac{65^2}{60,8769} + \frac{102^2}{90,7692} + \frac{118^2}{124,8677} + \frac{75^2}{79,4231}\right) - 520$$

$$= 4,2416$$

Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,05}\left((2-1)(3-1)\right) = \mathcal{X}^2_{0,05}\left(2\right) = 5,9915$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên không đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 5%, ta có thể khẳng định các loại thuốc không có ảnh hưởng gì đến hiệu quả chữa trị bệnh.

7 Đề thi kết thúc học phần học kỳ I, năm học 2023 - 2024 của các lớp học phần 64.DDT-1, 64.DDT-2, 64.TTQL (ngày 06/01/2024 - mã đề 2)

Câu 1 (2,5 điểm). Thời gian bảo hành của một loại sản phẩm được quy định là 3 năm. Biết tuổi thọ của mỗi sản phẩm là đại lượng ngẫu nhiên tuân theo phân phối chuẩn với tuổi thọ trung bình là $4,2$ năm và độ lệch chuẩn là $0,8$ năm.
a) Tính tỉ lệ máy bảo hành của loại sản phẩm này.
b) Một cửa hàng bán được 200 sản phẩm loại này trong tháng. Tính số sản phẩm trung bình phải bảo hành trong số đó.

Lời giải

a) Gọi X (năm) là tuổi thọ của mỗi sản phẩm thì ta có $X \sim \mathcal{N}\left(4,2;(0,8)^2\right)$. Khi đó tỉ lệ máy bảo hành của loại sản phẩm này

$$P\left(X \leq 3\right) = \phi\left(\frac{3-4,2}{0,8}\right) - \phi\left(-\infty\right) = \phi\left(-1,5\right) - \phi\left(-\infty\right) = -0,4332 + 0,5 = 0,0668$$

Vậy tỉ lệ máy bảo hành của loại sản phẩm này là $6,68\%$.
b) Gọi Y là số sản phẩm phải bảo hành trong 200 sản phẩm thì ta có $Y \sim \mathcal{B}\left(200; 0,0668\right)$. Suy ra

$$E\left[Y\right] = 200 \cdot 0,0668 = 13,36$$

Vậy trung bình có 14 sản phẩm phải bảo hành.

Câu 2 (4,5 điểm). Trong một đợt kiểm tra trọng lượng các bao bột mì đóng gói tại một phân xưởng, người ta kiểm tra 120 bao thì thấy trọng lượng trung bình của mỗi bao là $19,8$ kg và độ lệch mẫu là $0,32$ kg.
a) Với độ tin cậy 95%, hãy ước lượng trọng lượng trung bình của mỗi bao bột mì tại phân xưởng.
b) Nếu muốn độ chính xác của ước lượng ở câu a là $0,05$ kg và độ tin cậy là 98% thì cần kiểm tra thêm bao nhiêu bao.
c) Trong số 120 bao này, người ta thấy có 25 bao bị lỗi bao bì. Hãy ước lượng tỉ lệ bao bị lỗi bao bì tại phân xưởng này với độ tin cậy 96%.

Lời giải

a) Gọi X là trọng lượng một bao bột mì tại phân xưởng và μ là trọng lượng trung bình. Từ đề bài, ta có $n = 120$, $\overline{x} = 19,8$ và $s = 0,32$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{0,32}{\sqrt{120}} = 0,0572$$

Khoảng tin cậy 95% của μ là

$$\left(\overline{x} - \varepsilon; \overline{x} + \varepsilon\right) = \left(19,8 - 0,0572; 19,8 + 0,0572\right) = \left(19,7428; 19,8572\right)$$

Vậy trọng lượng trung bình của một bao bột mì tại phân xưởng với độ tin cậy 95% là khoảng từ $19,7428$ đến $19,8572$ (kg).
b) Độ tin cậy

$$1 - \alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Với độ chính xác $0,05$ kg thì

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n_1}} \Leftrightarrow n_1 = Z_{\frac{\alpha}{2}}^2 \cdot \frac{s^2}{\varepsilon^2} = 2,3263^2 \cdot \frac{0,32^2}{0,05^2} = 221,6621$$

Vậy cần khảo sát thêm $n_1 - n = 222 - 120 = 102$ bao bột mì.

c) Gọi p là tỉ lệ bao bị lỗi bao bì tại phân xưởng. Từ đề bài, ta có cỡ mẫu $n = 120$ và tỉ lệ mẫu $f = \frac{25}{120} = 0,2083$ thỏa $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 96\% = 0,96 \Leftrightarrow \alpha = 0,04 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,02} = 2,0537$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,0537 \cdot \sqrt{\frac{0,2083 \cdot 0,7917}{120}} = 0,0761$$

Suy ra khoảng tin cậy 96% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,2083 - 0,0761; 0,2083 + 0,0761) = (0,1322; 0,2844)$$

Vậy tỉ lệ bao bị lỗi bao bì tại phân xưởng với độ tin cậy 96% là khoảng từ $13,22\%$ đến $28,44\%$.

Câu 3 (3 điểm). Để điều tra mức độ xem TV của người dân một tỉnh, người ta chia mức độ xem phim thành 3 cấp: nhiều, vừa, ít. Kết quả điều tra 300 hộ như dưới đây.

Vùng \ Mức độ	Nhiều	Vừa	Ít
Thành phố	34	60	72
Ven nội	46	40	48

Với mức ý nghĩa $0,05$, có mối liên quan nào giữa mức độ xem TV và vùng địa lý không? Sai lầm loại I trong trường hợp này là gì.

Lời giải

Gọi X là mức độ xem TV, Y là vùng địa lý. Với mức ý nghĩa $\alpha = 0,05$ thì ta đặt giả thuyết kiểm định

$$\begin{cases} H_0 : X \text{ và } Y \text{ độc lập} \\ H_1 : X \text{ và } Y \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Vùng \ Mức độ	Nhiều	Vừa	Ít	n_i
Thành phố	34 (44,2667)	60 (55,3333)	72 (66,4)	166
Ven nội	46 (35,7333)	40 (44,6667)	48 (53,6)	134
m_j	80	100	120	$n = 300$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{34^2}{44,2667} + \frac{60^2}{55,3333} + \frac{72^2}{66,4} + \frac{46^2}{35,7333} + \frac{40^2}{44,6667} + \frac{48^2}{53,6}\right) - 300$$
$$= 7,2694$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}_{\alpha}^2\left((h-1)(k-1)\right) = \mathcal{X}_{0,05}^2\left((2-1)(3-1)\right) = \mathcal{X}_{0,05}^2(2) = 5,9915$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}_{\alpha}^2\left((h-1)(k-1)\right)$ nên đủ cơ sở để bác bỏ H_0, nhận H_1. Vậy với mức ý nghĩa $0,05$, ta có thể khẳng định có sự liên quan giữa mức độ xem TV và vùng địa lý.

Sai lầm loại I ở đây là việc ta khẳng định rằng có mối liên quan giữa mức độ xem TV và vùng địa lý trong khi trên thực tế chúng không liên quan với nhau.

8 Đề thi kết thúc học phần học kỳ I, năm học 2023 - 2024 của các lớp học phần 64.CNNL, 64.KTTT, 64.CTM (ngày 07/01/2024 - mã đề 2)

Câu 1 (5 điểm). Để nghiên cứu nhu cầu của một loại hàng hóa ở một khu vực, người ta tiến hành khảo sát một số hộ gia đình được chọn ngẫu nhiên trong khu vực. Kết quả cho ở bảng dưới

Nhu cầu (kg/tháng)	Số gia đình	Nhu cầu (kg/tháng)	Số gia đình
40 − 45	11	55 − 60	234
45 − 50	52	60 − 65	226
50 − 55	143	65 − 70	146

Hộ có nhu cầu hơn 50 (kg/tháng) là hộ không nghèo.
a) Hãy ước lượng tỉ lệ hộ không nghèo trong khu vực với độ tin cậy 97, 2%.
b) Hãy ước lượng nhu cầu trung bình của mỗi hộ không nghèo trong khu vực với độ tin cậy 95, 8%.

Lời giải

a) Gọi p là tỉ lệ hộ không nghèo trong khu vực. Từ bảng dữ liệu trên đề bài, ta tính được cỡ mẫu $n = 812$ và tỉ lệ mẫu $f = \dfrac{749}{812} = 0,9224$ thỏa $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 97,2\% = 0,972 \Leftrightarrow \alpha = 0,028 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,014} = 2,1973$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,1973 \cdot \sqrt{\frac{0,9224 \cdot 0,0776}{812}} = 0,0206$$

Suy ra khoảng tin cậy 97, 2% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,9224 - 0,0206; 0,9224 + 0,0206) = (0,9108; 0,943)$$

Vậy tỉ lệ hộ không nghèo trong khu vực với độ tin cậy 97, 2% là khoảng từ 91, 08% đến 94, 3%.
b) Gọi X là nhu cầu của mỗi hộ không nghèo trong khu vực và μ là nhu cầu trung bình. Từ đề bài, ta có $n = 749$, $\overline{x} = 60,0033$ và $s = 5,0573$. Độ tin cậy

$$1 - \alpha = 95,8\% = 0,958 \Leftrightarrow \alpha = 0,042 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,021} = 2,0335$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2,0335 \cdot \frac{5,0573}{\sqrt{749}} = 0,3748$$

Khoảng tin cậy 95, 8% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (60,0033 - 0,3758; 60,0033 + 0,3758) = (59,6275; 60,3791)$$

Vậy nhu cầu trung bình của mỗi hộ không nghèo trong khu vực với độ tin cậy 95, 8% là khoảng từ 59, 6275 đến 60, 3791 (kg/tháng).

Câu 2 (5 điểm). Một công ty có bốn loại kho chứa các sản phẩm cùng loại. Giám đốc công ty đó nói chất lượng sản phẩm và kho hàng là độc lập (tỉ lệ thành phần sản phẩm loại 1, 2 và 3 trong các kho hàng là như nhau). Kiểm tra ngẫu nhiên một sản phẩm ở các kho 1, 2, 3 và 4 có số liệu sau

Kho \ Chất lượng	Kho 1	Kho 2	Kho 3	Kho 4
Loại 1	112	124	106	119
Loại 2	213	211	227	225
Loại 3	165	197	180	161

Hãy kết luận về ý kiến trên với mức ý nghĩa 3%.

Lời giải

Gọi A là loại hàng, B là kho hàng. Với mức ý nghĩa $\alpha = 3\% = 0,03$ thì ta đặt giả thuyết kiểm định

$$\begin{cases} H_0 : A \text{ và } B \text{ độc lập} \\ H_1 : A \text{ và } B \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Kho / Chất lượng	Kho 1 (B_1)	Kho 2 (B_2)	Kho 3 (B_3)	Kho 4 (B_4)	n_i
Loại 1 (A_1)	112 (110, 7304)	124 (120, 2216)	106 (115, 9279)	119 (114, 1201)	461
Loại 2 (A_2)	213 (210, 4118)	211 (228, 4471)	227 (220, 2882)	225 (216, 8529)	876
Loại 3 (A_3)	165 (168, 8578)	197 (183, 3314)	180 (176, 7838)	161 (174, 027)	703
m_j	490	532	513	505	$n = 2040$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,3}$ và $\forall j = \overline{1,4}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=3}\sum_{j=1}^{k=4} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{112^2}{110,7304} + \frac{124^2}{120,2216} + \cdots + \frac{180^2}{176,7838} + \frac{161^2}{174,027}\right) - 2040 = 5,208$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,03}\left((3-1)(4-1)\right) = \mathcal{X}^2_{0,03}(6) = 13,9676$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên không đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 3%, ta có thể khẳng định chất lượng sản phẩm và kho hàng là độc lập với nhau.

9 Đề thi kết thúc học phần học kỳ I, năm học 2023 - 2024 của các lớp học phần 64.KT-1, 64.KT-2, 64.KT-3, 64.KTE, 64.QTKD (ngày 08/01/2024 - mã đề 1)

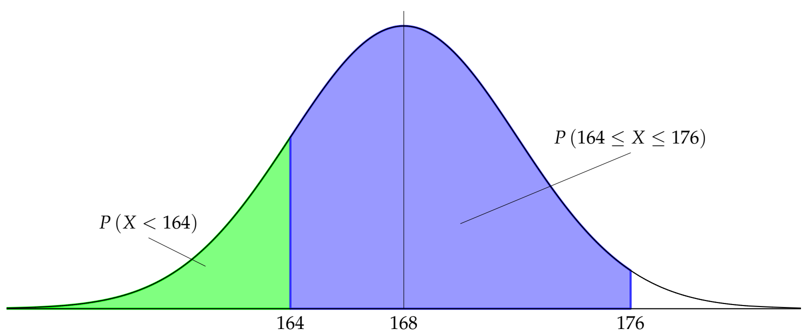
Câu 1 (3 điểm). Gọi X (cm) là chiều cao của nam giới Việt Nam là đại lượng ngẫu nhiên có phân phối chuẩn với trung bình là $\mu = 168$ (cm) và độ lệch chuẩn $\sigma = 4$ (cm).
a) Tính $P(X < 164)$, $P(164 \le X \le 176)$. Minh họa các kết quả bằng hình vẽ.
b) Một công ty bảo hiểm xe máy có 10.000 khách hàng. Mỗi chủ xe máy phải mua một bảo hiểm với mức giá 120.000 đồng/năm. Nếu xe bị tai nạn giao thông thì khách hàng sẽ được bồi thường trung bình 1.000.000 đồng. Theo thống kê của Ủy ban An toàn giao thông, tỉ lệ xe máy bị tai nạn giao thông trong một năm là $0,006$. Hỏi số tiền lời kì vọng của công ty thu được trong một năm (chưa tính chi phí tiền lương cho nhân viên, chi phí mặt bằng và các chi phí khác).

Lời giải

a) Theo đề bài thì ta có $X \sim \mathcal{N}\left(168; 4^2\right)$. Suy ra

- $P(X < 164) = \phi\left(\frac{164-168}{4}\right) - \phi(-\infty) = \phi(-1) - \phi(-\infty) = -0,3413 + 0,5 = 0,1587$.
- $P(164 \le X \le 176) = \phi\left(\frac{176-168}{4}\right) - \phi\left(\frac{164-168}{4}\right) = \phi(2) - \phi(-1) = \phi(2) + \phi(1) = 0,8185$.

Hình vẽ minh họa



b) Gọi Y là số lượng vụ tai nạn giao thông và T số tiền lời của công ty sau một năm bán bảo hiểm thì ta có $Y \sim \mathcal{B}(10000; 0,006)$ và $T = 10000 \cdot 120000 - Y \cdot 1000000$. Suy ra

$$E[T] = 10000 \cdot 120000 - 1000000 \cdot E[Y] = 10000 \cdot 120000 - 1000000 \cdot 10000 \cdot 0,006 = 1140000000$$

Vậy số tiền lời kì vọng của công ty thu được trong một năm là khoảng 1 tỷ 140 triệu đồng.

Câu 2 (4 điểm). Để khảo sát mức chi tiêu cho giáo dục của các gia đình Việt Nam năm 2023 (đơn vị: triệu/năm), tiến hành điều tra ngẫu nhiên 200 hộ gia đình và tính toán thì ta thu được số liệu sau: $\overline{x} = 16,05$ (tr), $s = 0,75$ (tr) và có 10 gia đình chi trên 100 (triệu/năm) cho giáo dục.
a) Với độ tin cậy 95%, hãy ước lượng mức chi tiêu cho giáo dục trung bình/năm của mỗi gia đình Việt Nam năm 2023.
b) Với độ tin cậy 96%, hãy ước lượng tỉ lệ hộ gia đình Việt Nam chi trên 100 (triệu/năm) cho giáo dục.
c) Với khoảng ước lượng ở câu b, để độ tin cậy là 98% thì cần khảo sát thêm bao nhiêu gia đình nữa.

Lời giải

a) Gọi X (triệu đồng) là mức chi tiêu cho giáo dục của mỗi gia đình Việt Nam năm 2023 và μ là mức chi tiêu trung bình. Từ đề bài, ta có $n = 200$, $\overline{x} = 16,05$ và $s = 0,75$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{0,75}{\sqrt{200}} = 0,1039$$

Khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (16,05 - 0,1039; 16,05 + 0,1039) = (15,9461; 16,1539)$$

Vậy mức chi tiêu cho giáo dục trung bình/năm của mỗi gia đình Việt Nam năm 2023 với độ tin cậy 95% là khoảng từ $15,9461$ đến $16,1539$ (triệu đồng).
b) Gọi p là tỉ lệ hộ gia đình Việt Nam chi trên 100 (triệu/năm) cho giáo dục. Từ đề bài, ta có $n = 200$ và tỉ lệ mẫu $f = \dfrac{10}{200} = 0,05$ thỏa $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 96\% = 0,96 \Leftrightarrow \alpha = 0,04 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,04} = 2,0537$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 2,0537 \cdot \sqrt{\frac{0,05 \cdot 0,95}{200}} = 0,0316$$

Suy ra khoảng tin cậy 96% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,05 - 0,0316; 0,05 + 0,0316) = (0,0184; 0,0816)$$

Vậy tỉ lệ hộ gia đình Việt Nam chi trên 100 (triệu/năm) cho giáo dục với độ tin cậy 96% là khoảng từ $1,84\%$ đến $8,16\%$.
c) Độ tin cậy

$$1 - \alpha = 98\% = 0,98 \Leftrightarrow \alpha = 0,02 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,01} = 2,3263$$

Với độ chính xác giữ nguyên thì

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} \Leftrightarrow n_1 = Z^2_{\frac{\alpha}{2}} \cdot \frac{f(1-f)}{\varepsilon^2} = 2,3263^2 \cdot \frac{0,05 \cdot 0,95}{0,0316^2} = 257,4251$$

Vậy cần khảo sát thêm $n_1 - n = 258 - 200 = 58$ gia đình.

Câu 3 (3 điểm). Điều tra ngẫu nhiên thu nhập (triệu đồng/năm) của 400 nhân viên kế toán làm việc ở Nha Trang và Đà Lạt, ta thu được kết quả sau

Thành phố \ Thu nhập	Dưới 90	90 − 130	Trên 130
Nha Trang	30	75	40
Đà Lạt	70	125	60

Với mức ý nghĩa 1%, hãy kết luận xem thu nhập của nhân viên kế toán có phụ thuộc vào thành phố mà họ làm việc không. Nêu ý nghĩa của miền bác bỏ của bài toán này.

Lời giải

Gọi X là thu nhập của nhân viên kế toán, Y là thành phố mà họ làm việc. Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta đặt giả thuyết kiểm định

$$\begin{cases} H_0 : X \text{ và } Y \text{ độc lập} \\ H_1 : X \text{ và } Y \text{ có liên quan} \end{cases}$$

Ta tính các tần số lý thuyết $n\widehat{P}_{ij}$ và viết vào bảng trên như sau

Thành phố \ Thu nhập	Dưới 90	90 − 130	Trên 130	m_j
Nha Trang	30 (36, 25)	75 (72, 5)	40 (36, 25)	145
Đà Lạt	70 (63, 75)	125 (127, 5)	60 (63, 75)	255
n_i	100	200	100	$n = 400$

Từ bảng trên ta thấy $n\widehat{P}_{ij} > 5, \forall i = \overline{1,2}$ và $\forall j = \overline{1,3}$. Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{h=2}\sum_{j=1}^{k=3} \frac{n_{ij}^2}{n\widehat{P}_{ij}} - n = \left(\frac{30^2}{36,25} + \frac{75^2}{72,5} + \frac{40^2}{36,25} + \frac{70^2}{63,75} + \frac{125^2}{127,5} + \frac{60^2}{63,75}\right) - 400 = 2,4341$$

Ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha\left((h-1)(k-1)\right) = \mathcal{X}^2_{0,01}\left((2-1)(3-1)\right) = \mathcal{X}^2_{0,01}(2) = 9,2103$$

So sánh thì ta có $\mathcal{X}^2 < \mathcal{X}^2_\alpha\left((h-1)(k-1)\right)$ nên không đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 1%, ta có thể khẳng định thu nhập của nhân viên kế toán không phụ thuộc vào thành phố mà họ làm việc.

Chương VII

Đề thi tham khảo từ các trường khác

1 Mã đề 2 - đề thi kết thúc học phần học kỳ I, năm học 2023 - 2024 của lớp học phần 2311MATH1703, Trường Đại học Sư phạm Thành phố Hồ Chí Minh

Câu 1 (3 điểm). Giả sử một khoa ở một trường đại học có 60% sinh viên nam và 40% sinh viên nữ. Số sinh viên thường trú ở TP.HCM chiếm tỉ lệ 70% trong tổng số sinh viên nam và chiếm tỉ lệ 30% trong tổng số sinh viên nữ.
a) Chọn ngẫu nhiên một sinh viên của khoa. Tính xác suất để chọn được một sinh viên thường trú ở TP.HCM.
b) Nếu biết rằng sinh viên vừa chọn thường trú ở TP.HCM thì xác suất để sinh viên đó là nam bằng bao nhiêu.
c) Chọn ngẫu nhiên hai sinh viên khoa này. Tính xác suất để có ít nhất một sinh viên thường trú ở TP.HCM, biết rằng khoa này có tất cả 300 sinh viên.

Lời giải

Phép thử: Chọn ngẫu nhiên một sinh viên của khoa.
Gọi N_1 là biến cố "Sinh viên được chọn là nam" và N_2 là biến cố "Sinh viên được chọn là nữ" thì $\{N_1, N_2\}$ tạo thành hệ biến cố đầy đủ với $P(N_1) = 0,6$ và $P(N_2) = 0,4$.
Gọi T là biến cố "Sinh viên được chọn thường trú ở TP.HCM" thì ta có $P(T|N_1) = 0,7$ và $P(T|N_2) = 0,3$. Suy ra

$$P(T) = P(N_1) \cdot P(T|N_1) + P(N_2) \cdot P(T|N_2) = 0,6 \cdot 0,7 + 0,4 \cdot 0,3 = 0,54$$

b) Ta có

$$P(N_1|T) = \frac{P(N_1) \cdot P(T|N_1)}{P(T)} = \frac{0,6 \cdot 0,7}{0,54} = \frac{7}{9}$$

c) Trong 300 sinh viên của khoa thì có $300 \cdot 0,54 = 162$ sinh viên thường trú ở TP.HCM.
Phép thử: Chọn ngẫu nhiên hai sinh viên khoa này. Gọi X là biến cố "Có ít nhất một sinh viên thường trú ở TP.HCM" thì $\overline{X}$ là biến cố "Không có ai thường trú ở TP.HCM", tức ta có

$$P(X) = 1 - P\left(\overline{X}\right) = 1 - \frac{C^2_{300-162}}{C^2_{300}} = 1 - \frac{C^2_{138}}{C^2_{300}} = 0,7892$$

Câu 2 (2 điểm). Gọi X là tuổi thọ của con người. Giả sử hàm mật độ xác suất của X là

$$f(x) = \begin{cases} cx^2(100-x)^2 & \text{nếu } 0 \le x \le 100 \\ 0 & \text{nếu } x < 0 \text{ hay } x > 100 \end{cases}$$

a) Xác định hằng số c.
b) Tính xác suất để một người sống hơn 60 năm, biết rằng hiện tại họ đã 50 tuổi.

Lời giải

a) Vì $f(x)$ là hàm mật độ xác suất nên ta có

$$\begin{cases} cx^2(100-x)^2 \geq 0, \forall x \in [0,100] \\ \int\limits_{-\infty}^{+\infty} cx^2(100-x)^2 \mathrm{d}x = 1 \end{cases} \Rightarrow \begin{cases} c \geq 0 \\ \int\limits_{0}^{100} cx^2(100-x)^2 \mathrm{d}x = 1 \end{cases} \Rightarrow c = 3 \cdot 10^{-9}$$

b) Ta có hàm phân phối xác suất là

$$F(x) = \int\limits_{-\infty}^{x} f(t)\mathrm{d}t = \begin{cases} 0 & x < 0 \\ 3 \cdot 10^{-9} \left(\dfrac{x^5}{5} - 50x^4 + \dfrac{10^4}{3} x^3 \right) & 0 \leq x \leq 100 \\ 1 & x > 100 \end{cases}$$

Khi đó ta có

$$P(X \geq 60 | X \geq 50) = \frac{P[(X \geq 60)(X \geq 50)]}{P(X \geq 50)} = \frac{P(X \geq 60)}{P(X \geq 50)} = \frac{1 - F(60)}{1 - F(50)} = \frac{1 - \frac{2133}{3125}}{1 - \frac{1}{2}} = \frac{1984}{3125} = 0,63488$$

Câu 3 (5 điểm). Để khảo sát doanh thu một làng nghề truyền thống làm nón lá tại địa phương B, người ta điều tra ngẫu nhiên 100 hộ gia đình sản xuất và kinh doanh loại mặt hàng này trong một tháng năm 2023 và thu được bảng số liệu

Doanh thu (triệu đồng)	20	24	28	32	36	40	44	48	52
Số hộ gia đình	5	10	17	25	20	10	8	3	2

a) Tính trung bình mẫu và độ lệch chuẩn hiệu chỉnh mẫu từ bảng dữ liệu trên.
b) Với độ tin cậy 95%, hãy ước lượng doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng nói trên. Để độ chính xác của ước lượng nhỏ hơn 2 triệu đồng thì cần điều tra ít nhất bao nhiêu hộ.
c) Theo số liệu điều tra năm 2022 thì tỉ lệ những hộ gia đình đạt doanh thu dưới 28 triệu đồng là 20%. Theo anh/chị, tỉ lệ này có giảm đi trong năm 2023 hay không. Hãy kết luận với mức ý nghĩa 1%.
d) Theo điều tra cách đây 2 năm thì doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng này là 30 triệu đồng/tháng. Hãy đánh giá xem doanh thu trung bình sau 2 năm có thay đổi không với mức ý nghĩa 5%.
e) Điều tra doanh thu của 200 hộ gia đình kinh doanh loại mặt hàng trên ở địa phương C năm 2023, người ta tính được doanh thu trung bình/tháng là 37 triệu đồng và độ lệch chuẩn hiệu chỉnh mẫu là 1, 1 triệu đồng. Doanh thu trung bình loại mặt hàng này ở địa phương C và B có như nhau hay không. Hãy kết luận với độ tin cậy 95%.

Lời giải

a) Gọi X (triệu đồng) là doanh thu của các hộ gia đình kinh doanh loại mặt hàng trên ở địa phương B. Trung bình mẫu của X là

$$\overline{x} = \frac{20 \cdot 5 + 24 \cdot 10 + 28 \cdot 17 + 32 \cdot 25 + 36 \cdot 20 + 40 \cdot 10 + 44 \cdot 8 + 48 \cdot 3 + 52 \cdot 2}{100} = 33,36$$

Độ lệch chuẩn mẫu hiệu chỉnh của X là

$$s_X = \sqrt{\frac{n}{n-1}\left[\overline{x^2} - (\overline{x})^2\right]} = \sqrt{\frac{100}{99} \cdot \left[1164,16 - (33,36)^2\right]} = 7,1964$$

b) Gọi μ là doanh thu trung bình.
• Vì $n = 100 > 30$ nên ta có độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Độ chính xác

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s_X}{\sqrt{n_X}} = 1,96 \cdot \frac{7,1964}{\sqrt{100}} = 1,4105$$

Khoảng tin cậy 95% của μ là

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (33,36 - 1,4105; 33,36 + 1,4105) = (31,9495; 34,7705)$$

Vậy doanh thu trung bình của các hộ gia đình kinh doanh loại mặt hàng nói trên với độ tin cậy 95% là khoảng từ 31,9495 đến 34,7705 (triệu đồng).

• Ta có độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s_X}{\sqrt{n_X}} = 1,96 \cdot \frac{7,1964}{\sqrt{n_X}} = \frac{14,105}{\sqrt{n_X}}$$

Theo đề bài, ta có

$$\varepsilon < 2 \Leftrightarrow \frac{14,105}{\sqrt{n_X}} \Leftrightarrow n_X > \left(\frac{14,105}{2}\right)^2 = 49,7378$$

Vậy cần điều tra ít nhất 50 hộ.

c) Gọi p là tỉ lệ những hộ gia đình đạt doanh thu dưới 28 triệu đồng năm 2023. Giả thuyết kiểm định

$$\begin{cases} H_0 : p = 0,2 \\ H_1 : p < 0,2 \end{cases}$$

Từ đề bài, ta có cỡ mẫu $n_X = 100$, tỉ lệ mẫu $f = \dfrac{15}{100} = 0,15$ và $p_0 = 0,2$ thỏa $\begin{cases} n_X p_0 > 5 \\ n_X(1-p_0) > 5 \end{cases}$. Tiêu chuẩn kiểm định

$$Z = \frac{f - p_0}{\sqrt{p_0(1-p_0)/\sqrt{n_X}}} = \frac{0,15 - 0,2}{\sqrt{0,2 \cdot 0,8/\sqrt{100}}} = -1,25$$

Với mức ý nghĩa $\alpha = 1\% = 0,01$ thì ta sẽ bác bỏ H_0 khi

$$Z < -Z_\alpha = -Z_{0,01} = -2,3263$$

So sánh thì ta có $Z > -Z_\alpha$ nên không đủ cơ sở để bác bỏ H_0.Vây ta không đủ cơ sở để khẳng định tỉ lệ những hộ gia đình đạt doanh thu dưới 28 triệu đồng giảm đi vào năm 2023 với mức ý nghĩa 1%.

d) Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$$

Từ câu a, ta có $n_X = 100$, $\overline{x} = 33,36$ và $s_X = 7,1964$. Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \mu_0}{s_X/\sqrt{n_X}} = \frac{33,36 - 30}{7,1964/\sqrt{100}} = 4,669$$

Ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

So sánh thì ta có $|Z| > Z_{\frac{\alpha}{2}}$ nên đủ cơ sở để bác bỏ H_0 và nhận H_1.Vậy với mức ý nghĩa 5%, ta đủ cơ sở để khẳng định doanh thu trung bình sau 2 năm có thay đổi.

e) Gọi Y (triệu đồng) là doanh thu của các hộ gia đình kinh doanh loại mặt hàng trên ở địa phương C. Giả thuyết kiểm định

$$\begin{cases} H_0 : E(X) = E(Y) \\ H_1 : E(X) \neq E(Y) \end{cases}$$

Từ đề bài, ta có $\overline{x} = 33,36$; $s_X = 7,1964$; $n_X = 100$ và $\overline{y} = 37$; $s_Y = 1,1$; $n_Y = 200$. Tiêu chuẩn kiểm định

$$Z = \frac{\overline{x} - \overline{y}}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}} = \frac{33,36 - 37}{\sqrt{(7,1964)^2/100 + (1,1)^2/200}} = -5,0288$$

Với mức ý nghĩa $\alpha = 1 - 95\% = 0,05$ thì ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

So sánh thì ta có $|Z| > Z_{\frac{\alpha}{2}}$ nên đủ cơ sở để bác bỏ H_0.Vậy đủ cơ sở để khẳng định doanh thu trung bình loại mặt hàng này ở địa phương C và B khác nhau với độ tin cậy 95%.

2 Mã đề 2 - đề thi kết thúc học phần học kỳ I năm học 2022 - 2023, mã học phần MAT1101 của trường Đại học Khoa học Tự nhiên - Đại học Quốc gia Hà Nội

Câu 1. Vòng bụng của nam giới $20-29$ tuổi có phân phối xấp xỉ chuẩn, với trung bình $92,5\ cm$ và độ lệch chuẩn là $13,7\ cm$. Sử dụng mô hình phân phối chuẩn để
a) Xác định tỷ lệ nam giới từ 20 đến 29 tuổi có vòng bụng nhỏ hơn $100\ cm$.
b) Xác định ngưỡng dưới của vòng bụng đại diện cho trên 90% tất cả các vòng bụng của nam giới từ 20 đến 29 tuổi.
c) Xác định khoảng giá trị đối xứng quanh $92,5\ cm$ của vòng bụng sao cho 90% tất cả các vòng bụng của nam giới độ tuổi $20-29$ nằm trong đó.

Lời giải

a) Gọi $X\ (cm)$ là vòng bụng của nam giới $20-29$ tuổi thì ta có $X \sim \mathcal{N}\left(92,5;(13,7)^2\right)$. Suy ra

$$P(X<100)=\phi\left(\frac{100-92,5}{13,7}\right)-\phi(-\infty)=\phi(0,5474)-\phi(-\infty)=0,20795+0,5=0,70795$$

b) Gọi $a\ (cm)$ là ngưỡng dưới của vòng bụng. Khi đó ta có

$$P(X\geq a)=\phi(+\infty)-\phi\left(\frac{a-92,5}{13,7}\right)=0,5-\phi\left(\frac{a-92,5}{13,7}\right)=0,9$$

$$\Leftrightarrow \phi\left(\frac{a-92,5}{13,7}\right)=-0,4=\phi(-1,28155)\Leftrightarrow \frac{a-92,5}{13,7}=-1,28155\Leftrightarrow a=74,942765$$

Vậy ngưỡng dưới của vòng bụng đại diện cho trên 90% tất cả các vòng bụng của nam giới từ 20 đến 29 tuổi là $74,942765\ cm$.
c) Gọi khoảng giá trị đối xứng quanh $92,5\ cm$ của vòng bụng là $92,5\pm\varepsilon\ (cm)$. Khi đó ta có

$$P(92,5-\varepsilon\leq X\leq 92,5+\varepsilon)=P(|X-\mu|\leq\varepsilon)=2\phi\left(\frac{\varepsilon}{13,7}\right)=0,9$$

$$\Leftrightarrow \phi\left(\frac{\varepsilon}{13,7}\right)=0,45=\phi(1,64485)\Leftrightarrow \frac{\varepsilon}{13,7}=1,64485\Leftrightarrow \varepsilon=22,534445$$

Vậy khoảng giá trị đối xứng quanh $92,5\ cm$ của vòng bụng sao cho 90% tất cả các vòng bụng của nam giới độ tuổi $20-29$ nằm trong đó là $92,5\pm 22,534445\ cm$.

Câu 2. Theo lý thuyết di truyền, tỉ lệ sinh con trai và con gái là 1 : 1. Giả sử các lần sinh là độc lập.
a) Gọi X là số con gái trong một gia đình 3 con. Tìm phân phối xác suất và các đặc trưng của X.
b) Quan sát thực tế trên 90 gia đình có 3 con thu được kết quả sau

Số bé gái	0	1	2	3
Số gia đình	20	40	25	5

Với mức ý nghĩa 5%, hãy kiểm tra xem dữ liệu quan sát thực tế có theo phân phối của X ở câu a hay không.

Lời giải

a) Do tỉ lệ sinh con gái là $0,5$ và các lần sinh độc lập với nhau nên ta có $X\sim\mathcal{B}(3;0,5)$. Các đặc trưng của X gồm:
Kỳ vọng mẫu: $E(X)=3\cdot 0,5=1,5$.
Phương sai mẫu: $D(X)=3\cdot 0,5\cdot(1-0,5)=0,75$.
Độ lệch mẫu: $\sigma(X)=\sqrt{D(X)}=0,866$.
Giá trị tin chắc nhất: $\text{Mod}(X)=1$ hoặc 2 vì $P(X=1)=P(X=2)=0,375$.
b) Giả thuyết kiểm định

$$\begin{cases} H_0: X \text{ có phân phối nhị thức} \\ H_1: X \text{ không có phân phối nhị thức} \end{cases}$$

Giả sử H_0 đúng thì ta có $X \sim \mathcal{B}(3; 0,5)$. Khi đó ta có bảng sau

Số bé gái	Xác suất (P_i)	Tần số lý thuyết (nP_i)	Số gia đình (n_i)
0	$0,125$	$11,25$	20
1	$0,375$	$33,75$	40
2	$0,375$	$33,75$	25
3	$0,125$	$11,25$	5
Tổng	1	90	$n = 90$

Tiêu chuẩn kiểm định

$$\mathcal{X}^2 = \sum_{i=1}^{4} \frac{n_i^2}{nP_i} - n = \left(\frac{20^2}{11,25} + \frac{40^2}{33,75} + \frac{25^2}{33,75} + \frac{5^2}{11,25} \right) - 90 = 13,7037$$

Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta sẽ bác bỏ H_0 khi

$$\mathcal{X}^2 > \mathcal{X}^2_\alpha (k - r - 1) = \mathcal{X}^2_{0,05} (4 - 1 - 1) = \mathcal{X}^2_{0,05} (2) = 5,991$$

So sánh thì ta có $\mathcal{X}^2 > \mathcal{X}^2_\alpha (k - r - 1)$ nên đủ cơ sở để bác bỏ H_0. Vậy với mức ý nghĩa 5%, ta có thể khẳng định dữ liệu quan sát thực tế không theo phân phối nhị thức của X ở câu a.

Câu 3. Giả sử rằng lực ném của bàn tay thuận của một người có phân phối chuẩn. Dữ liệu về lực ném bàn tay thuận (y) và độ tuổi (x) của một số người được thu thập như sau

x	15	16	28	61	53	43	16	25	28	34	37	41	43	49	53	61	68
y	65	60	58	60	46	66	56	75	46	45	58	70	73	45	60	56	30
Số người	2	4	5	3	3	4	4	5	2	3	1	2	5	4	6	2	4

a) Hãy ước lượng khoảng tin cậy 90% cho trung bình lực nắm bàn tay thuận ở người. Kiểm tra xem trung bình lực nắm bàn tay thuận ở người có bằng 65 hay không với mức ý nghĩa 5%.
b) Hãy so sánh lực nắm bàn tay thuận ở người trước 40 tuổi có nhỏ hơn sau 40 tuổi hay không với mức ý nghĩa 5%.
c) Một bác sĩ muốn xác định xem có mối quan hệ nào giữa tuổi và lực nắm bàn tay thuận của người hay không. Hãy tính toán để đưa ra câu trả lời cho bác sĩ bằng cách tìm hệ số tương quan mẫu giữa hai đại lượng trên và tìm đường hồi quy dự báo lực nắm bàn tay thuận theo độ tuổi. Tính sai số tiêu chuẩn của đường hồi quy dự báo. Dự báo xem lực nắm bàn tay thuận của người đó ở độ tuổi 65.

Lời giải

a) Gọi μ_Y là lực nắm trung bình bàn tay thuận ở người. Từ đề bài, ta có $n = 59, \overline{y} = 57,678$ và $s_Y = 11,9776$.
Độ tin cậy

$$1 - \alpha = 90\% = 0,9 \Leftrightarrow \alpha = 0,1 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,05} = 1,6449$$

Suy ra độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \frac{s_Y}{\sqrt{n}} = 1,6449 \cdot \frac{11,9776}{\sqrt{59}} = 2,565$$

Khoảng tin cậy 90% của μ_y là

$$(\overline{y} - \varepsilon; \overline{y} + \varepsilon) = (57,678 - 2,565; 57,678 + 2,565) = (55,113; 60,243)$$

Vậy lực nắm bàn tay thuận ở người với độ tin cậy 90% là khoảng từ $55,113$ đến $60,243$.
Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta đặt giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \mu_Y = 65 \\ H_1 : \mu_Y \neq 65 \end{cases}$$

Tiêu chuẩn kiểm định

$$Z = \frac{\overline{y} - \mu_0}{s_Y / \sqrt{n}} = \frac{57,678 - 65}{11,9776/\sqrt{59}} = -4,6955$$

Ta sẽ bác bỏ H_0 khi

$$|Z| > Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

So sánh thì ta có $|Z| > Z_{\frac{\alpha}{2}}$ nên đủ cơ sở để bác bỏ H_0 và nhận khác 65.
b) Gọi Y_1, Y_2 lần lượt là lực nắm bàn tay thuận ở người trước 40 tuổi và sau 40 tuổi. Gọi μ_1, μ_2 lần lượt là trung bình lực nắm bàn tay thuận ở người trước 40 tuổi và sau 40 tuổi. Giả thuyết kiểm định

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{cases}$$

Từ bảng dữ liệu trên đề bài, ta tính được $n_1 = \overline{x} = 3,6677$; $s_X = 0,2979$; $\overline{y} = 3,2692$; $s_Y = 0,4429$ và $n_1 = n_2 = 13$. Suy ra

$$s^2 = \frac{(n_1-1)s_X^2 + (n_2-1)s_Y^2}{n_1+n_2-2} = \frac{12 \cdot 0,2979^2 + 12 \cdot 0,4429^2}{24} = 0,1424$$

Tiêu chuẩn kiểm định

$$T = \frac{\overline{x} - \overline{y}}{\sqrt{s^2\left(1/n_1 + 1/n_2\right)}} = \frac{3,6677 - 3,2692}{\sqrt{0,1424\left(1/13 + 1/13\right)}} = 2,6923$$

Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta sẽ bác bỏ H_0 khi

$$T < -t_{\alpha}^{n_1+n_2-2} = -t_{0,05}^{24} = -1,7109$$

So sánh thì $T > -t_{\alpha}^{n_1+n_2-2}$ nên không đủ cơ sở để bác bỏ H_0.Vây không thể khẳng định trẻ sơ sinh của nhóm mẹ nghiện thuốc lá nhẹ cân hơn trẻ sơ sinh của nhóm mẹ không hút thuốc lá với mức ý nghĩa 5%.

3 Đề thi kết thúc học phần học kỳ 2023.1, mã học phần MI3180 của Đại học Bách khoa Hà Nội (ngày 03/02/2024 - đề 2)

Câu 1. *(2,0 điểm)* Cho biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} C(1+3x^2) & \text{nếu } x \in [2,6] \\ 0 & \text{nếu } x \notin [2,6] \end{cases}$$

a) Hãy tìm hằng số C và tính giá trị kỳ vọng của X.
b) Tính xác suất để trong 8 lần quan sát biến X, có đúng 2 lần X nhận giá trị trong khoảng $(4,5)$.

Lời giải

a) Vì $f(x)$ là hàm mật độ xác suất nên ta có

$$\begin{cases} C(1+3x^2) \geq 0, \forall x \in [2,6] \\ \int\limits_{-\infty}^{+\infty} C(1+3x^2)\mathrm{d}x = 1 \end{cases} \Rightarrow \begin{cases} C \geq 0 \\ \int\limits_{2}^{6} C(1+3x^2)\mathrm{d}x = 1 \end{cases} \Rightarrow C = \frac{1}{212}$$

Từ đây ta có hàm phân phối xác suất của X là

$$F(x) = \int\limits_{-\infty}^{x} f(t)\mathrm{d}t = \begin{cases} 0 & x < 2 \\ \dfrac{x^3+x-10}{212} & 2 \leq x \leq 6 \\ 1 & x > 6 \end{cases}$$

Khi đó ta có

$$E[X] = \int\limits_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \frac{1}{212}\int\limits_{2}^{6}\left(x+3x^3\right)\mathrm{d}x = \frac{244}{53} \approx 4,6038$$

b) Xác suất để X nhận giá trị trong khoảng $(4,5)$ là

$$P(4 < X < 5) = F(5) - F(4) = \frac{30}{53} - \frac{29}{106} = \frac{31}{106} \approx 0,29245$$

Gọi Y là số lần X nhận giá trị trong khoảng $(4,5)$ thì ta có $Y \sim \mathcal{B}(8; 0,29245)$. Suy ra

$$P(Y = 2) = C_8^2 \cdot (0,29245)^2 \cdot (1 - 0,29245)^6 \approx 0,3005$$

Câu 2. *(1,5 điểm)* Để ước lượng tỉ lệ vòng bi trục khuỷu động cơ ô tô có độ nhám bề mặt vượt quá thông số kỹ thuật cho phép, người ta quan sát ngẫu nhiên 150 vòng bi thì thấy có 12 vòng bi có độ nhám bề mặt vượt quá thông số kỹ thuật. Từ số liệu trên, hãy tìm khoảng tin cậy đối xứng cho tỉ lệ vòng bi có độ nhám bề mặt vượt quá thông số kỹ thuật với độ tin cậy 95%.

Lời giải

Gọi p là tỉ lệ vòng bi có độ nhám bề mặt vượt quá thông số kỹ thuật. Từ đề bài, ta tính được cỡ mẫu $n = 150$ và tỉ lệ mẫu $f = \dfrac{12}{150} = 0,08$ thỏa điều kiện $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$. Độ tin cậy

$$1 - \alpha = 95\% = 0,95 \Leftrightarrow \alpha = 0,05 \Leftrightarrow Z_{\frac{\alpha}{2}} = Z_{0,025} = 1,96$$

Từ đây ta tính được độ chính xác là

$$\varepsilon = Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{f(1-f)}{n}} = 1,96 \cdot \sqrt{\frac{0,08 \cdot 0,92}{150}} = 0,0434$$

Suy ra khoảng tin cậy 95% của p là

$$(f - \varepsilon; f + \varepsilon) = (0,08 - 0,0434; 0,08 + 0,0434) = (0,0366; 0,1234)$$

Vậy tỉ lệ vòng bi có độ nhám bề mặt vượt quá thông số kỹ thuật với độ tin cậy 95% là khoảng từ $3,66\%$ đến $12,34\%$.

Câu 3. *(2,0 điểm)* Để so sánh đường kính của các thanh thép được sản xuất bởi hai máy khác nhau, người ta lấy ngẫu nhiên gồm 2 mẫu độc lập gồm 17 thanh sản xuất bởi máy I và 15 thanh sản xuất bởi máy II thì thu được trung bình mẫu và phương sai mẫu lần lượt như sau: $\overline{x}_1 = 8,37$; $\overline{x}_2 = 8,65$; $s_1^2 = 0,36$; $s_2^2 = 0,39$. Giả sử đường kính của các thanh được sản xuất bởi 2 máy có phân phối chuẩn với phương sai bằng nhau. Ở mức ý nghĩa 5%, có thể cho rằng các thanh thép được sản xuất bởi hai máy có đường kính trung bình khác nhau hay không.

Lời giải

Gọi X_1, X_2 lần lượt là đường kính các thanh thép được sản xuất bởi máy I và máy II. Giả thiết kiểm định

$$\begin{cases} H_0 : E(X_1) = E(X_2) \\ H_1 : E(X_1) \neq E(X_2) \end{cases}$$

Từ đề bài, ta có $n_1 = 17$; $\overline{x_1} = 8,37$; $s_1^2 = 0,36$ và $n_2 = 15$; $\overline{x_2} = 8,65$, $s_2^2 = 0,39$. Suy ra

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{16 \cdot 0,36 + 14 \cdot 0,39}{30} = 0,374$$

Tiêu chuẩn kiểm định

$$\mathrm{T} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2\,(1/n_1 + 1/n_2)}} = \frac{8,37 - 8,65}{\sqrt{0,374\,(1/17 + 1/15)}} = -1,2925$$

Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta sẽ bác bỏ H_0 khi

$$|\mathrm{T}| > \mathrm{t}_{\frac{\alpha}{2}}^{n_1+n_2-2} = \mathrm{t}_{0,025}^{30} = 2,0423$$

So sánh thì ta có $|\mathrm{T}| < t_{\frac{\alpha}{2}}^{n_1+n_2-2}$ nên không đủ cơ sở để bác bỏ H_0.Vậy không đủ cơ sở để khẳng định rằng các thanh thép được sản xuất bởi hai máy có đường kính trung bình khác nhau với mức ý nghĩa 5%.

Câu 4. *(4,5 điểm)* Một kỹ sư hóa học nghiên cứu mối quan hệ giữa mức độ chuyển hóa của một sản phẩm từ nguyên liệu thô (Y) phụ thuộc vào nhiệt độ phản ứng (Z_1) và thời gian phản ứng (Z_2). Xét miền quy hoạch của 2 biến độc lập $\begin{cases} 26 \le Z_1 \le 30 \\ 2 \le Z_2 \le 8 \end{cases}$ và áp dụng phương pháp quy hoạch trực giao cấp 2 với các tham số mô hình như sau

$$k = 2; n_0 = 5; N = 2^k + n_0 + 2k = 13; a = \sqrt{\sqrt{N \cdot 2^{k-2}} - 2^{k-1}}; \lambda = \frac{2^k + 2a^2}{N}$$

Thực hiện các phép đổi biến $X_j = \dfrac{Z_j - Z_j^0}{\Delta Z_j}$ và $X_j' = X_j^2 - \lambda$ với $j = 1; 2$. Xét mô hình

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1' + \beta_{22} X_2' + \xi$$

a) Hoàn thiện bảng thực nghiệm trực giao cấp 2 sau

N_0	Z_1	Z_2	X_0	X_1	X_2	X_1X_2	X_1'	X_2'	Y
1	26	2	+						164,3
2	26	8	+						337,1
3	30	2	+						204,6
4	30	8	+						456,2
5	28	5	+						293,6
6	28	5	+						291,5
7	28	5	+						294,2
8	28	5	+						292,7
9	28	5	+						293,5
10			+	a	0				339,4
11			+	$-a$	0				236,3
12			+	0	a				428,2
13			+	0	$-a$				161,8

b) Xác định các hệ số hồi quy của phương trình hồi quy thực nghiệm.
c) Kiểm định sự có nghĩa của các hệ số hồi quy với mức ý nghĩa 5%.

Lời giải

a) Từ đề bài thì ta tính được $a = \sqrt{\sqrt{13} - 2}$ và $\lambda = \dfrac{2}{\sqrt{13}}$. Suy ra bảng thực nghiệm đã cho trở thành

N_0	X_0	X_1	X_2	X_1X_2	X_1'	X_2'	Y
1	+	−	−	+	0,4453	0,4453	164,3
2	+	−	+	−	0,4453	0,4453	337,1
3	+	+	−	−	0,4453	0,4453	204,6
4	+	+	+	+	0,4453	0,4453	456,2
5	+	0	0	0	−0,5547	−0,5547	293,6
6	+	0	0	0	−0,5547	−0,5547	291,5
7	+	0	0	0	−0,5547	−0,5547	294,2
8	+	0	0	0	−0,5547	−0,5547	292,7
9	+	0	0	0	−0,5547	−0,5547	293,5
10	+	1,2671	0	0	1,0508	−0,5547	339,4
11	+	−1,2671	0	0	1,0508	−0,5547	236,3
12	+	0	1,2671	0	−0,5547	1,0508	428,2
13	+	0	−1,2671	0	−0,5547	1,0508	161,8

b) Ta có

$$\beta_0 = \frac{\sum\limits_{i=1}^{N} Y_i}{N} = \frac{164,3 + 337,1 + 204,6 + \cdots + 236,3 + 428,2 + 161,8}{13} = 291,8$$

$$\beta_1 = \frac{\sum\limits_{i=1}^{N} X_{1i} \cdot Y_i}{2^k + 2a^2} = \frac{-164,3 - 337,1 + 204,6 + 456,2 + 1,2671 \cdot 339,4 - 1,2671 \cdot 236,3}{2\sqrt{13}} = 40,221$$

$$\beta_2 = \frac{\sum\limits_{i=1}^{N} X_{2i} \cdot Y_i}{2^k + 2a^2} = \frac{-164,3 + 337,1 - 204,6 + 456,2 + 1,2671 \cdot 428,2 - 1,2671 \cdot 161,8}{2\sqrt{13}} = 105,6642$$

$$\beta_{12} = \frac{\sum\limits_{i=1}^{N} X_{1i}X_{2i}Y_i}{2^k} = \frac{164,3 - 337,1 - 204,6 + 456,2}{4} = 19,7$$

$$\beta_{11} = \frac{\sum\limits_{i=1}^{N} X'_{1i}Y_i}{2a^4} = \frac{0,4453 \cdot 164,3 + 0,4453 \cdot 337,1 + \cdots - 0,5547 \cdot 428,2 - 0,5547 \cdot 161,8}{34 - 8\sqrt{13}} = -3,4356$$

$$\beta_{22} = \frac{\sum\limits_{i=1}^{N} X'_{2i}Y_i}{2a^4} = \frac{0,4453 \cdot 164,3 + 0,4453 \cdot 337,1 + \cdots + 1,0508 \cdot 428,2 + 1,0508 \cdot 161,8}{34 - 8\sqrt{13}} = 1,0175$$

Vậy phương trình hồi quy cần tìm là

$$Y = 291,8 + 40,221X_1 + 105,6642X_2 + 19,7X_1X_2 - 3,4356X'_1 + 1,0175X'_2$$

c) Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì ta chọn cặp giả thuyết kiểm định như sau

$$\begin{cases} H_0 : \beta_\theta = 0 \\ H_1 : \beta_\theta \neq 0 \end{cases}$$

Ta xét các thí nghiệm tại tâm như sau

y_i	$\overline{y_0}$	$y_i - \overline{y_0}$
293,6	293,1	0,5
291,5		−1,6
294,2		1,1
292,7		−0,4
293,5		0,4

Suy ra phương sai tái sinh là

$$s_{ts} = \sqrt{\frac{1}{n_0 - 1}\sum_{i=1}^{n_0}(y_i - \overline{y_0})^2} = \sqrt{\frac{(0,5)^2 + (-1,6)^2 + (1,1)^2 + (-0,4)^2 + (0,4)^2}{4}} = 1,0416$$

Từ đây ta tính được các giá trị quan sát t_{β_θ} như sau

$$t_{\beta_0} = \frac{\beta_0}{s_{ts}/\sqrt{N}} = \frac{291,8}{1,0416/\sqrt{13}} = 1010,0805$$

$$t_{\beta_1} = \frac{\beta_1}{s_{ts}/\sqrt{2^k + 2a^2}} = \frac{40,221}{1,0416/\sqrt{2\sqrt{13}}} = 103,6978$$

$$t_{\beta_2} = \frac{\beta_2}{s_{ts}/\sqrt{2^k + 2a^2}} = \frac{105,6642}{1,0416/\sqrt{2\sqrt{13}}} = 272,4129$$

$$t_{\beta_{12}} = \frac{\beta_{12}}{s_{ts}/\sqrt{2^k}} = \frac{19,7}{1,0416/\sqrt{4}} = 37,8264$$

$$t_{\beta_{11}} = \frac{\beta_{11}}{s_{ts}/\sqrt{2a^4}} = \frac{-3,4356}{1,0416/\sqrt{34 - 8\sqrt{13}}} = -7,4893$$

$$t_{\beta_{22}} = \frac{\beta_{22}}{s_{ts}/\sqrt{2a^4}} = \frac{1,0175}{1,0416/\sqrt{34 - 8\sqrt{13}}} = 2,2181$$

Với mức ý nghĩa $\alpha = 5\% = 0,05$ thì hệ số hồi quy có nghĩa (khác 0) khi và chỉ khi

$$\left|t_{\beta_\theta}\right| > \mathrm{t}_{\frac{\alpha}{2}}^{n_0-1} = t_{0,025}^{4} = 2,7764$$

So sánh thì ta thấy chỉ có $\left|t_{\beta_{22}}\right| < \mathrm{t}_{\frac{\alpha}{2}}^{n_0-1}$ nên $b_{22} = 0$. Vậy phương trình hồi quy mới là

$$Y = 291,8 + 40,221X_1 + 105,6642X_2 + 19,7X_1X_2 - 3,4356X_1'$$