# BÀI TẬP XÁC SUẤT THỐNG KÊ

Bộ môn Đại số - Xác suất Thống kê

8-2017

**Yêu cầu chung**

1. Các bài tập trong tài liệu này dành cho sinh viên các ngành Công trình, Cơ khí, Công nghệ thông tin, học môn Xác suất thống kê với thời lượng 2 tín chỉ.

2. Sinh viên chuẩn bị các bài tập sau đây cho các buổi học bài tập theo yêu cầu của giảng viên. Việc tổ chức kiểm tra lấy điểm quá trình sẽ do giảng viên thông báo với lớp.

**I. Phần đề bài các bài tập**

## 1 Thống kê mô tả và phân phối mẫu

**Bài 1.1.** Một mẫu quan sát về điểm thi kết thúc học phần của sinh viên có kết quả như sau:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2,3 | 3,5 | 7,8 | 5,3 | 6,7 | 5,9 | 6,9 | 8,2 | 9,2 | 5,4 |
| 3,8 | 8,9 | 7,3 | 6,6 | 8,5 | 5,9 | 6,4 | 8,0 | 7,1 | 6,2 |
| 4,2 | 6,5 | 5,7 | 7,6 | 9,1 | 3,9 | 5,2 | 8,4 | 4,5 | 6,6 |
| 4,2 | 7,2 | 5,5 | 8,7 | 6,4 | 4,9 | 7,5 | 4,8 | 2,7 | 0,5 |
| 1,2 | 6,8 | 7,4 | 8,0 | 7,2 | 6,9 | 4,3 | 2,1 | 1,5 | 3,6 |
| 7,7 | 4,5 | 9,0 | 8,6 | 8,3 | 6,9 | 7,4 | 7,6 | 7,5 | 6,1 |
| 4,3 | 0,5 | 6,8 | 6,7 | 7,4 | 2,2 | 3,5 | 4,2 | 7,4 | 5,4 |
| 6,8 | 7,6 | 8,2 | 7,3 | 4,2 | 3,1 | 3,5 | 7,6 | 8,2 | 7,4 |
| 9,5 | 6,8 | 7,8 | 4,5 | 8,5 | 4,6 | 7,7 | 8,0 | 3,5 | 5,6 |
| 4,8 | 7,5 | 7,5 | 8,5 | 5,5 | 5,0 | 6,0 | 8,2 | 6,3 | 6,7 |

a) Hãy cho biết kích thước của mẫu quan sát trên.
b) Hãy lập bảng phân bố tần số của mẫu trên theo các khoảng chia với độ dài mỗi khoảng là 2 và bắt đầu từ 0.
c) Tính trung bình mẫu, phương sai mẫu và độ lệch tiêu chuẩn của mẫu đã chia khoảng.

**Bài 1.2.** Cho mẫu cụ thể của biến ngẫu nhiên $X$ như sau:

| $X$ | 42 | 44 | 45 | 58 | 60 | 64 |
|---|---|---|---|---|---|---|
| $n_i$ | 4 | 5 | 20 | 10 | 8 | 3 |

a) Lập bảng phân phối tần suất.

b) Tính $\overline{x}$ và $s^2$.

**Bài 1.3.** Cho bảng số liệu sau:

| $X$ | $0-10$ | $10-20$ | $20-30$ | $30-50$ | $50-70$ | $70-100$ |
|---|---|---|---|---|---|---|
| $n_i$ | 8 | 18 | 16 | 23 | 11 | 9 |

a) Lập bảng phân bố tần suất.

b) Hãy tìm giá trị trung bình mẫu, độ lệch tiêu chuẩn của mẫu trên.

**Bài 1.4.** Một trạm trộn bê tông nhựa nóng sản xuất nhựa đường thông thường. Để bê tông nhựa đạt chất lượng tốt, nhiệt độ trộn hỗn hợp nhựa và cốt liệu được điều khiển tự động xung quanh mức 177°C. Nhiệt độ trộn được ghi lại trong quá trình làm việc như sau:

| | | | | |
|---|---|---|---|---|
| 169 | 172 | 163 | 167 | 170,9 |
| 170 | 168 | 175,1 | 174 | 169 |
| 166,5 | 176 | 171,2 | 174 | 168 |

(a) Hãy tính nhiệt độ trộn trung bình và độ lệch tiêu chuẩn của mẫu trên.

(b) Nếu coi nhiệt độ trộn là biến ngẫu nhiên chuẩn với trung bình 170°C và độ lệch tiêu chuẩn 3°C thì xác suất để nhiệt độ trộn vượt quá ngưỡng 177°C bằng bao nhiêu?

**Bài 1.5.** Đo đường kính (mm) của một loại đinh tán ta được bảng số liệu sau:

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

(a) Lập bảng phân phối tần số bằng cách chia thành 5 đoạn bắt đầu từ 8,0 với độ dài mỗi khoảng là 0,2.

(b) Vẽ biểu đồ tần số.

(c) Tính giá trị trung bình và độ lệch tiêu chuẩn mẫu.

**Bài 1.6.** Thống kê mức tiêu thụ nhiên liệu của một loại ô-tô chạy thử nghiệm trên một đoạn đường 100 km được cho trong bảng dưới đây:

| $x_i$ (lít) | 4,4-4,6 | 4,6-4,8 | 4,8-5,0 | 5,0-5,2 | 5,2-5,4 | 5,4-5,6 |
|---|---|---|---|---|---|---|
| $n_i$ | 2 | 8 | 14 | 16 | 8 | 4 |

Tính giá trị trung bình và độ lệch tiêu chuẩn của dữ liệu.

**Bài 1.7.** Kiểm tra ngẫu nhiên 36 điểm để đo chiều dày (cm) của lớp nhựa dải đường ta thu được bảng số liệu sau:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5,8 | 5,1 | 4,5 | 5,0 | 4,2 | 5,1 | 3,1 | 4,1 | 4,6 |
| 4,1 | 6,2 | 5,8 | 6,4 | 5,8 | 5,2 | 5,7 | 5,9 | 4,3 |
| 6,3 | 5,9 | 3,3 | 4,6 | 5,3 | 5,7 | 5,6 | 4,1 | 5,2 |
| 5,8 | 4,0 | 6,0 | 3,3 | 4,1 | 4,0 | 3,8 | 3,2 | 5,6 |

(a) Hãy tính giá trị trung bình và độ lệch tiêu chuẩn của mẫu trên.

(b) Bây giờ chia dữ liệu thành các khoảng, mỗi khoảng có độ rộng bằng 0,5 và bắt đầu từ 3,0. Lập bảng phân phối tần số của dữ liệu chia khoảng.

(c) Tính giá trị trung bình và độ lệch tiêu chuẩn dựa trên dữ liệu chia khoảng. So sánh các giá trị tính toán được và giải thích tại sao có sự khác biệt.

**Bài 1.8.** Điều tra mức lương hàng tháng (triệu đồng) của 100 kỹ sư công nghệ thông tin mới ra trường, ta thu được bảng số liệu sau:

| $x_i$ | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-10 |
|---|---|---|---|---|---|---|
| $n_i$ | 6 | 18 | 24 | 32 | 14 | 6 |

Tính mức lương trung bình hàng tháng của một kỹ sư công nghệ thông tin và độ lệch tiêu chuẩn.

**Bài 1.9.** Chiều cao của sinh viên được xấp xỉ bởi phân phối chuẩn với trung bình $\mu = 174,5$ (cm) và độ lệch tiêu chuẩn $\sigma = 6,9$ (cm).

(a) Kiểm tra ngẫu nhiên 25 sinh viên trong số sinh viên trên. Tính xác suất để chiều cao trung bình của 25 sinh viên được kiểm tra nằm trong khoảng (172,5; 175,8).

(b) Nếu kiểm tra 100 mẫu, mỗi mẫu gồm 25 sinh viên. Tính số mẫu trung bình có chiều cao trung bình nằm trong khoảng (172,5; 175,8).

(c) Xác định cỡ mẫu $n$ sao cho giá trị trung bình của mẫu lệch so với trung bình tập chính không quá 1 cm với xác suất ít nhất là 95%.

**Bài 1.10.** Cường độ chịu nén của bê tông mác 200 được xem như biến ngẫu nhiên chuẩn với trung bình $\mu = 90\ kG/cm^2$ (kilogram-lực trên 1 centimet vuông, 1kG = 9,81 N) và độ lệch tiêu chuẩn $\sigma = 4,5\ kG/cm^2$. Nếu kiểm tra 7 mẫu bê tông mác 200 thì xác suất để cường độ chịu nén trung bình của các mẫu này nằm trong khoảng (87,3; 92,5) là bao nhiêu?

**Bài 1.11.** Thiếc hàn là hợp kim có nhiệt độ nóng chảy khá thấp nên được sử dụng trong việc liên kết bề mặt các kim loại khác nhau, ứng dụng nhiều trong kỹ thuật điện, điện tử. Nhiệt độ nóng chảy của thiếc hàn được coi như biến chuẩn với giá trị trung bình là 185 $^oC$ và độ lệch tiêu chuẩn 5 $^oC$. Kiểm tra 10 mẫu thiếc hàn, tính xác suất để nhiệt độ nóng chảy trung bình của mẫu lớn hơn 190 $^oC$.

**Bài 1.12.** Một công ty điện lực quản lý một vùng dân cư có 20 nghìn hộ dân. Lượng điện tiêu thụ của mỗi hộ gia đình trong một tháng được xem như một biến ngẫu nhiên với trung bình 370 kWh và độ lệch tiêu chuẩn 350 kWh. Hãy ước tính xác suất để tổng lượng điện tiêu thụ không vượt quá 7,5 triệu kWh.

**Bài 1.13.** Một kết cấu thép có thể chịu được tải trọng tối đa 3,5 tấn. Giả sử rằng trọng lượng của mỗi kiện hàng là biến ngẫu nhiên với trung bình 59 kg và độ lệch tiêu chuẩn 15 kg. Hãy ước tính số lượng kiện hàng lớn nhất có thể đặt lên kết cấu đó để xác suất kết cấu an toàn cao hơn 99%.

**Bài 1.14.** Thời gian quét một phiếu thông tin và nhập vào cơ sở dữ liệu của một hệ thống kho vận (logistic) tự động là một biến ngẫu nhiên có phân phối chuẩn với giá trị trung bình $\mu = 10$ giây và độ lệch tiêu chuẩn $\sigma = 5$ giây. Tính xác suất để tổng thời gian quét và lưu trữ 10 phiếu thông tin của hệ thống không vượt quá 120 giây.

**Bài 1.15.** Một trường đại học hàng năm tuyển 1500 sinh viên cho tất cả các chuyên ngành mà trường đào tạo. Thống kê cho thấy số lượng sinh viên đến nhập trường bằng 85% số lượng giấy gọi nhập học được phát ra. Trong đợt tuyển sinh năm 2016, để đảm bảo số lượng sinh viên theo học như chỉ tiêu đã đăng ký, trường đại học dự kiến phát ra 1720 giấy gọi nhập học. Tính xác suất để số sinh viên nhập học lớn hơn số chỉ tiêu đăng ký.

**Bài 1.16.** Thống kê cho thấy, tỉ lệ sinh viên tốt nghiệp đại học năm 2016 từ bỏ công việc đầu tiên sau 3 tháng thử việc đạt 60%. Hỏi phải tiến hành thăm dò bao nhiêu sinh viên tốt nghiệp năm 2016 để tỉ lệ sinh viên gắn bó với công việc đầu tiên nhiều hơn 3 tháng nằm trong khoảng $40 \pm 5\%$, với xác suất nhiều hơn 90%.

**Bài 1.17.** Một nhà máy chế tạo một loại thiết bị điện tử có khả năng tự ngắt điện khi nhiệt độ tăng lên quá cao. Nghiên cứu trước đó cho thấy nhiệt độ làm cho thiết bị tự ngắt có phân phối chuẩn với phương sai $\sigma^2$. Hãy xác định giá trị $a$ sao cho

$$\mathbb{P}(S^2/\sigma^2 < a) = 0,9$$

trong đó $S^2$ là phương sai của mẫu gồm 5 thiết bị được chọn ngẫu nhiên để kiểm tra.

**Bài 1.18.** Một cánh tay rô-bốt sử dụng mắt điện tử để xác định vị trí vật thể trong không gian ba chiều. Quá trình định vị luôn làm nảy sinh ra sai số. Giả sử các sai số $X$, $Y$, $Z$ tương ứng theo ba chiều không gian là các biến ngẫu nhiên độc lập có cùng phân phối chuẩn với trung bình 0 mm và độ lệch tiêu chuẩn 3 mm. Hãy xác định giá trị $d$ sao cho với xác suất 0,95 điểm được định vị bởi mắt điện tử lệch ra khỏi vị trí thực tế không vượt quá $d$.

**Bài 1.19.** Ký hiệu bu lông M8x80 nghĩa là bu lông có đường kính ngoài của ren là 8 mm và chiều dài là 80 mm. Trong thực tế sản xuất chiều dài bu lông M8x80 có phân phối chuẩn với trung bình $\mu_1$ =80 mm và độ lệch tiêu chuẩn $\sigma_1$ =1 mm.

a) Hỏi phải kiểm tra mẫu gồm bao nhiêu bu lông M8x80 để xác suất chiều dài trung bình của mẫu lệch không quá giá trị tiêu chuẩn $\mu_1$ 0,5 mm không ít hơn 95%.

b) Đường kính bu lông M8x80 có phân phối chuẩn với trung bình là $\mu_2$ =8mm và độ lệch tiêu chuẩn $\sigma_2 = 0,1$mm. Một bu lông loại này được gọi là đạt tiêu chuẩn nếu đường kính $d$ của nó thỏa mãn $\mu_2 - 0,2 < d < \mu_2 + 0,2$. Tính tỉ lệ bu lông đạt tiêu chuẩn.

c) Tính xác suất để trong 1 lô 1000 bu lông loại M8x80 có ít hơn 50 bu lông không đạt tiêu chuẩn.

d) Tính xác suất để phương sai của 6 bu lông chọn ngẫu nhiên lớn hơn 0,03 mm$^2$.

## 2 Ước lượng tham số

*Ước lượng khoảng cho kỳ vọng của biến chuẩn trong trường hợp đã biết phương sai* $\mathbb{V}[X] = \sigma^2$

**Bài 2.1.** Cho mẫu

| $X$ | $19-21$ | $21-23$ | $23-25$ | $25-27$ | $27-29$ |
|---|---|---|---|---|---|
| $n_i$ | 5 | 6 | 7 | 6 | 4 |

Biết $X$ tuân theo luật $N(\mu, \sigma^2)$. Hãy tìm khoảng ước lượng của $EX$, biết $\gamma = 0,95$ và $\sigma = 3$.

**Bài 2.2.** Cho mẫu

| $X$ | $7-9$ | $9-11$ | $11-13$ | $13-15$ | $15-17$ |
|---|---|---|---|---|---|
| $n_i$ | 4 | 9 | 7 | 8 | 8 |

Biết $X$ tuân theo luật $N(\mu, \sigma^2)$. Hãy tìm khoảng ước lượng của $EX$, với $\gamma = 0,99$, biết $\sigma = 2,8$.

**Bài 2.3.** Chiều dài của một chi tiết máy do một phân xưởng sản xuất là một biến chuẩn $N(\mu, \sigma^2)$ với độ lệch bình phương trung bình $\sigma = 3$.

a) Lấy ngẫu nhiên 36 chi tiết đem đo và thu được độ dài trung bình $\overline{x} = 24,55$. Hãy tìm khoảng tin cậy của $\mu = E(X)$, với độ tin cậy $\gamma = 0,95$.

b) Cần phải lấy ngẫu nhiên bao nhiêu chi tiết để với độ tin cậy $\gamma = 0,99$, độ dài khoảng ước lượng của $\mu = E(X)$ không vượt quá $0,6$.

**Bài 2.4.** Gọi $X$ là mức xăng tiêu thụ cho ô tô con trên đoạn đường từ $A$ đến $B$. để ước lượng mức xăng hao phí trung bình, người ta lấy 36 chiếc và cho chạy thử thì tính được $\overline{x} = 28,45$ lít. Biết rằng độ lệch tiêu chuẩn $\sigma = 3$, mức xăng tiêu thụ $X$ là đại lượng ngẫu nhiên có phân phối chuẩn. Với độ tin cậy 95%, hãy xác định khoảng ước lượng cho mức xăng hao phí trung bình.

*Ước lượng khoảng cho kỳ vọng của biến chuẩn trong trường hợp phương sai $\mathbb{V}[X] = \sigma^2$ chưa biết và $n \leq 30$*

**Bài 2.5.** Để nghiệm thu đoạn đường do bên B thi công, bên A tiến hành khoan thăm dò 16 điểm ngẫu nhiên trên con đường và thu được dãy số liệu (tính bằng mm) chỉ độ dày của lớp bê tông nhựa trải đường như sau:

136; 139; 134; 137; 132; 133; 135; 138;
137; 141; 145; 142; 143; 137; 138; 133.

Với độ tin cậy 95% hãy tìm khoảng ước lượng chiều dầy trung bình của lớp bê tông nhựa đó. Biết chiều dày của lớp bê tông có phân phối chuẩn.

**Bài 2.6.** Khảo sát một mẫu 16 sinh viên cho thấy số lần họ đi xem phim trong một năm như sau:

9; 12; 15; 7; 7; 9; 12; 8; 6; 15; 13; 14; 10; 10; 8; 9.

Hãy tìm khoảng tin cậy 99% cho số lần trung bình mà mỗi sinh viên tới rạp chiếu phim trong một năm. Biết số lần sinh viên đi xem phim có phân phối chuẩn.

**Bài 2.7.** Để kiểm tra mức xăng hao phí của một loại xe ô tô. Người ta chọn ngẫu nhiên 28 chiếc xe và cho chạy trên cùng một đoạn đường 300 km. Kết quả thu được như sau:

| $X$ | $4,6-4,8$ | $4,8-5,0$ | $5,0-5,2$ | $5,2-5,4$ | $5,4-5,6$ |
|---|---|---|---|---|---|
| $n_i$ | 6 | 5 | 9 | 4 | 4 |

Với độ tin cậy $\gamma = 0,95$, hãy tìm khoảng ước lượng của lượng xăng hao phí trung bình. Biết mức xăng hao phí có phân phối chuẩn.

**Bài 2.8.** Người ta đo cường độ chịu nén (kG/cm$^2$) của 12 mẫu bê tông cùng loại và nhận được số liệu:

2216; 2237; 2249; 2204; 2225; 2301;
2281; 2263; 2318; 2255; 2275; 2295.

Giả sử cường độ chịu nén của bê tông là biến ngẫu nhiên có phân phối chuẩn. Hãy xây dựng khoảng tin cậy 95% cho cường độ chịu nén trung bình của loại bê tông này.

**Bài 2.9.** Một máy sản xuất các thanh kim loại được dùng trong hệ thống giảm xóc ô tô. Một mẫu gồm 15 thanh được chọn và người ta đo đường kính (mm) của chúng. Dữ liệu thu được như sau:

$8; 24;\ 8,25;\ 8,20;\ 8,23;\ 8,24;\ 8,21;\ 8,26;\ 8,26;$
$8,20;\ 8,25;\ 8,23;\ 8,23;\ 8,19;\ 8,28;\ 8,24.$

Giả sử đường kính của thanh kim loại là biến ngẫu nhiên có phân phối chuẩn. Hãy xây dựng khoảng tin cậy 95% cho đường kính trung bình của thanh kim loại.

*Ước lượng khoảng cho kỳ vọng của biến chuẩn trong trường hợp phương sai $\mathbb{V}[X] = \sigma^2$ chưa biết và $n > 30$*

**Bài 2.10.** Hao phí nguyên liệu cho một đơn vị sản phẩm là một đại lượng ngẫu nhiên tuân theo luật phân phối chuẩn. Sản xuất thử 36 sản phẩm và thu được bảng số liệu sau:

| $X$ | $29-31$ | $31-33$ | $33-35$ | $35-37$ | $37-39$ |
|---|---|---|---|---|---|
| $n_i$ | 5 | 9 | 12 | 6 | 4 |

Với độ tin cậy 95% hãy tìm khoảng ước lượng mức hao phí nguyên liệu trung bình cho một đơn vị sản phẩm.

**Bài 2.11.** Giả sử rằng thu nhập $X$ hàng tháng của một kỹ sư sau khi ra trường 5 năm là một đại lượng ngẫu nhiên có phân phối chuẩn $N(\mu, \sigma^2)$. Lấy một mẫu quan sát của $X$ có kích thước $n = 64$ ta có kết quả:

| $X$ (triệu) | $4-6$ | $6-8$ | $8-10$ | $10-12$ | $12-14$ |
|---|---|---|---|---|---|
| $n_i$ (người) | 12 | 24 | 18 | 7 | 3 |

Với độ tin cậy $\gamma = 0,99$ hãy xây dựng khoảng tin cậy cho thu nhập trung bình hàng tháng của kỹ sư .

**Bài 2.12.** Người ta ghi lại giá mặt hàng $A$ tại 50 cửa hàng và có kết quả sau:

| $X$ | 95 | 99 | 102 | 98 | 105 |
|---|---|---|---|---|---|
| $n_i$ | 5 | 15 | 15 | 10 | 5 |

Hãy tìm khoảng ước lượng giá trung bình của mặt hàng $A$ với độ tin cậy 95%. Giả thiết giá mặt hàng $A$ tuân theo luật chuẩn.

**Bài 2.13.** Chiều cao $X$ của trẻ em tuân theo phân phối chuẩn. Hãy ước lượng chiều cao trung bình của trẻ em với độ tin cậy 95%, nếu như đo 55 em có kết quả sau:

| Chiều cao | $1,49$ | $1,50$ | $1,51$ | $1,52$ | $1,55$ | $1,57$ | $1,58$ | $1,60$ |
|---|---|---|---|---|---|---|---|---|
| Số TE | 3 | 7 | 8 | 9 | 6 | 5 | 9 | 8 |

**Bài 2.14.** Với độ tin cậy 95%, hãy tìm khoảng ước lượng mức xăng hao phí trung bình cho một loại mô tô chạy trên cùng một đoạn đường từ $A$ đến $B$, dựa vào bảng số liệu sau và biết $X$ (mức xăng hao phí) là biến chuẩn.

| $X$ | $4,6-4,8$ | $4,8-5,0$ | $5,0-5,2$ | $5,2-5,4$ | $5,4-5,6$ |
|---|---|---|---|---|---|
| $n_i$ | 3 | 8 | 13 | 7 | 5 |

**Bài 2.15.** Thời gian hoàn thành 1 sản phẩm $X$ (phút) tuân theo phân phối chuẩn. Hãy tìm khoảng ước lượng cho thời gian trung bình hoàn thành 1 sản phẩm, với độ tin cậy 99%, khi ta theo dõi một nhóm công nhân làm việc và có số liệu sau:

| $X$ | 42 | 44 | 45 | 58 | 60 | 64 |
|---|---|---|---|---|---|---|
| $n_i$ | 4 | 5 | 20 | 10 | 8 | 3 |

*Ước lượng khoảng cho tỷ lệ*

**Bài 2.16.** Phỏng vấn 2500 người được chọn ngẫu nhiên trong một thành phố. Kết quả cho thấy có 980 người thường xuyên sử dụng Internet.

a) Với độ tin cậy $0,98$ hãy ước lượng tỷ lệ người dân trong thành phố có sử dụng Internet.

b) Nếu dân số của thành phố là 7 triệu người thì với độ tin cậy trên số dân sử dụng Internet trong thành phố là bao nhiêu?

**Bài 2.17.** Kiểm tra ngẫu nhiên 400 người đi xe máy ở khu vực có 500.000 người đi xe máy thấy có 360 người có bằng lái. Với độ tin cậy $0,95$ hãy ước lượng số người đi xe máy có bằng lái trong khu vực.

**Bài 2.18.** Để điều tra số cá có trong Hồ Tây, cơ quan quản lý đánh bắt 900 con, đánh dấu rồi thả lại hồ. Lần sau lại bắt 400 con thì có 94 con có dấu. Hãy ước lượng số cá trong hồ, với độ tin cậy 95%.

**Bài 2.19.** Thăm dò ý kiến của 400 người ở Liên hiệp đường sắt thấy có 236 người đồng ý với Nghị quyết nâng cao chất lượng chạy tàu và phục vụ hành khách. Hãy ước lượng số người đồng ý với Nghị quyết trên trong toàn Liên hợp với độ tin cậy $\gamma = 0,99$, biết rằng số người trong toàn Liên hợp là 24.000 người.

**Bài 2.20.** Trong cuộc thăm dò ý kiến 1600 khách hàng người ta thấy có 1315 người thích mặt hàng A. Hãy ước lượng tỷ lệ người tiêu dùng thích mặt hàng A với độ tin cậy 90%.

**Bài 2.21.** Trong cuộc thăm dò ý kiến 400 khách hàng của một hãng sản xuất hàng điện tử người ta thấy có 136 khách hàng chưa hài lòng với chính sách hậu mãi hiện có của hãng đó. Tìm khoảng tin cậy cho tỷ lệ khách hàng chưa hài lòng với chính sách hậu mãi hiện có với độ tin cậy 90%.

**Bài 2.22.** Cơ quan cảnh sát giao thông kiểm tra hệ thống phanh của 250 chiếc xe tải trên đường quốc lộ thì phát hiện 120 chiếc có phanh chưa an toàn.
a) Tìm khoảng ước lượng cho tỷ lệ xe tải có phanh chưa an toàn với độ tin cậy 95%.
b) Tìm khoảng ước lượng cho tỷ lệ xe tải có phanh tốt với độ tin cậy 98%.

**Bài 2.23.** Để ước lượng tỷ lệ phế phẩm của một dây chuyền sản suất, người ta chọn ngẫu nhiên 300 sản phẩm và kiểm tra thấy có 20 phế phẩm.

a) Với độ tin cậy $\gamma = 0,99$ hãy tìm khoảng ước lượng cho tỷ lệ phế phẩm của dây chuyền trên.

b) Để sai số của việc ước lượng không vượt quá 0,01 thì ta phải kiểm tra tối thiểu bao nhiêu sản phẩm với độ tin cậy 95%.

*Ước lượng khoảng cho phương sai của biến chuẩn*

**Bài 2.24.** Kích thước của một chi tiết máy là một đại lượng ngẫu nhiên có phân phối chuẩn. Trong một mẫu khảo sát gồm 30 chi tiết, người ta tính được $\bar{x} = 0,47$cm và $s = 0,032$cm. Hãy tìm khoảng tin cậy cho phương sai của kích thước chi tiết máy với độ tin cậy 95%.

**Bài 2.25.** Lấy 28 mẫu xi măng của một nhà máy sản xuất xi măng để kiểm tra về sức chịu lực (kG/cm$^3$), kết quả như sau:

| | | | | | | |
|---|---|---|---|---|---|---|
| 10,0 | 13,0 | 13,7 | 11,5 | 11,0 | 13,5 | 12,4 |
| 13,5 | 13,0 | 10,0 | 11,0 | 11,5 | 13,0 | 12,2 |
| 11,5 | 13,7 | 12,0 | 12,2 | 11,5 | 10,5 | 11,5 |
| 13,0 | 12,2 | 11,5 | 13,0 | 10,5 | 12,4 | 10,0 |

Với độ tin cậy 95%, hãy ước lượng:

a) Sức chịu lực trung bình của xi măng do nhà máy này sản xuất.

b) Phương sai của sức chịu lực.

**Bài 2.26.** Dung lượng (đơn vị Ampe-giờ) của 10 chiếc pin được ghi lại dưới đây

140, 136, 150, 144, 148, 152, 138, 141, 143, 151

a) Ước lượng giá trị phương sai tập chính $\sigma^2$.

b) Tìm khoảng tin cậy 98% cho $\sigma^2$.

**Bài 2.27.** Để khảo sát chất lượng đóng chai của một máy đóng chai tự động, người ta chọn ngẫu nhiên 16 chai trên dây chuyền sản xuất, đo lượng chất lỏng trong chai và tính được độ lệch tiêu chuẩn mẫu là $s = 0,0875$ (lít). Nếu độ lệch quá lớn thì sẽ ảnh hưởng tới chất lượng của việc đóng chai, tức là sẽ có những chai quá ít hoặc quá nhiều. Giả sử lượng chất lỏng được đóng ở mỗi chai là biến ngẫu nhiên chuẩn. Hãy tìm khoảng ước lượng cho phương sai của lượng chất lỏng đóng chai với độ tin cậy 95%.

## 3 Kiểm định giả thuyết thống kê

*Kiểm định giả thuyết về giá trị trung bình*

**Bài 3.1.** Hao phí nguyên liệu cho 1 đơn vị sản phẩm là đại lượng ngẫu nhiên chuẩn với độ lệch chuẩn $\sigma = 0,03$ và hao phí trung bình là $29,9$ gram. Nghi ngờ máy móc trục trặc làm cho hao phí nguyên liệu cho 1 đơn vị sản phẩm tăng lên. Người ta kiểm tra 36 sản phẩm thì có bảng số liệu sau:

| $x_i$ (gram) | $29,5-29,7$ | $29,7-29,9$ | $29,9-30,1$ | $30,1-30,3$ | $30,3-30,5$ |
|---|---|---|---|---|---|
| $n_i$ số SP | 7 | 9 | 12 | 5 | 3 |

Hãy kết luận ý kiến nêu trên với mức ý nghĩa $\alpha = 0,05$.

**Bài 3.2.** Tuổi thọ của một loại bóng hình của máy vô tuyến truyền hình là 1 đại lượng ngẫu nhiên $X$ tuân theo luật phân phối chuẩn với $EX = 3500$ giờ và độ lệch tiêu chuẩn là $\sigma = 20$ giờ. Nghi ngờ tuổi thọ bị thay đổi, người ta tiến hành theo dõi 25 bóng thấy tuổi thọ trung bình là 3422 giờ. Với mức ý nghĩa $\alpha = 0,05$, hãy kiểm định điều nghi ngờ trên.

**Bài 3.3.** Giả sử thời gian $X$ để hoàn thành một sản phẩm của công nhân là một đại lượng ngẫu nhiên tuân theo luật chuẩn $N(\mu, \sigma^2)$ với $\sigma^2 = 4$. Định mức thời gian để hoàn thành một sản phẩm là 25 phút. Có ý kiến cho rằng thời gian trung bình để hoàn thành một sản phẩm ít hơn định mức được cho. Hãy kiểm định ý kiến trên với mức ý nghĩa $\alpha = 0,05$, biết rằng khi lấy một mẫu quan sát có cỡ $n = 25$ ta thu được trung bình mẫu $\bar{x} = 24,12$ phút.

**Bài 3.4.** Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm $X$ là biến chuẩn với kỳ vọng $\mu = 100$ g, độ lệch chuẩn $\sigma = 1,8$. Qua một thời gian sản xuất người ta nghi ngờ trọng lượng sản phẩm có xu hướng tăng lên. Cân thử 100 sản phẩm ta có bảng số liệu sau:

| $x_i$ (gam) | $96-98$ | $98-100$ | $100-102$ | $102-104$ | $104-106$ |
|---|---|---|---|---|---|
| $n_i$ số SP | 10 | 20 | 45 | 15 | 10 |

Với mức ý nghĩa $\alpha = 0,05$, hãy kết luận về điều nghi ngờ nói trên.

**Bài 3.5.** Một máy sản xuất đinh ốc có đường kính là biến ngẫu nhiên $X$ tuân theo luật phân bố chuẩn $N(\mu, \sigma^2)$ với $\sigma = 0,12$ và đường kính trung bình là 2. Nghi ngờ máy hoạt động không bình thường làm cho đường kính của các đinh ốc bị thay đổi, người ta tiến hành đo thử 100 đinh ốc và được số liệu cho ở bảng dưới đây:

| $x_i$ | $1,8$ | $1,9$ | $2,0$ | $2,1$ | $2,2$ |
|---|---|---|---|---|---|
| $n_i$ | 10 | 20 | 25 | 35 | 10 |

Với mức ý nghĩa $0,05$, hãy kiểm định điều nghi ngờ nói trên.

**Bài 3.6.** Một hãng chuyên sản xuất ô tô cho ra đời một dòng xe ô tô điện mới và tuyên bố rằng xe loại này có khả năng tăng tốc lên 100 km/h chỉ trong vòng 6 giây. Người ta chọn ngẫu nhiên 9 xe để kiểm tra, tốc độ đạt được khi tăng tốc sau 6 giây được ghi lại dưới đây

91, 93, 92, 90, 98, 91, 92, 95, 96

Với mức ý nghĩa 0,01, có thể nói rằng tuyên bố của hãng xe hơi phóng đại hay không? Giả sử rằng tốc độ đạt được sau 6 giây có phân phối chuẩn.

**Bài 3.7.** Thời gian sử dụng của một loại bóng đèn chiếu sáng công cộng là đại lượng ngẫu nhiên $X$ (tính theo tháng) có phân phối chuẩn với trung bình là 36 tháng. Nghi ngờ các tác động của môi trường là giảm thời gian sử dụng người ta lấy một mẫu thực nghiệm và có kết quả

| $X$ | $27-30$ | $30-33$ | $33-36$ | $36-39$ | $39-42$ |
|---|---|---|---|---|---|
| $n_i$ | 8 | 7 | 7 | 4 | 3 |

Với mức ý nghĩa $0,01$, hãy kiểm định điều nghi ngờ nói trên.

**Bài 3.8.** Định mức thời gian hoàn thành 1 sản phẩm là 14 phút. Liệu có cần thay đổi định mức không nếu theo dõi thời gian hoàn thành 1 sản phẩm ở 25 công nhân ta có bảng số liệu sau:

| $x_i$ t.gian (phút) | $10-12$ | $12-14$ | $14-16$ | $16-18$ | $18-20$ |
|---|---|---|---|---|---|
| $n_i$ số CN | 2 | 6 | 7 | 7 | 3 |

Yêu cầu kết luận với mức ý nghĩa $\alpha = 0,05$. Biết rằng thời gian hoàn thành 1 sản phẩm "X" là đại lượng ngẫu nhiên có phân phối chuẩn.

**Bài 3.9.** Mức xăng tiêu thụ của 1 loại xe ô tô chạy từ Hà Nội đến Thanh Hoá là 1 đại lượng ngẫu nhiên $X$ tuân theo luật phân phối chuẩn với trung bình là 50 lít xăng. Đoạn đường được bảo dưỡng, người ta cho rằng mức xăng hao phí trung bình giảm xuống. Quan sát 30 ô tô cùng loại ta thu được số liệu sau:

| $x_i$ | $47-48$ | $48-49$ | $49-50$ | $50-51$ | $51-52$ |
|---|---|---|---|---|---|
| $n_i$ | 5 | 10 | 10 | 3 | 2 |

Với mức ý nghĩa $\alpha = 0,05$. Hãy cho kết luận về ý kiến đã nêu ra.

**Bài 3.10.** Trọng lượng những bao phân đạm do nhà máy sản xuất ra là 1 biến ngẫu nhiên $X$ tuân theo luật phân bố chuẩn, với $EX = 50$ kg. Khách hàng cho rằng trọng lượng đó đã thay đổi và ít hơn 50 kg. Cân thử 100 bao, ta có bảng số liệu dưới đây:

| $x_i$ | $47-48$ | $48-49$ | $49-50$ | $50-51$ | $51-52$ |
|---|---|---|---|---|---|
| $n_i$ | 30 | 40 | 20 | 5 | 5 |

Với mức ý nghĩa $\alpha = 0,05$, hãy cho kết luận về ý kiến của khách hàng.

**Bài 3.11.** Tỷ lệ Titan trong một loại hợp kim là thành tố quan trọng để xác định độ cứng của vật liệu. Kiểm tra ngẫu nhiên 20 mẫu vật liệu ta được bảng số liệu sau về tỷ lệ phần trăm titan trong mẫu:

8,32 8,05 8,93 8,65 8,25 8,46 8,52 8,35 8,36 8,41
8,42 8,30 8,71 8,75 8,60 8,83 8,50 8,38 8,29 8,46

Có ý kiến cho rằng tỷ lệ phần trăm trung bình của titan trong vật liệu trên thấp hơn $8,5$. Với mức ý nghĩa 5% hãy đánh giá ý kiến trên. Giả sử rằng tỷ lệ phần trăm của titan trong vật liệu trên tuân theo phân phối chuẩn.

**Bài 3.12.** Đường kính (mm) của vòng bi được đo bởi thước kẹp được kết quả cho bởi bảng sau:

2,65 2,63 2,66 2,67 2,67 2,65 2,67 2,67
2,65 2,68 2,63 2,64 2,65 2,64 2,67 2,63

Có ý kiến cho rằng đường kính trung bình của vòng vi trên bằng $2,65$ (mm). Với mức ý nghĩa 1%, hãy đánh giá ý kiến trên. Biết rằng đường kính của vòng bi có phân phối chuẩn.

**Bài 3.13.** Thời gian (giờ) để một loại sơn khô khi sơn tường là một tham số quan trọng trong quá trình sản suất. Có ý kiến cho rằng loại sơn mà một công ty sản suất có thời gian khô dưới 1,5 (giờ). Kiểm tra ngẫu nhiên 36 mẫu về thời gian khô ta có bảng số liệu sau:

| $x_i$ | 1,0-1,3 | 1,3-1,4 | 1,4-1,5 | 1,5-1,6 | 1,6-1,7 | 1,7-2,0 |
|---|---|---|---|---|---|---|
| $n_i$ | 6 | 10 | 9 | 6 | 3 | 2 |

Với mức ý nghĩa 5%, hãy đánh giá ý kiến trên. Biết rằng thời gian khô của loại sơn trên tuân theo luật chuẩn.

**Bài 3.14.** Có ý kiến cho rằng thời gian trung bình tự học ở nhà của sinh viên là 15 giờ mỗi tuần. Kiểm tra một lớp có 64 sinh viên ta được bảng số liệu sau:

| $x_i$ (giờ) | 0-4 | 4-8 | 8-12 | 12-16 | 16-20 | 20-24 |
|---|---|---|---|---|---|---|
| $n_i$ | 5 | 12 | 15 | 17 | 10 | 5 |

Với mức ý nghĩa 5% hãy đánh giá ý kiến trên.

**Bài 3.15.** Tiêu chuẩn nước an toàn về hàm lượng asen không quá 0,04(mg/lít). Nghi ngờ nước sinh hoạt của một khu vực không đạt tiêu chuẩn an toàn. Lấy ngẫu nhiên 36 mẫu nước và đo hàm lượng asen ta có bảng số liệu sau:

| $x_i$ | 0,02-0,03 | 0,03-0,04 | 0,04-0,05 | 0,05-0,06 | 0,06-0,07 | 0,07-0,08 |
|---|---|---|---|---|---|---|
| $n_i$ | 3 | 6 | 8 | 9 | 6 | 4 |

Với mức ý nghĩa $0,05$ hãy cho biết nước sinh hoạt ở khu vực trên có an toàn không.

*Kiểm định giả thuyết cho tỷ lệ*

**Bài 3.16.** Một tòa báo thanh niên thông báo có 35% học sinh phổ thông trung học là độc giả thường xuyên. Phỏng vấn 400 em thì có 136 em đọc báo đó thường xuyên. Hãy kiểm định tính chính xác của thông báo trên với mức ý nghĩa $0,05$.

**Bài 3.17.** Tỷ lệ phế phẩm do một nhà máy sản xuất là 5%. Kiểm tra ngẫu nhiên 640 sản phẩm thấy có 37 phế phẩm. Có ý kiến cho rằng máy móc đã xuống cấp, tỷ lệ phế phẩm có chiều hướng tăng lên. Với mức ý nghĩa $\alpha = 0,05$, hãy cho kết luận về ý kiến trên.

**Bài 3.18.** Có ý kiến cho rằng tỷ lệ sinh viên nghiện game online là $0,12$. Có ý kiến cho rằng đó là tỷ lệ thực tế. Chọn ngẫu nhiên 1600 sinh viên, kết quả cho thấy có 154 sinh viên nghiện game online. Với mức ý nghĩa $\alpha = 0,01$ hãy kết luận về ý kiến trên.

**Bài 3.19.** Công ty FPT cung cấp dịch vụ Internet thông báo rằng dịch vụ của họ cung cấp cho 70% hộ gia đình của một khu vực dân cư. Kiểm tra ngẫu nhiên 200 hộ gia đình của khu vực trên thấy có 125 hộ sử dụng dịch vụ Internet của công ty FPT. Với mức ý nghĩa $0,05$ có thể kết luận rằng thông báo của công ty quá sự thật không.

**Bài 3.20.** Cơ quan quản lý thị trường thông báo rằng số mũ bảo hiểm không đạt tiêu chuẩn chất lượng không quá 5%. Kiểm tra 200 người có sử dụng mũ bảo hiểm thấy có 16 chiếc không đạt tiêu chuẩn. Với mức ý nghĩa 0,02 hãy cho đánh giá về ý kiến trên.

*Kiểm định giả thuyết về hai giá trị trung bình*

**Bài 3.21.** Một xí nghiệp sử dụng hai dây chuyền tự động khác nhau để sản xuất. Sau khi theo dõi người ta thu được kết quả sau

| Dây chuyền | Số ca theo dõi | Số sản phẩm trung bình | Độ lệch tiêu chuẩn |
|---|---|---|---|
| I | 100 | 306 | 10 |
| II | 150 | 375 | 25 |

Với mức ý nghĩa $\alpha = 0,05$, có thể kết luận dây chuyền thứ hai có sản lượng tốt hơn dây chuyền thứ nhất hay không? Giả thiết sản lượng tuân theo phân phối chuẩn.

**Bài 3.22.** Có ý kiến cho rằng mức thu nhập $X$ (tính theo triệu đồng/tháng) của kỹ sư thuộc lĩnh vực xây dựng tốt hơn mức thu nhập $Y$ của kỹ sư thuộc lĩnh vực công nghệ. Người ta tiến hành lấy một mẫu quan sát và có kết quả.

| Ngành | Số kỹ sư theo dõi | Thu nhập trung bình | Độ lệch tiêu chuẩn |
|---|---|---|---|
| Xây dựng | 1500 | $8,1$ | $3,2$ |
| Công nghệ | 1600 | $7,5$ | $1,1$ |

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý kiến đã nêu.

**Bài 3.23.** So sánh trọng lượng trung bình của trẻ sơ sinh ở thành thị và nông thôn, người ta cân thử trọng lượng của 5000 cháu và có bảng số liệu sau:

| Khu vực | Số trẻ | Tăng trọng trung bình | độ lệch tiêu chuẩn |
|---|---|---|---|
| Nông thôn | 4000 | $3,0$ | $1,2$ |
| Thành thị | 1000 | $3,2$ | $0,5$ |

Với mức ý nghĩa 5% có thể coi trọng lượng trung bình của trẻ sơ sinh ở thành phố cao hơn nông thôn hay không? Giả sử trọng lượng trung bình của trẻ là đại lượng ngẫu nhiên chuẩn.

*Kiểm định giả thuyết về hai tỷ lệ*

**Bài 3.24.** Có ý kiến cho rằng tỷ lệ cận thị của học sinh thành thị cao hơn của học sinh nông thôn. Người ta tiến hành kiểm tra 1000 cháu và thu được kết quả

| Nhóm | Số học sinh kiểm tra | Số học sinh bị cận thị |
|---|---|---|
| Thành thị | 400 | 96 |
| Nông thôn | 600 | 98 |

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý kiến đã nêu.

**Bài 3.25.** Có ý kiến cho rằng tỷ lệ sinh viên khá giỏi của ngành Công trình và ngành Kinh tế là như nhau. Người ta chọn ngẫu nhiên 500 sinh viên và thu được kết quả

| Ngành | Số sinh viên được chọn | Số sinh viên khá giỏi |
|---|---|---|
| Công trình | 300 | 93 |
| Kinh tế | 200 | 71 |

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý kiến đã nêu.

**Bài 3.26.** Một kỹ sư đưa ra một quy trình sản xuất mới để giảm tỷ lệ phế phẩm của nhà máy. Kiểm tra về hai quy trình sản xuất ta có bảng số liệu sau

| Quy trình | Số sản phẩm kiểm tra | Số phế phẩm |
|---|---|---|
| Cũ | $n_1 = 250$ | $m_1 = 18$ |
| Mới | $n_2 = 350$ | $m_2 = 35$ |

Với mức ý nghĩa 0,05 hãy đánh giá ý kiến của kỹ sư trên.

## 4 Tương quan và hồi quy

**Bài 4.1.** Một mẫu quan sát của đại lượng ngẫu nhiên hai chiều $(X, Y)$ có giá trị như sau

$$(2,1; 4,12),\ (2,2; 4,34),\ (2,4; 4,56),\ (2,5; 4,63)$$
$$(2,25; 4,38),\ (2,45; 4,75),\ (2,16; 4,4),\ (2,34; 4,62)$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

**Bài 4.2.** Người ta lấy một mẫu thực nghiệm của đại lượng ngẫu nhiên hai chiều $(X, Y)$ và thu được kết quả:

| $X$ | $3,6$ | $3,8$ | $4,3$ | $4,5$ | $4,9$ | $5,2$ | $5,4$ |
|---|---|---|---|---|---|---|---|
| $Y$ | $7,1$ | $7,83$ | $9,62$ | $10,05$ | $10,7$ | $11,6$ | $12.3$ |

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

**Bài 4.3.** Số liệu về lượng vận chuyển của một công ty vận tải trong các năm qua (tính theo triệu tấn) là như sau:

| Năm | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|
| Khối lượng | 28 | 31 | $35,5$ | 36 | $37,5$ | 39 | $41,5$ |

a) Tìm hàm xu thế tuyến tính biểu thị năng lực vận chuyển của công ty đó.

b) Dự báo khối lượng vận chuyển năm 2017 và tìm khoảng tin cậy 95% cho giá trị đó.

**Bài 4.4.** Phân tích chi phí bảo dưỡng cho xe tải trong 8 năm sử dụng đầu tiên (tính theo triệu đồng) ta có kết quả:

| Năm thứ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Chi phí TB | 6 | $8,2$ | $8,7$ | $10,5$ | 12 | $14,4$ | 17 | $19,2$ |

a) Tìm hàm xu thế tuyến tính biểu thị chi phí bảo dưỡng xe.

b) Dự báo chi phí bảo dưỡng xe trong năm sử dụng thứ 10 và tìm khoảng tin cậy 90% cho giá trị đó.

**Bài 4.5.** Số liệu về dân số (tính theo nghìn người) thành phố Hồ Chí Minh trong các năm gần đây được thống kê như sau:

| Năm | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|
| Số dân | $7498,4$ | $7660,3$ | $7820,0$ | $7981,9$ | $8146,3$ | $8320,1$ |

a) Tìm hàm xu thế tuyến tính biểu thị dân số của thành phố Hồ Chí Minh.

b) Dự báo số dân năm 2017 của thành phố này và tìm khoảng tin cậy 98% cho giá trị đó.

**Bài 4.6.** Để nghiên cứu về quan hệ giữa khối lượng bốc dỡ $X$ (nghìn tấn) và thời gian bốc dỡ $Y$ (giờ) người ta lấy một mẫu thực nghiệm và thu được kết quả:

$$(10;5,5),\ (12;6,5),\ (11;6,3),\ (9;4,5)$$
$$(9,5;5,3),\ (8;4,0),\ (12;7,0),\ (8,5;5,0)$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

**Bài 4.7.** Để nghiên cứu về quan hệ giữa khoảng cách $X$ (km) từ nhà tới nơi làm việc và thời gian đi lại $Y$ (phút), người lấy một mẫu thực nghiệm và có kết quả

$$(10;45),\ (12;54),\ (11;48),\ (9;45)$$
$$(7;30),\ (8;32),\ (7,5;40),\ (8,5;42)$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

**Bài 4.8.** Người ta lấy một mẫu thực nghiệm của đại lượng ngẫu nhiên $(X,Y)$ trong đó $X$ là số giờ vắng mặt trên lớp và $Y$ là điểm thi của 7 sinh viên và thu được kết quả:

$$(8;6,1),(10;6,0),(15;5,5),(20;4,2),$$
$$(25;1,3),(24;3,5),(21;2,7).$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

**Bài 4.9.** Người ta lấy một mẫu thực nghiệm của đại lượng ngẫu nhiên hai chiều $(X,Y)$ trong đó $X$ là số tiền đầu tư và $Y$ là doanh thu tương ứng của 7 dự án trong lĩnh vực cầu đường (tính theo nghìn tỷ đồng) và thu được kết quả:

$$(2,3;3,08),(4,5;5,12),(3,7;4,63),(7,1;9,04),$$
$$(12;13,2),(8,5;9,6),(10;11,3).$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

**Bài 4.10.** Người ta lấy một mẫu thực nghiệm của đại lượng ngẫu nhiên $(X,Y)$ trong đó $X$ là số tháng được sử dụng của máy in và $Y$ là số trang đã in (tính theo nghìn trang) của 8 máy in văn phòng và thu được kết quả:

$$(8;3,2),(10;4,1),(11;4,6),(14;5,2),$$
$$(18;7,3),(24;8,5),(21;8,7),(15;6,3).$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

## II. Đáp số các bài tập

### Thống kê mô tả và phân phối mẫu

**1.1** c) $\bar{x}=6,18; s^2=4,4521; s=2,11.$

**1.2** b) $\bar{x}=50,8; s^2=59,8367$

**1.3** $\bar{x}=35,9412; s=23,3694.$

**1.4** a) $\bar{x}=170,2467;\ s=3,5817;$ b) 0,01

**1.5** c) $\bar{x}=8,5067;\ s=0,2318$

**1.6** $\bar{x}=5,0231;\ s=0,251$

**1.7** a) $\bar{x}=4,9083;\ s=0,9691;$ c) $\bar{x}=4,9028;\ s=0,9548$

**1.8** $\bar{x}=6,01;\ s=1,339$

**1.9** a) 0,7529; b) 75,29; c) $n\geq 183$

**1.10** 0,8733

**1.11** 0,0008

**1.12** 0,9783

**1.13** $n\leq 54$

**1.14** 0,8962

**1.15** 0,0051

**1.16** $\geq 260$

**1.17** $a=1,945$

**1.18** $d=8,3866$

**1.19** a) $n\geq 16$; b) 0,9544; c) 0,7486; d) xấp xỉ 0,01

### Ước lượng tham số

**2.1** $(22,7459;24,9684).$

**2.2** $(11,1868;13,591).$

**2.3** a) $(23,57;25,53).$ b) $n\geq 665.$

**2.4** $(27,47;29,43).$

**2.5** $(135,4877;139,5123).$

**2.6** $(8,1062;12,3938).$

**2.7** $(4,9608;5,1678).$

**2.8** $2237,317 < \mu < 2282,516$

**2.9** $8,1639 < \mu < 8,4375$

**2.10** $(32,9390; 34,5055)$.

**2.11** $(7,2174; 8,5951)$.

**2.12** $(99,144; 100,656)$.

**2.13** $(1,5337; 1,5536)$.

**2.14** $(5,0412; 5,1922)$.

**2.15** $(47,9776; 53,6224)$.

**2.16** a) $(0,3693; 0,4147)$. b) $(2.585.100; 2.902.900)$.

**2.17** $(435.300; 464.700)$ người.

**2.18** $(3254, 4652)$.

**2.19** $(12.637, 15.682)$.

**2.20** $83,76\%$.

**2.21** $(0,3012; 0,3788)$.

**2.22** a) $(0,4181; 0,5419)$. b) $(0,4064; 0,5536)$

**2.23** a) $(0,0295; 0,1038)$, b) $n \geq 2391$.

**2.24** $0,00065 < \sigma^2 < 0,00185$

**2.25** a) $11,5114 < \mu < 12,4029$, b) $0,7973 < \sigma^2 < 2,3633$

**2.26** a) $s^2 = 32,2333$, b) $13,3896 < \sigma^2 < 138,9368$

**2.27** b) $0,0042 < \sigma^2 < 0,0183$

### Kiểm định giả thuyết thống kê

**3.1** $W_\alpha = (1,64; \infty)$, $t_{qs} = 6,6667$. Bác bỏ giả thuyết $H_o$.

**3.2** $W_\alpha = (-\infty; -1,96) \cup (1,96; \infty)$, $t_{qs} = -19,5$. Bác bỏ $H_o$.

**3.3** $W_\alpha = (-\infty; -1,64)$, $t_{qs} = -2,2$. Bác bỏ giả thuyết $H_o$.

**3.4** $W_\alpha = (1,64; \infty)$, $t_{qs} = 5$. Bác bỏ $H_o$.

**3.5** $W_\alpha = (-\infty; -1,96) \cup (1,96; \infty), t_{qs} = 0,75$. Chưa bác bỏ $H_o$.

**3.6** $H_0 : \mu = 100$ $H_1 : \mu < 100$; $t_{qs} = -7,7499$; bác bỏ $H_0$

**3.7** $W_\alpha = (-\infty; -2,467), t_{qs} = -3,8535$. Bác bỏ $H_o$.

**3.8** $W_\alpha = (-\infty; -2,064) \cup (2,064; \infty), t_{qs} = 2,6582$. Bác bỏ $H_o$.

**3.9** $W_\alpha = (-\infty; -1.699), t_{qs} = -4,6291$. Bác bỏ $H_o$.

**3.10** $W_\alpha = (-\infty; -1.64), t_{qs} = -12,5001$. Bác bỏ $H_o$.

**3.11** $H_0 : \mu = 8,5$ $H_1 : \mu < 8,5$; $t_{qs} = -0,4784$; chưa bác bỏ $H_0$

**3.12** $H_0 : \mu = 2,65$ $H_1 : \mu \neq 2,65$; $t_{qs} = 0,889$; chưa bác bỏ $H_0$

**3.13** $H_0 : \mu = 1,5$ $H_1 : \mu < 1,5$; $t_{qs} = -2,42$; bác bỏ $H_0$

**3.14** $H_0 : \mu = 15$ $H_1 : \mu \neq 15$; $t_{qs} = -4,5297$; bác bỏ $H_0$

**3.15** $H_0 : \mu = 0,04$ $H_1 : \mu > 0,04$; $t_{qs} = 4,4478$; bác bỏ $H_0$

**3.16** $W_\alpha = (-\infty; -1,96) \cup (1,96; \infty), t_{qs} = -0,4193$. Chưa bác bỏ $H_o$.

**3.17** $W_\alpha = (1,64; \infty), t_{qs} = 0,9068$. Chưa bác bỏ $H_o$.

**3.18** $W_\alpha = (-\infty; -2,58) \cup (2,58; \infty), t_{qs} = -2,9234$. Bác bỏ $H_o$.

**3.19** $H_0 : p = 0,7$ $H_1 : p < 0,7$; $t_{qs} = -2,3146$; bác bỏ $H_0$

**3.20** $H_0 : p = 0,05$ $H_1 : p < 0,05$; $t_{qs} = 1,9467$; bác bỏ $H_0$

**3.21** $W_\alpha = (-\infty; -1,64), t_{qs} = -30,356$. Bác bỏ $H_o$.

**3.22** $W_\alpha = (1,64; \infty), t_{qs} = 6,8902$. Bác bỏ $H_o$.

**3.23** $W_\alpha = (-\infty; -1,64), t_{qs} = -8,0978$. Bác bỏ $H_o$.

**3.24** $W_\alpha = (1,64; \infty), t_{qs} = 3,3036$. Bác bỏ $H_o$.

**3.25** $W_\alpha = (-\infty; -1,96) \cup (1,96; \infty), t_{qs} = -1,04998$. Chưa bác bỏ $H_o$.

**3.26** $H_0 : p_1 = p_2$ $H_1 : p_1 < p_2$; $t_{qs} = -1,1915$; chưa bác bỏ $H_0$

### Tương quan và hồi quy

**4.1** a) $r_{tn} = 0,8736$, b) $y = 1,3016\,x + 1,4937$

**4.2** a) $r_{tn} = 0,9950$, b) $y = 1,0593\,x + 0,7167$.

**4.3** a) $y = 2,0893\,x - 4170,2321$; b) $y(2017) = 43,8571$; $(40,7446;\ 46,9697)$.

**4.4** a) $y = 1,8452\,x + 3,6964$; b) $y(10) = 22,1484$; $(20,8491;\ 23,4477)$.

**4.5** a) $y = 233,5943\,x - 462113,901$; b) $y(2017) = 8477,34$; $(8456,6797;\ 8498,0003)$.

**4.6** a) $r_{tn} = 0,9619$, b) $y = 0,6455\,x - 0,9420$

**4.7** a) $r_{tn} = 0,9013$, b) $y = 4,1170\,x + 4,4327$

**4.8** a) $r_{tn} = 0,9928$, b) $y = 2,7549\,x - 2,5899$.

**4.9** a) a) $r_{tn} = 0,9950$, b) $y = 1,0593\,x + 0,7167$.

**4.10** a) $r_{tn} = 0,9786$, b) $y = 0,3602\,x + 0,5399$

## III. Một số đề thi mẫu

Bộ môn Đại số và XSTK trân trọng giới thiệu một số mẫu đề thi kết thúc học phần. Để có sự chuẩn bị tốt cho kỳ thi sinh viên cần lưu ý các điểm sau:

1. Thời gian làm bài đối với mỗi đề thi là 70 phút.
2. Không được mang tài liệu trong phòng thi. Không mang điện thoại vào phòng thi.
3. Mang thẻ sinh viên khi đi thi, mang máy tính và các bảng tra để sử dụng trong giờ thi.
4. Sinh viên không được nháp vào đề thi, phải nộp lại đề thi cùng bài làm khi hết giờ làm bài.

### ĐỀ SỐ 1

**Bài 1.** Một cánh tay rô-bốt sử dụng mắt điện tử để xác định vị trí vật thể trong không gian ba chiều. Quá trình định vị luôn làm nảy sinh ra sai số. Giả sử các sai số $X$, $Y$, $Z$ tương ứng theo ba chiều không gian là các biến ngẫu nhiên độc lập có cùng phân phối chuẩn với trung bình 0 mm và độ lệch tiêu chuẩn 5 mm. Hãy tính xác suất để điểm được định vị bởi mắt điện tử lệch ra khỏi vị trí thực tế không vượt quá 14 mm.

**Bài 2.** Gọi $X$ là mức hao phí nguyên liệu của một loại máy xúc trong một ca làm việc. Để ước lượng mức hao phí nguyên liệu trung bình, người ta theo dõi 64 ca làm việc của các máy xúc và thu được kết quả sau

| $X$ | $50-52$ | $52-54$ | $54-56$ | $56-58$ | $58-60$ |
|---|---|---|---|---|---|
| $n_i$ | 6 | 15 | 20 | 17 | 6 |

Cho biết $X$ là đại lượng ngẫu nhiên có phân phối chuẩn với $\sigma = 5$ lít. Với độ tin cậy 95%, hãy xác định khoảng ước lượng cho mức hao phí nguyên liệu trung bình.

**Bài 3.** Có ý kiến cho rằng tỷ lệ cận thị của học sinh thành thị cao hơn của học sinh nông thôn. Người ta tiến hành kiểm tra 2000 cháu và thu được kết quả

| Nhóm | Số học sinh kiểm tra | Số học sinh bị cận thị |
|---|---|---|
| Thành thị | 900 | 187 |
| Nông thôn | 1100 | 192 |

Với mức ý nghĩa $\alpha = 0,05$ hãy kết luận về ý kiến đã nêu.

**Bài 4.** Số liệu về lượng vận chuyển của một công ty vận tải trong các năm qua (tính theo triệu tấn) là như sau:

| Năm | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|
| Khối lượng | 42 | 44 | $45,5$ | 46 | $38,5$ | 50 | 51 |

a) Tìm hàm xu thế tuyến tính biểu thị năng lực vận chuyển của công ty đó.

b) Dự báo khối lượng vận chuyển năm 2011 và tìm khoảng tin cậy 95% cho giá trị đó.

### ĐỀ SỐ 2

**Bài 1.** Một công ty sở hữu chuỗi 10 cửa hàng tại các địa điểm khác nhau trong thành phố. Doanh thu mỗi ngày của mỗi cửa hàng được coi như một biến ngẫu nhiên chuẩn với giá trị trung bình 50 (triệu đồng) và độ lệch tiêu chuẩn 8 (triệu đồng). Hãy tính xác suất để doanh thu của công ty do chuỗi 10 cửa hàng mang lại lớn hơn 550 triệu đồng.

**Bài 2.** Để điều tra số cá có trong một hồ lớn, cơ quan quản lý đánh bắt 1200 con, đánh dấu rồi thả lại hồ. Lần sau lại bắt 800 con thì trong đó có 114 con có dấu. Hãy ước lượng số lượng cá trong hồ, với độ tin cậy 98%.

**Bài 3.** Định mức thời gian hoàn thành 1 sản phẩm là 24 phút. Liệu có cần thay đổi định mức không nếu theo dõi thời gian hoàn thành 1 sản phẩm ở 25 công nhân ta có bảng số liệu sau:

| $x_i$ t.gian (phút) | $22-23$ | $23-24$ | $24-25$ | $25-26$ | $26-27$ |
|---|---|---|---|---|---|
| $n_i$ số CN | 3 | 7 | 8 | 4 | 3 |

Yêu cầu kết luận với mức ý nghĩa $\alpha = 0,05$. Biết rằng thời gian hoàn thành 1 sản phẩm "X" là đại lượng ngẫu nhiên có phân phối chuẩn.

**Bài 4.** Số liệu về số lượt nghe một bài hát của ca sĩ A sau khi bài hát được đưa lên youtube như sau:

| Ngày thứ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Số lượt nghe | 2112 | 2523 | 2265 | 2032 | 1983 | 1928 | 1765 |

a) Tìm hàm xu thế tuyến tính biểu thị số lượt nghe của bài hát theo số ngày đưa lên youtube.

b) Dự báo lượt nghe bài hát ở ngày thứ 10 và tìm khoảng ước lượng cho giá trị này với độ tin cậy 90%.

### ĐỀ SỐ 3

**Bài 1.** Điểm thi môn Vật lý trong kỳ thi tuyển đầu vào đại học năm 2016 được xem như biến ngẫu nhiên chuẩn với kỳ vọng $\mu = 6$ điểm và độ lệch tiêu chuẩn $\sigma = 2,5$ điểm. Chọn ngẫu nhiên một mẫu gồm $n$ điểm thi. Xác định giá trị của $n$ sao cho trung bình mẫu lệch so với kỳ vọng $\mu$ không quá 1 điểm với xác suất lớn hơn 95%.

**Bài 2.** Để khảo sát chất lượng đóng chai của một máy đóng chai tự động, người ta chọn ngẫu nhiên 16 chai trên dây chuyền sản xuất, đo lượng chất lỏng trong chai và tính được độ lệch tiêu chuẩn mẫu là $s = 0,0525$ (lít). Nếu độ lệch quá lớn thì sẽ ảnh hưởng tới chất lượng của việc đóng chai, tức là sẽ có những chai quá ít hoặc quá nhiều. Giả sử lượng chất lỏng được đóng ở mỗi chai là biến ngẫu nhiên chuẩn. Hãy tìm khoảng ước lượng cho phương sai của lượng chất lỏng đóng chai với độ tin cậy 90%.

**Bài 3.** Trọng lượng những bao phân đạm do nhà máy sản xuất ra là 1 biến ngẫu nhiên $X$ tuân theo luật phân bố chuẩn, với $EX = 50$ kg. Khách hàng cho rằng trọng lượng đó đã thay đổi và ít hơn 50 kg. Cân thử 100 bao, ta có bảng số liệu dưới đây:

| $x_i$ | $47-48$ | $48-49$ | $49-50$ | $50-51$ | $51-52$ |
|---|---|---|---|---|---|
| $n_i$ | 22 | 35 | 30 | 8 | 5 |

Với mức ý nghĩa $\alpha = 0,05$, hãy cho kết luận về ý kiến của khách hàng.

**Bài 4.** Để nghiên cứu về quan hệ giữa khối lượng đào đắp $X$ (nghìn $m^2$) và thời gian thi công $Y$ (giờ) người ta lấy một mẫu thực nghiệm và thu được kết quả:

$$(10;25),\ (12;28),\ (11;27),\ (9;23)$$
$$(9,5;24),\ (8;20),\ (12;30),\ (8,5;22)$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

## ĐỀ SỐ 4

**Bài 1.** Một công ty lữ hành thường tổ chức tour du lịch thám hiểm đến các vùng đất mới. Để đảm bảo lợi nhuận và an toàn cho du khách, mỗi tour như vậy được tổ chức cho khoảng 25 người. Tuy nhiên trên website của công ty lại cho phép đặt trước nhiều hơn 25 chỗ. Lý do là chỉ có khoảng 70% khách đặt chỗ thực sự mua tour. Giả sử rằng website của công ty cho phép đặt chỗ lên đến 32 khách. Hãy tính xác suất để có nhiều hơn 25 khách đặt chỗ và xác nhận mua tour.

**Bài 2.** Để ước lượng năng suất trung bình của một giống lúa đặc sản của một huyện A được canh tác theo phương thức mới, người ta chọn ngẫu nhiên 100 thửa ruộng trồng giống lúa đó để thu hoạch và thu được bảng số liệu sau:

| $X$ | $40-41$ | $41-42$ | $42-43$ | $43-44$ | $44-45$ |
|---|---|---|---|---|---|
| $n_i$ | 12 | 20 | 36 | 22 | 10 |

Giả sử đại lượng ngẫu nhiên chỉ năng suất $X$ tuân theo luật chuẩn $N(\mu,\sigma^2)$. Tìm khoảng tin cậy của $\mu=E(X)$, với độ tin cậy $\gamma=0,95$.

**Bài 3.** Tỷ lệ phế phẩm do một máy tự động sản xuất là 8%. Kiểm tra ngẫu nhiên 400 sản phẩm thấy có 36 phế phẩm. Từ đó có ý kiến cho rằng tỷ lệ phế phẩm do máy đó sản xuất có chiều hướng tăng lên. Hãy kết luận ý kiến nêu trên với mức ý nghĩa $\alpha=0,05$.

**Bài 4.** Người ta lấy một mẫu thực nghiệm của đại lượng ngẫu nhiên hai chiều $(X,Y)$ và thu được kết quả:

| $X$ | $4,15$ | $4,46$ | $4,65$ | $4,98$ | $5,12$ | $5,25$ |
|---|---|---|---|---|---|---|
| $Y$ | $18,2$ | $19,6$ | $19,7$ | $20,1$ | $22,3$ | $22,9$ |

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

## ĐỀ SỐ 5

**Bài 1.** Một kết cấu thép được dùng làm giá để hàng có thể chịu được tải trọng 5 tấn. Nếu khối lượng của mỗi kiện hàng là biến ngẫu nhiên với trung bình 100 kg và độ lệch tiêu chuẩn 15 kg thì số lượng kiện hàng tối đa có thể xếp lên giá đỡ để kết cấu vẫn an toàn với xác suất ít nhất 95% là bao nhiêu?

**Bài 2.** Để nghiệm thu đoạn đường do bên B thi công, bên A tiến hành khoan thăm dò 16 điểm ngẫu nhiên trên con đường và thu được dãy số liệu (tính bằng mm) chỉ độ dày của lớp bê tông nhựa trải đường như sau:

143; 137; 135; 136; 132; 143; 139; 138;
136; 141; 138; 142; 140; 140; 139; 137.

Với độ tin cậy 95% hãy tìm khoảng ước lượng chiều dầy trung bình của lớp bê tông nhựa đó. Biết chiều dày của lớp bê tông có phân phối chuẩn.

**Bài 3.** Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm $X$ là biến chuẩn với kỳ vọng $\mu=100$ g, độ lệch chuẩn $\sigma=2,2$ g. Qua một thời gian sản xuất người ta nghi ngờ trọng lượng sản phẩm có xu hướng tăng lên. Cân thử 100 sản phẩm ta có bảng số liệu sau:

| $x_i$ (gam) | $96-98$ | $98-100$ | $100-102$ | $102-104$ | $104-106$ |
|---|---|---|---|---|---|
| $n_i$ số SP | 8 | 22 | 35 | 25 | 10 |

Với mức ý nghĩa $\alpha=0,05$, hãy kết luận về điều nghi ngờ nói trên.

**Bài 4.** Người ta lấy một mẫu thực nghiệm của đại lượng ngẫu nhiên hai chiều $(X,Y)$ trong đó $X$ là số tiền đầu tư và $Y$ là doanh thu tương ứng của 7 dự án trong lĩnh vực cầu đường (tính theo nghìn tỷ đồng) và thu được kết quả:

$$(3,2;4,5),(3,8;4,8),(3,7;4,63),(6,6;9,8),$$
$$(7;10,2),(8,5;11,6),(12;14,3).$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

## III. Lời giải mẫu của các bài tập

**Bài 1.** Chiều cao sinh viên của sinh viên trường Đại học GTVT được xấp xỉ bởi phân phối chuẩn với trung bình $\mu = 175$ (cm) và độ lệch tiêu chuẩn $\sigma = 10$ (cm). Tính xác suất để chiều cao trung bình của 16 sinh viên được chọn ngẫu nhiên nằm trong khoảng (173; 177).

**Lời giải.** Gọi $X_i$ là chiều cao của sinh viên thứ $i$ được chọn ($i = 1, 2, \ldots, 16$). Theo đề bài, $X_i \sim N(\mu = 175; \sigma^2 = 10^2)$. Chiều cao trung bình của 16 sinh viên là

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_{16}}{16}.$$

$\bar{X}$ có phân phối chuẩn với kỳ vọng $\mu = 175$ (cm) và độ lệch tiêu chuẩn $\dfrac{\sigma}{\sqrt{n}} = 2,5$.

Vậy, xác suất để chiều cao trung bình của 16 sinh viên nằm trong khoảng (173,177) là

$$\begin{aligned} P\left(173 < \bar{X} < 177\right) &= P\left(\frac{173-175}{2,5} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{177-175}{2,5}\right) \\ &= P\left(-0,8 < Z < 0,8\right) \end{aligned}$$

với $Z$ là biến ngẫu nhiên chuẩn tắc. Do đó

$$\begin{aligned} P\left(173 < \bar{X} < 177\right) &= \Phi\left(Z < 0,8\right) - \Phi\left(Z < -0,8\right) \\ &= 0,7881 - 0,1867 = 0,6014 \end{aligned}$$

**Bài 2.** Thời gian để một bộ vi xử lý trung tâm (CPU) giải quyết một loại công việc được xem như biến ngẫu nhiên chuẩn với kỳ vọng 20 (giây) và độ lệch tiêu chuẩn là 4 (giây). Nếu quan sát ngẫu nhiên 9 công việc loại này thì hãy tìm $a$ để phương sai mẫu không vượt quá $a$ với xác suất bằng 0,9.

**Lời giải.** Ta có kích thước mẫu $n = 9$ và phương sai tập chính $\sigma^2 = 4^2$. Gọi $S^2$ là phương sai mẫu ngẫu nhiên. Giá trị $a$ cần tìm thỏa mãn hệ thức

$$P(S^2 \leq a) = 0,9$$

Dựa vào phân phối của phương sai mẫu, biến đổi tương đương hệ thức trên ta thu được

$$\begin{aligned} 0,9 &= P\left(\frac{(n-1)S^2}{\sigma^2} \leq \frac{8a}{4^2}\right) \\ &= P\left(\chi_8^2 \leq \frac{a}{2}\right) \\ &= 1 - P\left(\chi_8^2 > \frac{a}{2}\right). \end{aligned}$$

Do đó

$$P\left(\chi_8^2 > \frac{a}{2}\right) = 1 - 0,9 = 0,1.$$

Vậy $\dfrac{a}{2} = 13,362$ (tra bảng phân vị khi bình phương) và ta tìm được $a = 26,724$.

**Bài 3.** Để kiểm tra mức xăng hao phí của một loại xe ô tô. Người ta chọn ngẫu nhiên 28 chiếc xe và cho chạy trên cùng một đoạn đường 300 km. Kết quả thu được như sau:

| $X$ | $4,6-4,8$ | $4,8-5,0$ | $5,0-5,2$ | $5,2-5,4$ | $5,4-5,6$ |
|---|---|---|---|---|---|
| $n_i$ | 6 | 5 | 9 | 4 | 4 |

Với độ tin cậy $\gamma = 0,95$, hãy tìm khoảng ước lượng của lượng xăng hao phí trung bình. Biết mức xăng hao phí có phân phối chuẩn.

**Lời giải.** Gọi $X$ là đại lượng ngẫu nhiên chỉ lượng xăng hao phí. Theo giả thiết $X \sim N(\mu, \sigma^2)$ với $\mu = EX$ là tham số cần phải ước lượng và phương sai $\mathbb{V}[X] = \sigma^2$ chưa biết. Để ước lượng $\mu$ ta xét đại lượng ngẫu nhiên

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S}$$

Ứng với $n = 28 < 30$, đại lượng ngẫu nhiên $T$ có phân phối student với $n-1$ bậc tự do.

Với độ tin cậy $\gamma$ cho trước ta đặt $\alpha = 1 - \gamma$. Khi đó ta cần sử dụng phân vị $t_{(n-1,\alpha/2)}$ được xác định bởi ràng buộc

$$P(|T| < t_{(n-1,\alpha/2)}) = \gamma$$

Biến đổi tương đương, ta xác định được công thức ước lượng

$$\left(\overline{X} - \frac{t_{(n-1,\alpha/2)}S}{\sqrt{n}}; \overline{X} + \frac{t_{(n-1,\alpha/2)}S}{\sqrt{n}}\right)$$

Với $\gamma = 0,95$ và $n = 28$ ta có $\alpha = 1-\gamma = 0,05$ và $t_{(n-1,\alpha/2)} = t_{(27;0,025)} = 2,052$.

Tiếp theo chúng ta tính các đặc trưng thực nghiệm từ mẫu được cho. Đặt $x_i = \dfrac{a_{i-1} + a_i}{2}$ và lập bảng tính như sau

| $x_i$ | $n_i$ | $n_i x_i$ | $n_i x_i^2$ |
|---|---|---|---|
| 4,7 | 6 | 28,2 | 132,54 |
| 4,9 | 5 | 24,5 | 120,05 |
| 5,1 | 9 | 45,9 | 234,09 |
| 5,3 | 4 | 21,2 | 112,36 |
| 5,5 | 4 | 22 | 121 |
| $\Sigma$ | 28 | 141,8 | 720,04 |

Từ bảng trên ta thu được các đặc trưng thực nghiệm như sau:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{141,8}{28} \approx 5,0643$$

$$s^2 = \tfrac{n}{n-1}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2\right] = \tfrac{28}{27}\left[\tfrac{720,04}{28} - (5,0643)^2\right] \approx 0,0711,$$

$$s = \sqrt{s^2} = \sqrt{0,0711} \approx 0,2667$$

Như vậy thay các số liệu thực nghiệm vào công thức ước lượng ta thu được khoảng ước lượng thực nghiệm

$$\left(5,0643 - \frac{2,052 \times 0,2667}{\sqrt{28}}; 5,0643 + \frac{2,052 \times 0,2667}{\sqrt{28}}\right)$$

Rút gọn chúng ta thu được kết quả

$$(4,9609; 5,1677)$$

**Bài 4.** Năng suất lúa trên một đơn vị diện tích tuân theo luật chuẩn. Để ước lượng năng suất lúa trung bình của một giống lúa mới người ta trồng thử trên 100 thửa ruộng trong điều kiện như nhau và thu được bảng số liệu sau :

| $x_i$ (tạ/ha) | $40-42$ | $42-44$ | $44-46$ | $46-48$ | $48-50$ | $50-52$ |
|---|---|---|---|---|---|---|
| $n_i$ | 7 | 13 | 25 | 35 | 15 | 5 |

Với độ tin cậy 95% hãy ước lượng năng suất lúa trung bình của giống lúa mới đó.

**Lời giải.** Gọi $X$ là đại lượng ngẫu nhiên chỉ năng suất thu hoạch. Theo giả thiết $X \sim N(\mu, \sigma^2)$ với $\mu = EX$ là tham số cần phải ước lượng và phương sai $\mathbb{V}[X] = \sigma^2$ chưa biết. Để ước lượng $\mu$ ta xét đại lượng ngẫu nhiên

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S}$$

Ứng với $n = 100 > 30$, đại lượng ngẫu nhiên $T$ có phân phối xác suất xấp xỉ luật chuẩn tắc $N(0,1)$.
Với độ tin cậy $\gamma$ cho trước ta sử dụng phân vị $z_{\alpha/2} = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$ $(\alpha = 1-\gamma)$ được xác định bởi đẳng thức:

$$P(|T| < z_{\alpha/2}) = \gamma$$

Biến đổi tương đương, ta xác định được công thức ước lượng

$$\left(\overline{X} - \frac{z_{\alpha/2}S}{\sqrt{n}}; \overline{X} + \frac{z_{\alpha/2}S}{\sqrt{n}}\right)$$

Thay $\gamma = 0,95$ ta thu được $z_{0,025} = \Phi^{-1}(0,975) = 1,96$. Tiếp theo chúng ta tính các đặc trưng thực nghiệm từ mẫu được cho. Đặt $x_i = \dfrac{a_{i-1} + a_i}{2}$ và lập bảng tính như sau

| $x_i$ | $n_i$ | $n_i x_i$ | $n_i x_i^2$ |
|---|---|---|---|
| 41 | 7 | 287 | 11.767 |
| 43 | 13 | 559 | 24.037 |
| 45 | 25 | 1125 | 50.625 |
| 47 | 35 | 1645 | 77.315 |
| 49 | 15 | 735 | 36.015 |
| 51 | 5 | 255 | 13.005 |
| $\Sigma$ | 100 | 4.606 | 212.764 |

Từ bảng trên ta thu được các đặc trưng thực nghiệm như sau:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{4606}{100} \approx 46,06$$

$$s^2 = \tfrac{n}{n-1}\left[\tfrac{1}{n}\textstyle\sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2\right] = \tfrac{100}{99}\left[\tfrac{212764}{100} - (46,06)^2\right] \approx 6,1782,$$

$$s = \sqrt{s^2} = \sqrt{6,1782} \approx 2,4856$$

Như vậy thay các số liệu thực nghiệm vào công thức ước lượng ta thu được khoảng ước lượng thực nghiệm

$$\left(46,06 - \frac{1,96 \times 2,4856}{\sqrt{100}}; 46,06 + \frac{1,96 \times 2,4856}{\sqrt{100}}\right)$$

Rút gọn chúng ta thu được kết quả

$$(45,5728; 46,5472)$$

**Bài 5.** Phỏng vấn 2500 người được chọn ngẫu nhiên trong một thành phố. Kết quả cho thấy có 980 người thường xuyên sử dụng Internet.
a) Với độ tin cậy 0,98 hãy ước lượng tỷ lệ người dân trong thành phố có sử dụng Internet.
b) Nếu dân số của thành phố là 7 triệu người thì với độ tin cậy trên số dân sử dụng Internet trong thành phố là bao nhiêu?

**Lời giải.** Ta ký hiệu tỷ lệ dân sử dụng Internet là $p$. Theo giả thiết kích thước của mẫu thực nghiệm là $n = 2500$. Tần suất thực nghiệm tương ứng là

$$f = \frac{980}{2500} = 0,392$$

Kiểm tra điều kiện đối với kích thước $n$:

$$nf = 2500 \times 0,392 = 980 > 10,$$
$$n(1-f) = 2500(1-0,392) = 1520 > 10$$

Ký hiệu tần suất ngẫu nhiên là $\hat{p}$. Khi đó ta chọn đại lượng ngẫu nhiên

$$T = \frac{(\hat{p} - p)\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}}$$

Do $n$ đủ lớn nên $T$ có phân phối xác suất xấp xỉ luật chuẩn tắc $N(0,1)$.
Với độ tin cậy $\gamma$ cho trước ta sử dụng phân vị $z_{\alpha/2} = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$ $(\alpha = 1-\gamma)$ được xác định bởi đẳng thức:

$$P(|T| < z_{\alpha/2}) = \gamma$$

Biến đổi tương đương, ta xác định được công thức ước lượng

$$\left(\hat{p} - \frac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}; \hat{p} + \frac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

Thay $\gamma = 0,98$ ta thu được $z_{0,01} = \Phi^{-1}(0,99) = 2,33$. Thay các giá trị thực nghiệm $f = 0,392, n = 2500, z_{0,01} = 2,33$ ta thu được khoảng ước lượng:

$$\left(0,392 - \frac{2,33\sqrt{0,392(1-0,392)}}{\sqrt{2500}};\right.$$
$$\left.0,392 + \frac{2,33\sqrt{0,392(1-0,392)}}{\sqrt{2500}}\right)$$

Rút gọn chúng ta thu được kết quả

$$(0,3668;\ 0,4172)$$

b) Ký hiệu $N$ là số dân của thành phố. Ký hiệu $M$ là số dân thành phố sử dụng Internet. Khi đó tỷ lệ dân sử dụng Internet là

$$p = \frac{M}{N}$$

Theo giả thiết ta có $N = 7.10^6$. Áp dụng kết quả câu a) ta có:

$$0,3668 < p < 0,4172$$
$$\Leftrightarrow 0,3668 < \frac{M}{7.10^6} < 0,4172$$
$$\Leftrightarrow 2.567.600 < M < 2.920.400$$

Vậy số dân sử dụng Internet nằm trong khoảng (2.567.600; 2.920.400).

**Bài 6.** Lượng đường trong mỗi hộp sữa tươi LiF của Công ty Cổ phần Sữa Quốc tế được xem như một biến ngẫu nhiên có phân phối chuẩn với phương sai $\sigma^2$. Kiểm tra ngẫu nhiên 10 hộp sữa loại này, người ta xác định được độ lệch tiêu chuẩn mẫu của lượng đường trong mỗi hộp là $s = 2,5$ miligam. Hãy tìm khoảng tin cậy 95% cho phương sai tập chính $\sigma^2$.

**Lời giải.** Gọi $X$ là biến ngẫu nhiên chỉ lượng đường trong mỗi hộp sữa LiF. Ta có $X \sim N(\mu, \sigma^2)$. Ta phải tìm khoảng ước lượng cho phương sai $\sigma^2$.

Ký hiệu $S^2$ là phương sai mẫu ngẫu nhiên kích thước $n$ được chọn. Ta có tiêu chuẩn ước lượng

$$T = \frac{(n-1)S^2}{\sigma^2}.$$

$T$ tuân theo luật phân phối khi bình phương với $n-1$ bậc tự do.
Với độ tin cậy $\gamma = 1-\alpha$, ta tìm được các phân vị $\chi^2_{n-1,1-\alpha/2}$ và $\chi^2_{n-1,\alpha/2}$ thỏa mãn đẳng thức

$$P\left(\chi^2_{n-1,1-\alpha/2} < T < \chi^2_{n-1,\alpha/2}\right) = \gamma.$$

Biến đổi tương đương, ta xác định được khoảng ước lượng

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}; \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right).$$

Với $\gamma = 0,95;\ n = 10$, ta suy ra $\alpha = 1-0,95 = 0,05$ và $\chi^2_{9;0,975} = 2,7;\ \chi^2_{9;0,025} = 19,023$. Thay phương sai mẫu $s^2 = 2,5^2$ vào biểu thức ta thu được khoảng ước lượng

$$\left(\frac{9\times 2,5^2}{19,023}; \frac{9\times 2,5^2}{2,7}\right) = (2,9569; 20,8333).$$

**Bài 7.** Hao phí nguyên liệu cho 1 đơn vị sản phẩm là đại lượng ngẫu nhiên chuẩn với độ lệch chuẩn $\sigma = 2$ và hao phí trung bình là 65 gram. Nghi ngờ máy móc trục trặc làm cho hao phí nguyên liệu cho 1 đơn vị sản phẩm tăng lên. Người ta kiểm tra 36 sản phẩm thì có bảng số liệu sau:

| $x_i$ (gram) | $60-62$ | $62-64$ | $64-66$ | $66-68$ | $68-70$ |
|---|---|---|---|---|---|
| $n_i$ số SP | 5 | 7 | 10 | 8 | 6 |

Hãy kết luận ý kiến nêu trên với mức ý nghĩa $\alpha = 0,05$.

**Lời giải.** Gọi $X$ là đại lượng ngẫu nhiên chỉ mức hao phí nguyên liệu để sản xuất một sản phẩm. Theo giả thiết $X \sim N(\mu,\sigma^2)$ với $\mu = EX$ là mức hao phí trung bình và phương sai $\mathbb{V}[X] = \sigma^2$ đã biết. Theo yêu cầu được đưa ra chúng ta có bài toán kiểm định một phía.

Giả thuyết $H_o:\quad \mu = 65$
Đối thuyết $H_1:\quad \mu > 65$

Do đã biết phương sai $\mathbb{V}[X] = \sigma^2 = 4$ nên chúng ta chọn tiêu chuẩn kiểm định

$$T = \frac{(\overline{X}-65)\sqrt{n}}{\sigma}$$

Trên cơ sở giả thuyết $H_o$ đúng thì đại lượng ngẫu nhiên $T$ có phân phối xác suất chuẩn tắc $N(0,1)$.
Với mức ý nghĩa $\alpha$ và đối thuyết $H_1:\ \mu > 65$, ta chọn miền bác bỏ là

$$W_\alpha = (z_\alpha; +\infty)$$

trong đó $z_\alpha = \Phi^{-1}(1-\alpha)$. Thay $\alpha = 0,05$ ta có $z_{0,05} = \Phi^{-1}(0,95) = 1,64$ (tra bảng). Vậy ta xây dựng được

$$W_\alpha = (1,64; +\infty)$$

Tiếp theo ta cần tính giá trị thực nghiệm của tiêu chuẩn kiểm định $T$. Từ mẫu thực nghiệm ta đặt $x_i = \dfrac{a_{i-1}+a_i}{2}$ và lập bảng tính như sau

| $x_i$ | $n_i$ | $n_ix_i$ |
|---|---|---|
| 61 | 5 | 305 |
| 63 | 7 | 441 |
| 65 | 10 | 650 |
| 67 | 8 | 536 |
| 69 | 6 | 414 |
| $\Sigma$ | 36 | 2.346 |

Từ bảng trên ta thu được các đặc trưng thực nghiệm như sau:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_ix_i = \frac{2346}{36} \approx 65,1667$$

Như vậy ta nhận được

$$t_{qs} = \frac{(\bar{x}-65)\sqrt{n}}{\sigma} = \frac{(65,1667-65)\sqrt{36}}{2} = 0,5$$

Do $t_{qs} \notin W_\alpha$ nên ta chưa có cơ sở bác bỏ $H_o$.

**Bài 8.** Định mức thời gian hoàn thành 1 sản phẩm là 14 phút. Liệu có cần thay đổi định mức không nếu theo dõi thời gian hoàn thành 1 sản phẩm ở 25 công nhân ta có bảng số liệu sau:

| $x_i$ t.gian (phút) | $10-12$ | $12-14$ | $14-16$ | $16-18$ | $18-20$ |
|---|---|---|---|---|---|
| $n_i$ số CN | 2 | 6 | 7 | 7 | 3 |

Yêu cầu kết luận với mức ý nghĩa $\alpha = 0,05$. Biết rằng thời gian hoàn thành 1 sản phẩm "X" là đại lượng ngẫu nhiên có phân phối chuẩn.

**Lời giải.** Gọi $X$ là đại lượng ngẫu nhiên chỉ thời gian để hoàn thành một sản phẩm. Theo giả thiết $X \sim N(\mu,\sigma^2)$ với $\mu = EX$ là thời gian trung bình để hoàn thành một sản phẩm và phương sai $\mathbb{V}[X] = \sigma^2$ chưa biết.
Theo yêu cầu được đưa ra chúng ta có bài toán kiểm định hai phía.

Giả thuyết $H_o:\quad \mu = 14$
Đối thuyết $H_1:\quad \mu \neq 14$

Do chưa biết phương sai $\mathbb{V}[X] = \sigma^2$ nên chúng ta chọn tiêu chuẩn kiểm định

$$T = \frac{(\overline{X}-14)\sqrt{n}}{S}$$

Trên cơ sở giả thuyết $H_o$ đúng thì đại lượng ngẫu nhiên $T$ có phân phối xác suất theo luật student với $n-1$ bậc tự do.
Với mức ý nghĩa $\alpha$ và đối thuyết $H_1:\ \mu \neq 14$ ta chọn miền bác bỏ là

$$W_\alpha = (-\infty; -t_{(n-1,\alpha/2)}) \cup (t_{(n-1,\alpha/2)}; +\infty)$$

trong đó $t_{(n-1,\alpha/2)}$ là phân vị student. Thay $\alpha = 0,05$ và $n = 25$ ta có $t_{(n-1,\alpha/2)} = t_{(24;0,025)} = 2,064$ (tra bảng). Vậy ta xây dựng được

$$W_\alpha = (-\infty; -2,064) \cup (2,064; +\infty)$$

Tiếp theo ta cần tính giá trị thực nghiệm của tiêu chuẩn kiểm định $T$. Từ mẫu thực nghiệm ta đặt $x_i = \dfrac{a_{i-1}+a_i}{2}$ và lập bảng tính như sau

| $x_i$ | $n_i$ | $n_ix_i$ | $n_ix_i^2$ |
|---|---|---|---|
| 11 | 2 | 22 | 242 |
| 13 | 6 | 78 | 1014 |
| 15 | 7 | 105 | 1575 |
| 17 | 7 | 119 | 2023 |
| 19 | 3 | 57 | 1083 |
| $\sum$ | 25 | 381 | 5937 |

Từ bảng trên ta thu được các đặc trưng thực nghiệm như sau:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{381}{25} = 15,24$$

$$s^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2\right] = \frac{25}{24}\left[\frac{5937}{25} - (15,24)^2\right] \approx 5,44,$$

$$s = \sqrt{s^2} = \sqrt{5,44} \approx 2,3324$$

Như vậy thay các số liệu thực nghiệm để tính giá trị thực nghiệm của tiêu chuẩn $T$ ta nhận được

$$t_{qs} = \frac{(\bar{x}-14)\sqrt{n}}{s} = \frac{(15,24-15)\sqrt{25}}{2,3324} = 2,6582$$

Do $t_{qs} \in W_\alpha$ nên ta bác bỏ $H_o$ và thay thế bởi $H_1$.

**Bài 9.** Trọng lượng những bao phân đạm do nhà máy sản xuất ra là 1 biến ngẫu nhiên $X$ tuân theo luật phân bố chuẩn, với $EX = 50$ kg. Khách hàng cho rằng trọng lượng đó đã thay đổi và ít hơn 50 kg. Cân thử 100 bao, ta có bảng số liệu dưới đây:

| $x_i$ | $47-48$ | $48-49$ | $49-50$ | $50-51$ | $51-52$ |
|---|---|---|---|---|---|
| $n_i$ | 30 | 40 | 20 | 5 | 5 |

**Lời giải.** Gọi $X$ là đại lượng ngẫu nhiên chỉ trọng lượng của bao phân đạm. Theo giả thiết $X \sim N(\mu, \sigma^2)$ với $\mu = EX$ là trọng lượng trung bình và phương sai $\mathbb{V}[X] = \sigma^2$ chưa biết. Theo yêu cầu được đưa ra chúng ta có bài toán kiểm định một phía.

Giả thuyết $H_o: \quad a = 50$
Đối thuyết $H_1: \quad a < 50$

Do chưa biết phương sai $\mathbb{V}[X] = \sigma^2$ nên chúng ta chọn tiêu chuẩn kiểm định

$$T = \frac{(\overline{X}-50)\sqrt{n}}{S}$$

Trên cơ sở giả thuyết $H_o$ đúng thì đại lượng ngẫu nhiên $T$ có phân phối xác suất xấp xỉ luật phân phối chuẩn tắc $N(0,1)$. Với mức ý nghĩa $\alpha$ và đối thuyết $H_1: \ \mu < 50$, ta chọn miền bác bỏ là

$$W_\alpha = (-\infty; -z_\alpha)$$

trong đó $z_\alpha = \Phi^{-1}(1-\alpha)$. Thay $\alpha = 0,05$ ta có $z_{0,05} = \Phi^{-1}(0,95) = 1,64$ (tra bảng).
Vậy ta xây dựng được

$$W_\alpha = (-\infty; -1,64)$$

Tiếp theo ta cần tính giá trị thực nghiệm của tiêu chuẩn kiểm định $T$. Từ mẫu thực nghiệm ta đặt $x_i = \dfrac{a_{i-1}+a_i}{2}$ và lập bảng tính như sau

| $x_i$ | $n_i$ | $n_i x_i$ | $n_i x_i^2$ |
|---|---|---|---|
| 47,5 | 30 | 1425 | 67687,5 |
| 48,5 | 40 | 1940 | 94090 |
| 49,5 | 20 | 990 | 49005 |
| 50,5 | 5 | 252,5 | 12751,25 |
| 51,5 | 5 | 257,5 | 13261,25 |
| $\sum$ | 100 | 4865 | 236795 |

Từ bảng trên ta thu được các đặc trưng thực nghiệm như sau:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{4865}{100} = 48,65$$

$$s^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2\right] = \frac{100}{99}\left[\frac{236795}{100} - (48,65)^2\right] \approx 1,1389,$$

$$s = \sqrt{s^2} = \sqrt{1,1389} \approx 1,0672$$

Như vậy thay các số liệu thực nghiệm để tính giá trị thực nghiệm của tiêu chuẩn $T$ ta nhận được

$$t_{qs} = \frac{(\bar{x}-50)\sqrt{n}}{s} = \frac{(48,65-50)\sqrt{100}}{1,0672} = -12,6501$$

Do $t_{qs} \in W_\alpha$ nên ta bác bỏ $H_o$ và thay thế bởi $H_1$.

**Bài 10.** Tỷ lệ phế phẩm do một máy tự động sản xuất là 6%. Kiểm tra ngẫu nhiên 400 sản phẩm thấy có 27 phế phẩm. Từ đó có ý kiến cho rằng tỷ lệ phế phẩm do máy đó sản xuất có chiều hướng tăng lên. Hãy kết luận ý kiến nêu trên với mức ý nghĩa $\alpha = 0,05$.

**Lời giải.** Gọi $p$ là tỷ lệ phế phẩm của máy. Đặt $p_0 = 0,06$. Theo yêu cầu chúng ta có bài toán kiểm định

Giả thuyết $H_o: \quad p = 0,06$
Đối thuyết $H_1: \quad p > 0,06$

Từ số liệu được cho ta có kích thước mẫu thực nghiệm $n = 400$. Kiểm tra điều kiện đối với kích thước $n$

$$np_0 = 400 \times 0,06 = 24 > 5$$
$$n(1-p_0) = 400 \times (1-0,06) = 376 > 5$$

Ta chọn tiêu chuẩn kiểm định là

$$T = \frac{(\hat{p}-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$$

trong đó $\hat{p}$ là tần suất ngẫu nhiên.
Vì $n$ đủ lớn nên trên cơ sở giả thuyết $H_o$ đúng đại lượng ngẫu nhiên $T$ có phân phối xác suất xấp xỉ luật phân phối chuẩn tắc $N(0,1)$.
Với mức ý nghĩa $\alpha$ và đối thuyết của bài toán là $H_1: \ p > p_0$, ta chọn miền bác bỏ là

$$W_\alpha = (z_\alpha; +\infty)$$

trong đó $z_\alpha = \Phi^{-1}(1-\alpha)$. Thay $\alpha = 0,05$ ta có $z_{0,05} = \Phi^{-1}(0,95) = 1,64$ (tra bảng). Vậy ta xây dựng được

$$W_\alpha = (1,64; +\infty)$$

Tiếp theo ta cần tính giá trị thực nghiệm của tiêu chuẩn kiểm định $T$. Từ mẫu thực nghiệm ta có tần suất thực nghiệm là

$$f = \frac{m}{n} = \frac{27}{400} = 0,0675$$

Thay các giá trị thực nghiệm $f = 0,0675, \ n = 400$ và $p_0 = 0,06$ ta thu được giá trị thực nghiệm của tiêu chuẩn $T$ như sau:

$$t_{qs} = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \frac{(0,0675-0,06)\sqrt{400}}{\sqrt{0,06(1-0,06)}} \approx 0,6316$$

Do $t_{qs} \notin W_\alpha$ nên ta chưa có cơ sở bác bỏ $H_o$.

**Bài 11.** Một mẫu quan sát của đại lượng ngẫu nhiên hai chiều $(X, Y)$ có giá trị như sau

$$(2,1; 4,12), \ (2,2; 4,34), \ (2,4; 4,56), \ (2,5; 4,63)$$
$$(2,25; 4,38), \ (2,45; 4,75), \ (2,16; 4,4), \ (2,34; 4,62)$$

a) Hãy tính hệ số tương quan thực nghiệm của mẫu trên.
b) Hãy xây dựng hàm hồi quy tuyến tính của $Y$ theo $X$.

c) Ước lượng giá trị $y_0 = E[Y|2,9]$ và tìm khoảng tin cậy 90% cho giá trị đó.

**Lời giải.** Từ số liệu được cho ta lập bảng tính

| STT | $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_iy_i$ |
|---|---|---|---|---|---|
| 1 | 2,1 | 4,12 | 4,41 | 16,9744 | 8,652 |
| 2 | 2,2 | 4,34 | 4,84 | 18,8356 | 9,548 |
| 3 | 2,4 | 4,56 | 5,76 | 20,7936 | 10,944 |
| 4 | 2,5 | 4,63 | 6,25 | 21,4369 | 11,575 |
| 5 | 2,25 | 4,38 | 5,0625 | 19,1844 | 9,855 |
| 6 | 2,45 | 4,75 | 6,0025 | 22,5625 | 11,6375 |
| 7 | 2,16 | 4,4 | 4.6656 | 19,36 | 9,504 |
| 8 | 2,34 | 4,62 | 5,4756 | 21,3444 | 10,8108 |
| $\sum$ | 18,4 | 35,8 | 42,4662 | 160,4918 | 82,5263 |

Từ bảng tính ta được hệ số tương quan thực nghiệm

$$\bar{x} = \frac{\sum x_i}{n} = \frac{18,4}{8} = 2,3$$
$$\bar{y} = \frac{\sum y_i}{n} = \frac{35,8}{8} = 4,475$$
$$S_{xy} = \sum x_iy_i - \frac{1}{n}\sum x_i \sum y_i = 82,5263 - \frac{1}{8}18,4 \times 35,8 = 0,1863$$
$$S_{xx} = \sum x_i^2 - \frac{1}{n}(\sum x_i)^2 = 42,4662 - \frac{1}{8}18,4^2 = 0,1462$$
$$S_{yy} = \sum y_i^2 - \frac{1}{n}(\sum y_i)^2 = 160,4918 - \frac{1}{8}35,8^2 = 0,2868$$

Do đó, hệ số tương quan thực nghiệm là

$$r_{tn} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{0,1863}{\sqrt{0,1462 \times 0,2868}} \approx 0,9098$$

b) Giả sử hàm hồi quy tuyến tính phải tìm là $y = ax + b$. Chúng ta sẽ ước lượng các hệ số $a, b$ của hàm hồi quy theo phương pháp bình phương tối thiểu. Cụ thể ta phải xác định các hệ số $\hat{a}, \hat{b}$ để hàm số sau đạt giá trị nhỏ nhất

$$Q(a,b) = \sum_{i=1}^{n}(y_i - \hat{a}x_i - \hat{b})^2.$$

Như vậy $\hat{a}, \hat{b}$ được xác định bởi hệ phương trình sau:

$$\begin{cases} \dfrac{\partial Q}{\partial \hat{a}} = 0 \\ \dfrac{\partial Q}{\partial \hat{b}} = 0 \end{cases} \Leftrightarrow \begin{cases} \Big(\sum_{i=1}^{n} x_i^2\Big)\hat{a} + \Big(\sum_{i=1}^{n} x_i\Big)\hat{b} = \sum_{i=1}^{n} x_iy_i \\ \Big(\sum_{i=1}^{n} x_i\Big)\hat{a} + n\hat{b} = \sum_{i=1}^{n} y_i \end{cases}$$

Thay các số liệu thực nghiệm vào hệ trên ta thu được

$$\begin{cases} 42,4662\hat{a} + 18,4\hat{b} = 82,5263 \\ 18,4\hat{a} + 8\hat{b} = 35,8 \end{cases}$$

Giải hệ ta nhận được kết quả

$$\begin{cases} \hat{a} \approx 1,2743 \\ \hat{b} \approx 1,5442 \end{cases}$$

Như vậy ta nhận được hàm hồi quy tuyến tính $y = 1,2743x + 1,5442$.

c) Thay $x_0 = 2,9$ vào hàm hồi quy tuyến tính tìm được ở (b), ta nhận được

$$\hat{y}_0 = 1,2743 \times 2,9 + 1,5442 \approx 5,2397$$

là điểm ước lượng của $y_0 = E[Y|2,9]$.

Khoảng tin cậy cho giá trị dự báo $y_0 = E[Y|2,9]$ là

$$\left(\hat{y}_0 - t_{n-2,\alpha/2}\sqrt{s^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}\;;\; \hat{y}_0 + t_{n-2,\alpha/2}\sqrt{s^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}\right)$$

trong đó $\alpha = 1 - 0,9 = 0,1$, $t_{n-2,\alpha/2} = t_{6;0,05} = 1,943$ và

$$s^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{1}{6}\left(0,2868 - \frac{0,1863^2}{0,1462}\right) = 0,0082$$

Khi đó ta có khoảng tin cậy 90% cho giá trị dự báo

$$(4,9561; 5,5233)$$