

# Mục lục

<b>1</b>	<b>Biến cố ngẫu nhiên và xác suất</b>	<b>1</b>
1.1	Khái niệm	1
1.2	Mô hình xác suất cổ điển	3
1.3	Mô hình xác suất hình học	7
1.4	Công thức cộng và nhân xác suất	9
1.5	Công thức xác suất đầy đủ và công thức Bayes	15
1.6	Dãy thử Bernoulli	17
<b>2</b>	<b>Đại lượng ngẫu nhiên</b>	<b>22</b>
2.1	Khái niệm	23
2.2	Hàm phân bố xác suất	26
2.3	Hàm phụ thuộc đại lượng ngẫu nhiên	28
2.4	Các đặc trưng số của đại lượng ngẫu nhiên	31
2.5	Các phân bố xác suất thường gặp	37
<b>3</b>	<b>Véc tơ ngẫu nhiên</b>	<b>45</b>
3.1	Khái niệm	46
3.2	Hàm phân bố xác suất đồng thời	50
3.3	Xác định luật phân bố thành phần	52
3.4	Các đại lượng ngẫu nhiên độc lập	53
3.5	Phân bố có điều kiện	55
3.6	Tổng các đại lượng ngẫu nhiên	59
3.7	Momen tương quan và Hệ số tương quan	63
<b>4</b>	<b>Các định lý giới hạn</b>	<b>70</b>
<b>5</b>	<b>Mẫu và phân bố mẫu</b>	<b>74</b>
5.1	Mẫu ngẫu nhiên đơn giản	74
5.2	Các đặc trưng mẫu	75
5.3	Các phân bố thường gặp trong thống kê	80

5.4 Phân bố mẫu . . . . .	82
<b>6 Ước lượng tham số</b>	<b>84</b>
<b>7 Kiểm định giả thuyết thống kê</b>	<b>86</b>
7.1 Khái niệm . . . . .	86
7.2 Kiểm định giả thuyết về giá trị trung bình và xác suất . . . . .	87
7.3 Tiêu chuẩn phù hợp $\chi^2$ . . . . .	94
<b>8 Tương quan và hồi quy</b>	<b>107</b>
8.1 Hồi quy . . . . .	107
8.2 Hồi quy tuyến tính . . . . .	108
8.3 Dữ liệu lớn và học máy . . . . .	112
<b>1 Biến cố ngẫu nhiên và xác suất</b>	<b>122</b>
<b>2 Đại lượng ngẫu nhiên</b>	<b>125</b>
<b>3 Vectơ ngẫu nhiên</b>	<b>129</b>
<b>4 Các định lý giới hạn</b>	<b>136</b>
<b>5 Mẫu và phân bố mẫu</b>	<b>138</b>
<b>7 Kiểm định giả thuyết thống kê</b>	<b>140</b>
<b>8 Tương quan và hồi quy</b>	<b>144</b>
<b>Phụ lục</b>	<b>145</b>
<b>A Python</b>	<b>146</b>
A.1 Thư viện, môđun, phương thức . . . . .	146

## Chương 7

# Kiểm định giả thuyết thống kê

---

7.1	Khái niệm . . . . .	86
7.2	Kiểm định giả thuyết về giá trị trung bình và xác suất . . . . .	87
7.3	Tiêu chuẩn phù hợp $\chi^2$ . . . . .	94
7.3.1	Bài toán . . . . .	94
7.3.2	Kiểm tra phân bố của đại lượng ngẫu nhiên . . . . .	95
7.3.3	Kiểm tra tính độc lập . . . . .	101

---

### 7.1 Khái niệm

---

\* **Giả thuyết thống kê:** là giả thuyết về phân bố của đại lượng ngẫu nhiên, về các đặc trưng số của đại lượng ngẫu nhiên, về tính độc lập giữa các đại lượng ngẫu nhiên,...

Ký hiệu:  $H_0, H_1, \dots$

\* **Bài toán kiểm định** gồm cặp giả thuyết ngược nhau  $H_0 - H_1$  ( $H_1$  gọi là đối thuyết của  $H_0$ ). Để đơn giản, ta sắp xếp cặp  $H_0 - H_1$  sao cho mục tiêu của bài toán là bác bỏ  $H_0$ .

\* **Quyết định:**

- 1) bác bỏ  $H_0$  (chấp nhận  $H_1$ ).
- 2) chấp nhận  $H_0$  (tạm thời chấp nhận  $H_0$  hay chưa đủ cơ sở để bác bỏ  $H_0$ ).

\* **Sai lầm:**

- 1) loại I: bác bỏ  $H_0$  khi  $H_0$  đúng.

2) loại II: chấp nhận  $H_0$  khi  $H_0$  sai.

**\* Mục tiêu:**

1)  $P(\text{sai lầm loại I}) \leq \alpha$ .

2)  $P(\text{sai lầm loại II}) \rightarrow \min$ .

trong đó  $\alpha$  gọi là mức ý nghĩa của bài toán.

**\* Giải quyết:** Từ mẫu ngẫu nhiên thu được, xây dựng tiêu chuẩn kiểm định  $T$  và miền bác bỏ  $S$  sao cho:

1)  $T \in S \Rightarrow \text{bác bỏ } H_0$ .

2)  $T \notin S \Rightarrow \text{chấp nhận } H_0$ .

Với mỗi mẫu cụ thể, ta có 1 giá trị quan sát được của tiêu chuẩn kiểm định, ký hiệu  $t_{qs}$ .

## 7.2 Kiểm định giả thuyết về giá trị trung bình và xác suất

Một số giả thiết và thông tin khi kiểm định đối với:

1)  $EX, EY: X, Y \sim N$ , các mẫu ngẫu nhiên tương ứng  $(X_1, X_2, \dots, X_n)$  và  $(Y_1, Y_2, \dots, Y_m)$ .

2)  $p = P(A)$ :  $m$  là số lần  $A$  xảy ra trong  $n$  lần thử.

3)  $p_1 = P(A_1), p_2 = P(A_2)$ :  $m_i$  là số lần  $A_i$  xảy ra trong  $n_i$  lần thử,  $i = 1, 2$ .

$H_0$	$H_1$	Dấu hiệu	Tiêu chuẩn	ĐK bác bỏ $H_0$	Ghi chú
$EX = a_0$	$EX \neq a_0$	$DX = \sigma^2$ đã biết	$z_{qs} = \frac{\bar{x} - a_0}{\sigma} \sqrt{n} (*)$	$ z_{qs}  > z_0$	$\Phi(z_0) = 1 - \frac{\alpha}{2}$
	$EX > a_0$			$z_{qs} > z_0$	$\Phi(z_0) = 1 - \alpha$
	$EX < a_0$			$z_{qs} < -z_0$	
như trên		$DX$ chưa biết	$t_{qs} = \frac{\bar{x} - a_0}{s} \sqrt{n - 1} (**)$	$ t_{qs}  > t_0$	$t_0 = t_{\alpha}^{n-1}$
				$t_{qs} > t_0$	$t_0 = t_{2\alpha}^{n-1}$
				$t_{qs} < -t_0$	
$p = p_0$	$p \neq p_0$ $p > p_0$ $p < p_0$	$n \gg 1$	$z_{qs} = \frac{\frac{m}{n} - p_0}{\sqrt{p_0(1 - p_0)}} \sqrt{n}$	như (*)	
$EX = EY$	$EX \neq EY$	$DX, DY$ đã biết	$z_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{DX}{n} + \frac{DY}{m}}}$	như (*)	
	$EX > EY$				
	$EX < EY$				
như trên		$DX = DY$ chưa biết	$t_{qs} = (1)$	như (**), thay $n - 1$ bởi $n + m - 2$	
$p_1 = p_2$	$p_1 \neq p_2$ $p_1 > p_2$ $p_1 < p_2$	$n_1, n_2 \gg 1$	$z_{qs} = (2)$	như (*)	

$$(1) = \frac{\bar{x} - \bar{y}}{\sqrt{ns_X^2 + ms_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}, (2) = \left( \frac{m_1}{n_1} - \frac{m_2}{n_2} \right) \sqrt{\frac{n_1 n_2 (n_1 + n_2)}{(m_1 + m_2)(n_1 + n_2 - m_1 - m_2)}}. \quad (7.1)$$

**Ví dụ 7.1.** Để đánh giá hai loại phân bón, người ta tổng kết năng suất lúa  $X, Y$  (tấn/ha), tương ứng khi sử dụng mỗi loại phân bón, ở 50 địa phương

Năng suất $X, Y$	< 4.5	4.5 – 4.6	4.6 – 4.7	4.7 – 4.8	4.8 – 4.9	> 4.9
Số quan sát của $X$	5	10	15	10	7	3
Số quan sát của $Y$	3	11	13	10	8	5

a) Khi sử dụng loại phân bón thứ nhất:

- 1) Cho một ước lượng về năng suất trung bình, và về tỷ lệ các địa phương có năng suất trên 4.8 tấn/ha.
- 2) Ước lượng năng suất trung bình với độ tin cậy 90%.
- 3) Với độ tin cậy 95%, nên cho rằng tỷ lệ địa phương có năng suất trên 4.8 tấn/ha cao nhất là bao nhiêu %?
- 4) Với mức ý nghĩa 5%, có thể khẳng định năng suất trung bình đạt trên 4.6 tấn/ha không?

- b) Giả sử  $DX = DY$ . Với mức ý nghĩa 5%, năng suất khi sử dụng hai loại phân bón có khác biệt quá không?

*Giải.* Lấy giá trị đại diện trên mỗi khoảng: 4.45, 4.55, 4.65, 4.75, 4.85, 4.95.

- a) 1) Ước lượng năng suất trung bình  $EX$ :

$$\bar{x} = \frac{5 \cdot 4.45 + 10 \cdot 4.55 + 15 \cdot 4.65 + 10 \cdot 4.75 + 7 \cdot 4.85 + 3 \cdot 4.95}{50} = 4.676.$$

Ước lượng tỷ lệ các địa phương có năng suất trên 4.8 tấn/ha:

$$p^* = \frac{7 + 3}{50} = 0.2.$$

```

1 import numpy as np
2 X = np.array([4.45]*5 + [4.55]*10 + [4.65]*15 +
3              [4.75]*10 + [4.85]*7 + [4.95]*3)
4
5 X.mean()
6
7 X[X > 4.8]
8 len(X[X > 4.8])
9 p = 10 / len(X)
10 p

```

- 2) Khoảng tin cậy của  $EX$  ( $DX$  chưa biết):

$$\left( \bar{x} - t_0 \frac{s}{\sqrt{n-1}}, \bar{x} + t_0 \frac{s}{\sqrt{n-1}} \right).$$

$$\bar{x} = 4.676.$$

$$s^2 = \frac{5 \cdot 4.45^2 + 10 \cdot 4.55^2 + 15 \cdot 4.65^2 + 10 \cdot 4.75^2 + 7 \cdot 4.85^2 + 3 \cdot 4.95^2}{50} -$$

$$4.676^2 = 0.01832.$$

$$s = \sqrt{0.01832} = 0.1354.$$

$$t_0 = t_{1-\gamma}^{n-1} = t_{1-90\%}^{50-1} = t_{0.1}^{49} = 1.6766.$$

Khoảng tin cậy trên là (4.6436, 4.7084) (tấn/ha).

**Cách 1:**

```

1 from scipy.stats import t
2 t0 = t.isf(0.1 / 2, 50 - 1)
3
4
5 from math import sqrt
6 X.mean() - t0 * X.std() / sqrt(50-1), X.mean() + t0 * X.
   std() / sqrt(50-1)

```

**Cách 2:**

```
1 from scipy.stats import sem
2 t.interval(alpha=0.9, df=len(X)-1, loc=X.mean(), scale=
    sem(X))
```

3) Tỷ lệ địa phương có năng suất trên 4.8 tấn/ha cao nhất

$$p^* + z_0 \sqrt{\frac{p^*(1-p^*)}{n}}.$$

$$p^* = 0.2.$$

$$\Phi(z_0) = \frac{1+\gamma}{2} = \frac{1+95\%}{2} = 0.975 \Rightarrow z_0 = 1.960.$$

$$\text{Tỷ lệ cần tìm } 0.3109 = 31.09\%.$$

**Cách 1:**

```
1 from scipy.stats import norm
2 z0 = norm.ppf(0.975)
3 z0
5 0.2 + z0 * sqrt(0.2*0.8/50)
```

**Cách 2:**

```
1 from statsmodels.stats.proportion import
    proportion_confint
2 proportion_confint(10, 50, 0.05)
```

4)  $H_0 : EX = 4.6, H_1 : EX > 4.6, \alpha = 5\%$ .

$$DX \text{ chưa biết} \Rightarrow t_{qs} = \frac{\bar{x} - a_0}{s} \sqrt{n-1} = \frac{4.676 - 4.6}{0.1354} \sqrt{49} = 3.9301.$$

$$t_0 = t_{2\alpha}^{n-1} = t_{2.5\%}^{50-1} = t_{0.1}^{49} = 1.6766.$$

$$t_{qs} > t_0 \Rightarrow \text{bác bỏ } H_0, \text{ tức là, năng suất trung bình đạt trên 4.6 tấn/ha.}$$

**Cách 1:**

```
1 (X.mean() - 4.6) / X.std() * sqrt(50 - 1)
2 t.isf(0.1 / 2, 50 - 1)
```

**Cách 2:**

```
1 from scipy.stats import ttest_1samp
2 ttest_1samp(X, 4.6) # → tqs
```

b)  $H_0 : EX = EY, H_1 : EX \neq EY, \alpha = 5\%$ .

$$DX = DY \text{ nhưng chưa biết} \Rightarrow t_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{ns_X^2 + ms_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

$$\bar{x} = 4.676, s_X^2 = 0.01832.$$

$$\bar{y} = \frac{3 \cdot 4.45 + 11 \cdot 4.55 + 13 \cdot 4.65 + 10 \cdot 4.75 + 8 \cdot 4.85 + 5 \cdot 4.95}{50} = 4.698.$$

$$s_Y^2 = \frac{3 \cdot 4.45^2 + 11 \cdot 4.55^2 + 13 \cdot 4.65^2 + 10 \cdot 4.75^2 + 8 \cdot 4.85^2 + 5 \cdot 4.95^2}{50} - 4.698^2 = 0.01970.$$

Suy ra  $t_{qs} = -0.7898$ .

$$t_0 = t_{\alpha}^{n+m-2} = t_{5\%}^{50+50-2} = t_{0.05}^{98} = 1.9845.$$

$|t_{qs}| < t_0 \Rightarrow$  chấp nhận  $H_0$ , tức là, năng suất khi dùng hai loại phân bón không khác biệt nhiều.

### Cách 1:

```
1 Y = np.array([4.45]*3 + [4.55]*11 + [4.65]*13 + [4.75]*10 +
2             [4.85]*8 + [4.95]*5)
3 Y.mean(), Y.var()
4 (X.mean() - Y.mean()) / sqrt(50*X.var() + 50*Y.var()) * sqrt
   (50*50*(50+50-2)/(50+50))
5 t.isf(0.05/2, 50+50-2)
```

### Cách 2:

```
1 from scipy.stats import ttest_ind
2 ttest_ind(X, Y) # → tqs
```

**Cách 3:** có thể giải cả bài toán kiểm định một phía và hai phía.

```
1 from statsmodels.stats import weightstats
2 weightstats.ttest_ind(X, Y)
```

□

## Bài tập 7.2

### 7.1. Thời gian gia công của 30 sản phẩm (phút) cho ở bảng sau

Thời gian gia công 1 sản phẩm	10 – 12	12 – 14	14 – 16	16 – 18
Số sản phẩm	3	12	11	4

Giả sử thời gian gia công một sản phẩm là một đại lượng ngẫu nhiên có phân bố chuẩn với  $\sigma = 2$ .

- Hãy ước lượng thời gian gia công trung bình một sản phẩm bằng khoảng tin cậy với độ tin cậy 98%.
- Định mức thời gian gia công một sản phẩm là 15 phút. Hãy kiểm định ở mức ý nghĩa 2% xem định mức thời gian trên có nhiều quá không?



**7.2.** Một mẫu cỡ  $n = 9$  từ phân bố chuẩn cho ta các giá trị 3, 4, 5, 3, 6, 11, 8, 2, 3.

- Tính kỳ vọng mẫu. Tính ước lượng không chệch của phương sai.
- Tìm khoảng tin cậy của kỳ vọng của  $X$  với độ tin cậy 97%.
- Kiểm định giả thuyết  $H_0 : EX = 3, H_1 : EX > 3$  với mức ý nghĩa 5%.

**7.3.** Giả sử thời gian của một chuyến tàu chạy từ A đến B là một đại lượng ngẫu nhiên có phân bố chuẩn. Quan sát 25 lần chạy từ A đến B của chuyến tàu đó ta có bảng số liệu

Thời gian 1 lần chạy (giờ)	15 – 15.5	15.5 – 16	16 – 16.5	16.5 – 17
Số lần	1	8	14	2

Định mức thời gian chạy từ A đến B là 17 giờ. Hãy kiểm định ở mức ý nghĩa 1% xem định mức thời gian trên có nhiều quá không?

**7.4.** Giả sử  $X$ , lượng điện mà các gia đình ở quận A sử dụng trong 1 tháng, là một đại lượng ngẫu nhiên có phân bố chuẩn. Xem xét lượng điện sử dụng của 30 gia đình trong 1 tháng ta có số liệu

Lượng điện sử dụng (kWh)	150 – 160	160 – 170	170 – 180	180 – 190
Số gia đình	5	11	10	4

- Hãy tìm các ước lượng không chệch của  $EX$  và  $DX$ .
- Tìm khoảng tin cậy của  $DX$  với độ tin cậy 97%.
- Kiểm định ở mức ý nghĩa 2% xem có phải lượng điện trung bình mà các hộ dân sử dụng trong 1 tháng ở quận A vượt quá 167 kWh không.

**7.5.** Giả sử khối lượng  $X$  của một viên gạch do một xí nghiệp sản xuất là đại lượng ngẫu nhiên có phân bố chuẩn. Cân 31 viên gạch ta được bảng số liệu

Khối lượng (kg)	2.25 – 2.3	2.3 – 2.35	2.35 – 2.4	2.4 – 2.45
Số viên	4	11	13	3

- Tìm ước lượng không chệch của  $EX$  và  $DX$ .
- Hãy cho một ước lượng khoảng của khối lượng trung bình của viên gạch với độ tin cậy 99%.
- Kiểm định ở mức ý nghĩa 2%:  $H_0 : EX = 2.4, H_1 : EX < 2.4$ .

**7.6.** Một mẫu gồm 11 giá trị quan sát được của đại lượng ngẫu nhiên có phân bố chuẩn  $X$  được cho ở bảng

Giá trị của $X$	3	3.5	4	4.5
Số lần $n_i$	1	2	4	4

- Hãy ước lượng giá trị trung bình và độ lệch chuẩn của  $X$ .
- Tìm khoảng tin cậy của  $EX$  với độ tin cậy 96%.
- Kiểm định ở mức ý nghĩa 4%:  $H_0 : EX = 3.7, H_1 : EX > 3.7$ .

**7.7.** Hai mẫu ngẫu nhiên độc lập cỡ tương ứng là 20 và 30 từ  $X$  và  $Y$  cho ta các giá trị sau  $\bar{x} = 10, \bar{y} = 13, s'_X = s'_Y = 6$ . Kiểm định ở mức ý nghĩa 10%, giả thuyết cho rằng  $EX = EY$ , giả sử  $DX = DY$ .

**7.8.** Giả sử cường độ chịu kéo của các dây thép cùng một loại trên một công trường xây dựng là một đại lượng ngẫu nhiên có phân bố chuẩn. Đo ngẫu nhiên cường độ chịu kéo của 29 cây thép trên công trường đó, ta có bảng số liệu

Cường độ ( $\text{kg/cm}^2$ )	3570 – 3580	3580 – 3590	3590 – 3600	3600 – 3610
Số cây thép $n_i$	5	9	8	7

- Hãy cho một ước lượng khoảng của cường độ chịu kéo trung bình với độ tin cậy 99%.
- Quy định về kỹ thuật đòi hỏi cường độ chịu kéo trung bình phải đạt  $3600 \text{ kg/cm}^2$ . Dư luận cho rằng công trường sử dụng thép chưa đạt yêu cầu kỹ thuật. Hãy kiểm định ở mức ý nghĩa 1% xem dư luận trên có đúng không?

**7.9.** Giả sử những năm trước tỷ lệ sinh viên phải học lại môn xác suất là 20%. Năm nay, trong một lớp 100 sinh viên thì có 13 sinh viên phải học lại môn xác suất. Hãy kiểm định ở mức ý nghĩa 5% giả thuyết nói rằng học sinh năm nay nói chung khá hơn năm ngoái.

**7.10.** Khối lượng các bao xi măng là đại lượng ngẫu nhiên có phân bố chuẩn  $N(50, 0.01)$ . Có nhiều ý kiến khách hàng phản ánh là khối lượng bị thiếu. Để kiểm tra, người ta đã cân ngẫu nhiên 25 bao xi măng trong kho, kết quả như sau:

Khối lượng bao	48 – 48.5	48.5 – 49	49 – 49.5	49.5 – 50	50 – 50.5
Số bao	2	5	10	6	2

Hãy xem ý kiến khách hàng có đúng không bằng cách kiểm tra giả thuyết  $H_0 : EX =$

50,  $H_1 : EX < 50$  ở mức ý nghĩa 0.05.

**7.11.** Với mức ý nghĩa 0.08 hãy kiểm định  $H_0 : EX = EY$ ,  $H_1 : EX > EY$ , trong đó  $X$  và  $Y$  là hai đại lượng ngẫu nhiên có phân bố chuẩn. Biết rằng hai mẫu độc lập cỡ  $n = 17$  từ  $X$  và  $m = 13$  từ  $Y$  cho ta số liệu sau:

$x_i$	21.7	21.9	22.1	22.2	22.5
$n_i$	1	4	6	5	1

$y_i$	21.8	22	22.2
$m_i$	5	7	1

Cho biết  $DX = 0.03$ ,  $DY = 0.02$ .

**7.12.** Để so sánh cường độ chịu nén của gạch do hai nhà máy A và B sản xuất, người ta lấy hai mẫu cỡ 8 và 6 tương ứng từ tổng số gạch do nhà máy A và B, được số liệu sau

Cường độ chịu nén của gạch nhà máy A (kg/cm <sup>2</sup> )	Cường độ chịu nén của gạch nhà máy B (kg/cm <sup>2</sup> )
70.25, 70.5, 71.75, 69.5, 70.25, 73, 72.5, 71.25	70.25, 68.25, 69.25, 70.25, 69.25, 69

Giả sử cường độ chịu nén của gạch do hai nhà máy A và B sản xuất là các đại lượng ngẫu nhiên có phân bố chuẩn với phương sai bằng nhau.

- Hãy tìm khoảng tin cậy của cường độ chịu nén trung bình của gạch của nhà máy A với độ tin cậy 95%.
- Hãy kiểm định giả thuyết  $H_0$ : “cường độ chịu nén trung bình của gạch do hai nhà máy sản xuất là như nhau”, với đối thuyết  $H_1$ : “cường độ chịu nén trung bình của gạch nhà máy A cao hơn của gạch nhà máy B” ở mức ý nghĩa  $\alpha = 0.01$ .

## 7.3 Tiêu chuẩn phù hợp $\chi^2$

### 7.3.1 Bài toán

Cho hệ biến cố đầy đủ  $A_1, A_2, \dots, A_h$ . Giả sử  $n_i$  = số lần  $A_i$  xảy ra trong  $n$  lần thử ( $n = \sum_{i=1}^h n_i$ ). Kiểm định ở mức ý nghĩa  $\alpha$ :

$$H_0 : P(A_i) = p_{i0} \quad \forall i; \quad H_1 : \exists i \quad P(A_i) \neq p_{i0}. \quad (7.2)$$

1)  $e_i = np_{i0}$ .

2) Tiêu chuẩn kiểm định  $\chi^2 = \sum_{i=1}^h \frac{(n_i - e_i)^2}{e_i}$ .

3) Điều kiện bác bỏ  $H: \chi^2 > \chi^2(\alpha, h - 1)$ .

**Ví dụ 7.2.** Cho nhóm biễn cổ đầy đủ gồm  $A_1, A_2, A_3$ . Thực hiện phép thử 100 lần ta có kết quả sau

Biễn cổ	$A_1$	$A_2$	$A_3$
Tần số	52	34	14

Kiểm định ở mức ý nghĩa 5% giả thuyết  $P(A_1) = 0.5, P(A_2) = 0.3$  và  $P(A_3) = 0.2$ .

*Giải.*

$n_i$	$p_{i0}$	$e_i$	$\frac{(n_i - e_i)^2}{e_i}$	Cách tính
52	0.5	50	0.08	$100 \cdot 0.5$
34	0.3	30	0.5333	$\frac{(34 - 30)^2}{30}$
14	0.2	20	1.8	
$\Sigma$			2.4133	

$$\chi_{qs}^2 = 2.4133.$$

$$\chi_0^2 = \chi^2(\alpha, h - 1) = \chi^2(0.05, 2) = 5.9914.$$

$$\chi_{qs}^2 < \chi_0^2 \Rightarrow \text{chấp nhận giả thuyết đã nêu.}$$

□

```

1 import numpy as np
2 n = np.array([52, 34, 14])
3 p0 = np.array([0.5, 0.3, 0.2])

5 e = 100 * p0
6 (n - e)**2 / e
7 sum(_)

9 from scipy.stats import chi2
10 chi2.isf(0.05, 3-1)

```

### 7.3.2 Kiểm tra phân bố của đại lượng ngẫu nhiên

Cho mẫu ngẫu nhiên  $(X_1, X_2, \dots, X_n)$  rút từ đại lượng ngẫu nhiên  $X$ . Kiểm định ở mức ý nghĩa  $\alpha$ :

$$H_0 : X \sim F(x); H_1 : X \not\sim F(x). \quad (7.3)$$

trong đó  $F(x)$  là hàm phân bố xác suất cho trước.

1) Phân hoạch  $\text{Im}X = S_1 \cup S_2 \cup \dots \cup S_h$ .  $n_i$  = số phần tử trong mẫu  $\in S_i$ ,  $(n = \sum_{i=1}^h n_i)$ .

2)  $H_0 \Leftrightarrow X$  có bảng phân bố xác suất  $\frac{X}{P} \left| \begin{array}{ccc} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{array} \right.$  hoặc hàm mật độ xác suất  $f(x)$ .

$$p_{i0} = P(X \in S_i | H) = \begin{cases} \sum_{x_k \in S_i} p_k \\ \int_{S_i} f(x) dx. \end{cases}$$

3) Áp dụng tiêu chuẩn  $\chi^2$  ở mục 7.3.1

**Chú ý:** Nếu luật phân bố phụ thuộc  $r$  tham số chưa biết:  $F(x) = F(x, \theta_1, \theta_2, \dots, \theta_r)$  thì

1) Tìm ước lượng  $\theta_i^*$  của  $\theta_i$  bằng phương pháp bình phương nhỏ nhất. Bảng sau cho một số ước lượng theo phương pháp này:

Phân bố	Ký hiệu	$\theta$	$\theta^*$
Nhị thức	$X \sim B(n, p)$	$p$	$p^*$
Poisson	$X \sim P_\lambda$	$\lambda$	$\bar{X}$
Đều	$X \sim U(a, b)$	$a, b$	$\bar{X} - S\sqrt{3}, \bar{X} + S\sqrt{3}$
Mũ	$X \sim \varepsilon_\lambda$	$\lambda$	$1/\bar{X}$
Chuẩn	$X \sim N(\mu, \sigma^2)$	$\mu, \sigma^2$	$\bar{X}, S^2$

(7.4)

2) Kiểm định ở mức ý nghĩa  $\alpha$ :

$$H_0^* : X \sim F(x, \theta_1^*, \theta_2^*, \dots, \theta_r^*)$$

$$H_1^* : X \not\sim F(x, \theta_1^*, \theta_2^*, \dots, \theta_r^*)$$

với điều kiện bác bỏ  $H_0^*$  (và do đó bác bỏ  $H_0$ ) là  $\chi^2 > \chi^2(\alpha, h - r - 1)$ .

**Ví dụ 7.3.** Bảng sau cho số liệu thống kê mẫu cỡ 100 của đại lượng ngẫu nhiên  $X$

Khoảng của $X$	0 – 0.5	0.5 – 1.0	1.0 – 1.5	1.5 – 2.0
Số giá trị	9	21	34	36

Kiểm định ở mức ý nghĩa 10% giả thuyết cho rằng

a)  $X$  có hàm mật độ

$$f(x) = \begin{cases} \frac{x}{2}, & \text{khi } x \in [0, 2] \\ 0, & \text{khi } x \notin [0, 2]. \end{cases}$$

b)  $X$  có phân bố đều.

*Giải.* a)

$S_i$	$n_i$	$p_{i0}$	$e_i$	$\frac{(n_i - e_i)^2}{e_i}$	Cách tính
$(-\infty, 0.5)$	9	0.0625	6.25	1.21	$\int_{-\infty}^{0.5} f(x)dx$
$[0.5, 1)$	21	0.1875	18.75	0.27	
$[1, 1.5)$	34	0.3125	31.25	0.242	
$[1.5, \infty)$	36	0.4375	43.75	1.3729	
$\Sigma$				3.0949	

$$\chi_{qs}^2 = 3.0949.$$

$$\chi_0^2 = \chi^2(\alpha, h-1) = \chi^2(0.1, 3) = 6.2514.$$

$$\chi_{qs}^2 < \chi_0^2 \Rightarrow \text{chấp nhận } X \text{ có hàm mật độ } f(x).$$

```

1 from sympy import *
2 import numpy as np

4 s = [-oo, 0.5, 1, 1.5, oo]
5 n = [9, 21, 34, 36]

7 x = symbols('x')
8 f = Piecewise((x/2, (0 <= x) & (x <= 2)), (0, True))

10 p0 = np.array([ f.integrate((x, s[i], s[i+1])) for i in
11                  range(4) ])
11 e = 100 * p0
12 (n - e)**2 / e
13 sum(_)

15 from scipy.stats import chi2
16 chi2.isf(0.1, 4-1)

```

$$b) H_0 : X \sim U(a, b), H_1 : X \not\sim U(a, b), \alpha = 10\%.$$

\* Bảng đại diện

$X$	0.25	0.75	1.25	1.75
$n_i$	9	21	34	36

$$\bar{x} = \frac{9 \cdot 0.25 + 21 \cdot 0.75 + 34 \cdot 1.25 + 36 \cdot 1.75}{100} = 1.235.$$

$$s^2 = \frac{9 \cdot 0.25^2 + 21 \cdot 0.75^2 + 34 \cdot 1.25^2 + 36 \cdot 1.75^2}{100} - 1.235^2 = 0.2323 \Rightarrow s = 0.4819.$$

Ước lượng của  $a, b$  theo phương pháp bình phương nhỏ nhất:

$$a^* = \bar{x} - s\sqrt{3} = 0.4002, b^* = \bar{x} + s\sqrt{3} = 2.0698.$$

```

1 X = np.array([0.25]*9 + [0.75]*21 + [1.25]*34 + [1.75]*36)
2 X.mean(), X.var(), X.std()

4 a = X.mean() - X.std() * np.sqrt(3)
5 b = X.mean() + X.std() * np.sqrt(3)
6 a, b

```

\*  $H_0^* : X \sim U(a^*, b^*)$ ,  $H_1^* : X \not\sim U(a^*, b^*)$ ,  $\alpha = 10\%$ , trong đó  $H_0^*$  có nghĩa  $X$  có hàm mật độ

$$f(x) = \begin{cases} \frac{1}{b^* - a^*}, & \text{khi } x \in [a^*, b^*] \\ 0, & \text{khi } x \notin [a^*, b^*]. \end{cases}$$

$S_i$	$n_i$	$p_{i0}$	$e_i$	$\frac{(n_i - e_i)^2}{e_i}$
$(-\infty, 0.5)$	9	0.05975	5.975	1.5310
$[0.5, 1)$	21	0.2995	29.95	2.6739
$[1, 1.5)$	34	0.2995	29.95	0.5480
$[1.5, \infty)$	36	0.3413	34.13	0.1028
$\sum$				4.8557

$$\chi_{qs}^2 = 4.8557.$$

$$\chi_0^2 = \chi^2(\alpha, h - r - 1) = \chi^2(0.1, 4 - 2 - 1) = 2.7055.$$

$\chi_{qs}^2 > \chi_0^2 \Rightarrow$  bác bỏ  $H_0$ , tức là,  $X$  không có phân bố đều.

```

1 from sympy.stats import Uniform, P
2 X = Uniform('x', a, b)

4 p = np.array([ P((s[i] <= X) & (X < s[i+1])) for i in range
                (4) ])
5 e = p * n.sum()
6 (n - e)**2 / e
7 sum(_)

9 chi2.isf(0.1, 4-2-1)

```

□

**Ví dụ 7.4.** Trong [Ví dụ 5.2 Trang 78](#), xét biểu đồ tần số của chiều cao (gộp cả nam và nữ).

- Với mức ý nghĩa 5%, có thể khẳng định chiều cao có phân bố chuẩn không, biết khoảng chứa chiều cao được chia đều thành 5 đoạn bằng nhau?
- Vẽ biểu đồ mô tả giá trị của  $\chi_{qs}^2$  và  $\chi_0^2$  khi số khoảng chia thay đổi từ 5 đến 50?
- Thực hiện yêu cầu ở ý (a) và (b) cho chiều cao của nam giới?

Giải.

```

1 import pandas as pd
2 df = pd.read_csv('https://raw.githubusercontent.com/Dataweekends/
  zero_to_deep_learning_video/master/data/weight-height.csv')

```

a)

$S_i$	$n_i$	$p_{i0}$	$e_i$	$\frac{(n_i - e_i)^2}{e_i}$
$-\infty \rightarrow 59.21$	241	0.03143	314.2618	17.0790
$59.21 \rightarrow 64.16$	2832	0.2514	2514.0810	40.2026
$64.16 \rightarrow 69.10$	4363	0.4787	4787.3203	37.6093
$69.10 \rightarrow 74.05$	2377	0.2155	2155.2883	22.8072
$74.05 \rightarrow \infty$	187	0.02290	229.0487	7.7193
$\Sigma$				125.4174

$$\chi_{qs}^2 = 125.4174.$$

$$\chi_0^2 = \chi^2(\alpha, h - r - 1) = \chi^2(0.05, 5 - 2 - 1) = 5.9915.$$

$\chi_{qs}^2 > \chi_0^2 \Rightarrow$  chiều cao không có phân bố chuẩn.

```

1 X = df['Height']
2 a, b = X.min(), X.max()

4 h = 5

6 s = [a + i * (b-a)/h for i in range(h+1)]
7 from sympy import oo
8 s[0], s[-1] = -oo, oo

10 n = [len( X[(X >= s[i]) & (X < s[i+1])] ) for i in range(h)]

12 mu, sigma = X.mean(), X.std()
13 from sympy.stats import Normal, P
14 X = Normal('x', mu, sigma)

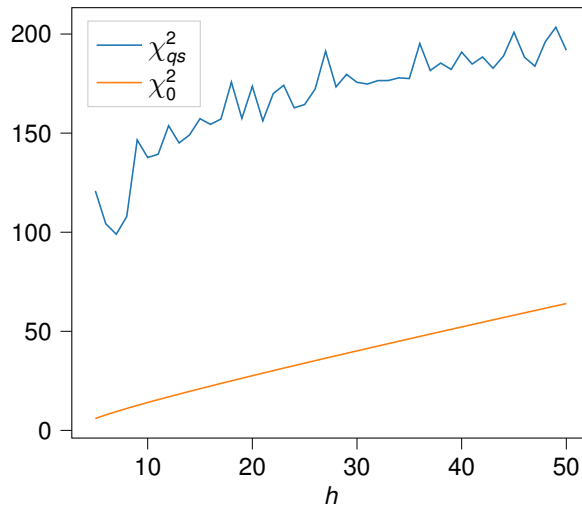
16 p0 = np.array( [ round( P((X > s[i]) & (X < s[i+1]))), 10)
  for i in range(h) ] )
17 e = sum(n) * p0
18 (n - e)**2 / e
19 sum(_)

21 from scipy.stats import chi2
22 chi2.isf(0.05, h-1)

```



b)

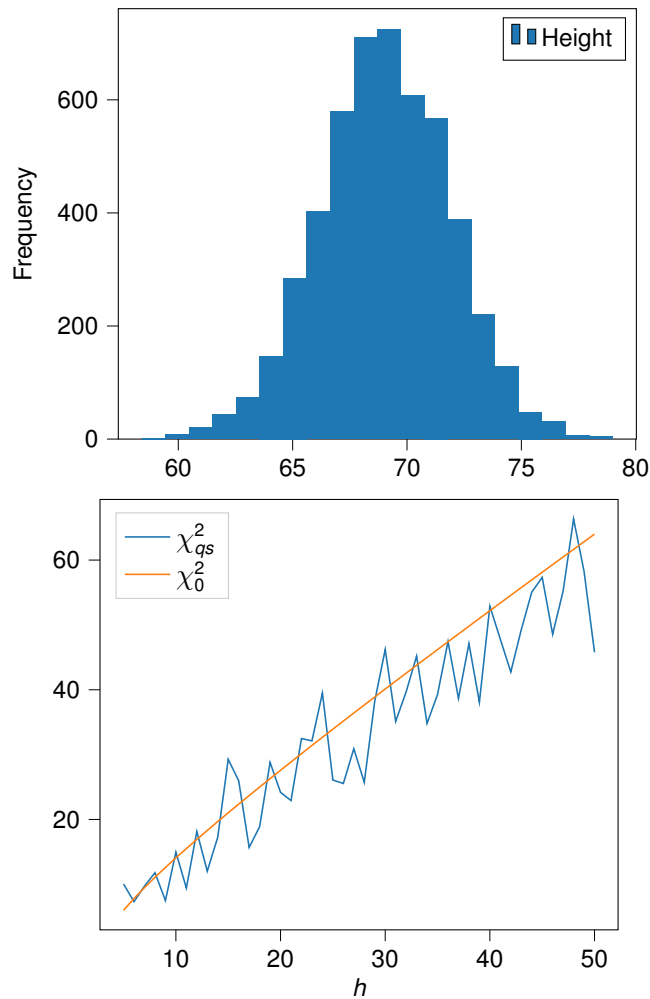


```

1 chi, chi0 = [], []
2 for h in range(5, 51): # ≈15 phút
3     X = df['Height']
4     s = [a + i * (b-a)/h for i in range(h+1)]
5     n = [len( X[(X >= s[i]) & (X < s[i+1])] ) for i in range(h
6         )]
7     X = Normal('x', mu, sigma)
8     p0 = np.array( [ round( P((X > s[i]) & (X < s[i+1])), 10)
9         for i in range(h)] )
10    e = sum(n) * p0
11
12    chi.append( sum((n - e)**2 / e) )
13    chi0.append( chi2.isf(0.05, h-2-1) )
14
15 import matplotlib.pyplot as plt
16 h = range(5, 51)
17 plt.plot(h, chi, label='χqs2')
18 plt.plot(h, chi0, label='χ02')
19 plt.legend()
20 plt.xlabel('h')

```

c)



Với đa số các cách chia khoảng,  $\chi_{qs}^2 < \chi_0^2$ , nên ta có thể chấp nhận giả thuyết chiều cao nam giới có phân bố chuẩn.

```
1 # Các lệnh giống ý (a) và (b), chỉ thay lệnh X = df['Height'] bởi
2 X = df[ df['Gender'] == 'Male' ]['Height']
3 X.plot(kind='hist', bins=20)
```

□

### 7.3.3 Kiểm tra tính độc lập

Cho mẫu ngẫu nhiên  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  rút từ véc tơ ngẫu nhiên  $(X, Y)$ . Kiểm định ở mức ý nghĩa  $\alpha$ :

$$H_0 : X, Y \text{ độc lập}; H_1 : X, Y \text{ không độc lập.} \quad (7.5)$$

1) Phân hoạch  $\text{Im}X = S_1 \cup S_2 \cup \dots \cup S_h$ ,  $\text{Im}Y = T_1 \cup T_2 \cup \dots \cup T_k$ .

$n_{ij}$  = số phần tử trong mẫu  $\in S_i \times T_j$ , ( $n = \sum_{i,j} n_{ij}$ ).

$$2) \quad n_{i*} = \sum_j n_{ij}, \quad n_{*j} = \sum_i n_{ij}.$$

$$3) \quad e_{ij} = \frac{n_{i*} \times n_{*j}}{n}.$$

$$4) \quad \text{Tiêu chuẩn kiểm định } \chi^2 = \sum_{i,j} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = n \left( \sum_{i,j} \frac{n_{ij}^2}{n_{i*} \times n_{*j}} - 1 \right).$$

Số phép toán của hai tổng này lần lượt là  $8mn - m - n - 1$  và  $6mn - m - n + 1$ .

5) Điều kiện bác bỏ  $H_0: \chi^2 > \chi^2[\alpha, (h-1)(k-1)]$ .

$\begin{matrix} Y \\ X \end{matrix}$	$T_1$	$T_2$	$\dots$	$T_k$	$\Sigma$
$S_1$	$n_{11}$ $e_{11}$	$n_{12}$ $e_{12}$	$\dots$	$n_{1k}$ $e_{1k}$	$n_{1*}$
$S_2$	$n_{21}$ $e_{21}$	$n_{22}$ $e_{22}$	$\dots$	$n_{2k}$ $e_{2k}$	$n_{2*}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$S_h$	$n_{h1}$ $e_{h1}$	$n_{h2}$ $e_{h2}$	$\dots$	$n_{hk}$ $e_{hk}$	$n_{h*}$
$\Sigma$	$n_{*1}$	$n_{*2}$	$\dots$	$n_{*k}$	$n$

**Ví dụ 7.5.** Số liệu cỡ 200 khảo sát mối liên hệ giữa thói quen hút thuốc và giới tính

$\begin{matrix} \text{Thói quen} \\ \text{Giới tính} \end{matrix}$	Nam	Nữ
Nghiện	35	10
Không nghiện	75	80

Với mức ý nghĩa 0.01, có thể khẳng định gì mối liên hệ này?

*Giải.*

Giới tính Thói quen	Nam	Nữ	$\Sigma$
Nghện	35 24.75	10 20.25	45
Không nghiện	75 85.25	80 69.75	155
$\Sigma$	110	90	

$$\chi_{qs}^2 = \frac{(35 - 24.75)^2}{24.75} + \frac{(10 - 20.25)^2}{20.25} + \frac{(75 - 85.25)^2}{85.25} + \frac{(80 - 69.75)^2}{69.75} = 12.17.$$

$$\chi_0^2 = \chi^2[\alpha, (h-1)(k-1)] = \chi^2[0.01, (2-1)(2-1)] = 6.6349.$$

$\chi_{qs}^2 > \chi_0^2 \Rightarrow$  thói quen hút thuốc phụ thuộc vào giới tính.

```

1 import numpy as np
2 n = np.array([[35, 10], [75, 80]])

4 nx = n.sum(axis=1)
5 ny = n.sum(axis=0)
6 e = np.array([ [nx[i] * ny[j] / n.sum() for j in range(2)] for i
    in range(2) ])
7 (n - e)**2 / e
8 _ .sum()

10 from scipy.stats import chi2
11 chi2.isf(0.01, (2-1)*(2-1))

```

□

**Ví dụ 7.6.** Trong [Ví dụ 5.2](#), với mức ý nghĩa 5%.

- Kiểm định giả thuyết về sự phụ thuộc giữa chiều cao vào giới tính, biết khoảng chiều cao chia thành ba phần.
- Khi số khoảng phân chia chiều cao thay đổi từ 3 đến 30, vẽ biểu đồ sự thay đổi giá trị của  $\chi_{qs}^2$  và  $\chi_0^2$ .

*Giải.*

a)

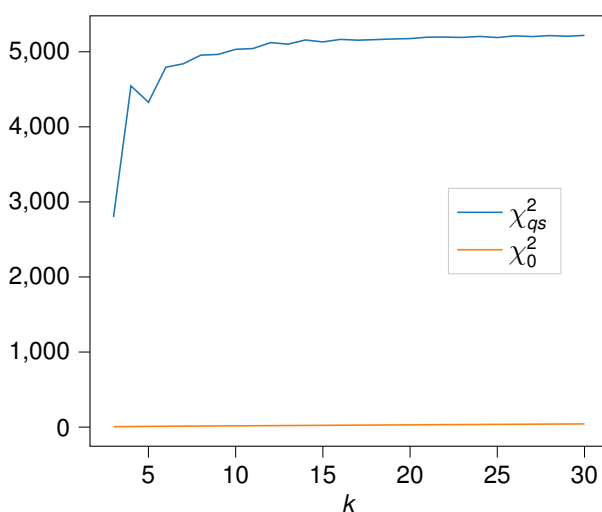
Chiều cao Giới tính	$[-\infty, 62.51)$	$[62.51, 70.75)$	$[70.75, \infty)$	$\Sigma$
Nam	74 861.5	3529 3433.0	1397 705.5	5000
Nữ	1649 861.5	3337 3433.0	14 705.5	5000
$\Sigma$	1723	6866	1411	

$$\chi_{qs}^2 = \frac{(74 - 861.5)^2}{861.5} + \frac{(3529 - 3433.0)^2}{3433.0} + \dots + \frac{(14 - 705.5)^2}{705.5} = 2800.6374.$$

$$\chi_0^2 = \chi^2[\alpha, (h-1)(k-1)] = \chi^2[0.05, (2-1)(3-1)] = 5.9915.$$

$\chi_{qs}^2 \gg \chi_0^2 \Rightarrow$  chiều cao phụ thuộc vào giới tính.

b)



```

1 chi, chi0 = [], []
2 for k in range(3, 31):
3     t = [a + i*(b-a)/k for i in range(k+1)]
4     n = np.array([ [ len(X[ (X == s[i]) & (Y >= t[j]) & (Y < t
5         [j+1]))]) for j in range(k)] for i in range(2) ])
6     ny = n.sum(axis=0)
7     e = [[i * j / n.sum() for j in ny] for i in nx]
8     chi.append( ((n - e)**2 / e).sum() )
9     chi0.append( chi2.isf(0.05, (2-1)*(k-1)) ) )
10
11 import matplotlib.pyplot as plt
12 k = range(3, 31)

```

```

12 plt.plot(k, chi, label='χqs2')
13 plt.plot(k, chi0, label='χ02')
14 plt.legend()
15 plt.xlabel('k')

```

□

### Bài tập 7.3

**7.13.** Số liệu thống kê của một mẫu cỡ  $n = 600$  từ  $X$  cho ở bảng sau:

$X$	10 – 11	11 – 12	12 – 13	13 – 14	14 – 15	15 – 16
Số lần $n_i$	95	104	96	106	93	106

Hãy kiểm định ở mức ý nghĩa 3% xem có phải  $X$  có phân bố đều trên đoạn  $[10, 16]$  hay không.

**7.14.** Mẫu cỡ  $n = 120$  từ  $X$  cho số liệu sau

Khoảng của $X$	10 – 11	11 – 12	12 – 14	14 – 17	17 – 19	19 – 22
Số lần $n_i$	6	16	24	24	22	28

Hãy kiểm định ở mức ý nghĩa 3% giả thuyết cho rằng  $X$  có phân bố đều trên đoạn  $[10, 22]$ .

**7.15.** Theo dõi thời gian gia công của 200 chi tiết máy, ta thu được số liệu sau

Thời gian gia công 1 chi tiết (phút)	10 – 12	12 – 14	14 – 16	16 – 18	18 – 20	20 – 22
Số chi tiết	16	35	49	50	30	20

Với mức ý nghĩa 10%, hãy kiểm định xem giả thuyết cho rằng thời gian gia công một chi tiết máy có phân bố chuẩn  $N(16, 3^2)$  có đúng hay không.

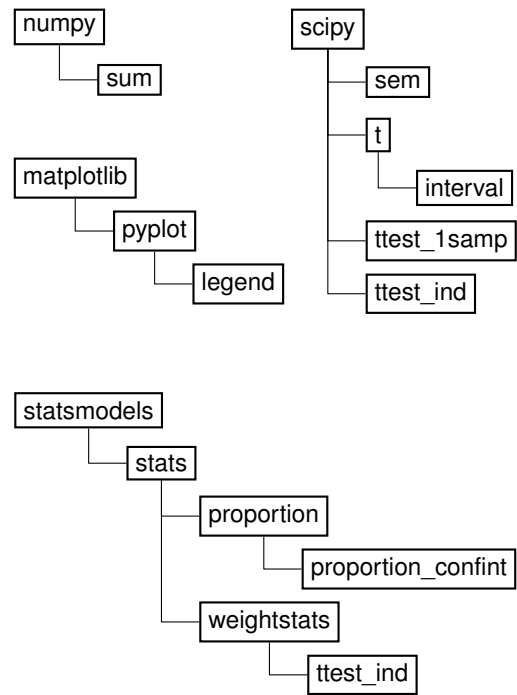
**7.16.** Mẫu cỡ  $n = 100$  từ đại lượng ngẫu nhiên  $X$  cho ta số liệu sau:

Giá trị của $X$	10 – 11	11 – 12	12 – 13	13 – 14
Số lần	12	43	30	15

Hãy kiểm tra giả thuyết cho rằng  $X$  là đại lượng ngẫu nhiên có phân bố chuẩn ở mức ý nghĩa 0.05.

Tóm tắt về Python

---



## Chương 7

# Kiểm định giả thuyết thống kê

7.1 a)  $\left( \bar{x} - z_0 \frac{\sigma}{\sqrt{n}}, \bar{x} + z_0 \frac{\sigma}{\sqrt{n}} \right) = (13.2172, 14.9161) \text{ phút.}$

b)

Định mức thời gian gia công một sản phẩm 15 phút là nhiều

$\Leftrightarrow$  thời gian gia công trung bình  $< 15$  phút

$\Leftrightarrow EX < 15.$

Xét bài toán  $H_0 : EX = 15, H_1 : EX < 15, \alpha = 2\%.$

$z_{qs} = -2.5560, z_0 = 2.0537.$

$z_{qs} < -z_0 \Rightarrow$  bác bỏ  $H_0$  (chấp nhận  $H_1$ ), tức là định mức thời gian gia công một sản phẩm 15 phút là nhiều.

7.2 a)  $\bar{x} = 5, s'^2 = 8.5$

b)  $\left( \bar{x} - t_0 \frac{s}{\sqrt{n-1}}, \bar{x} + t_0 \frac{s}{\sqrt{n-1}} \right) = (2.4403, 7.5596)$

c)  $t_{qs} = 2.0580, t_0 = 1.8595. t_{qs} > t_0 \Rightarrow$  bác bỏ  $H_0.$

7.3

Định mức thời gian chạy 17 giờ là nhiều

$\Leftrightarrow$  Thời gian chạy trung bình  $< 17$  giờ

$\Leftrightarrow EX < 17.$

trong đó  $X$  là thời gian của một chuyến tàu chạy từ A đến B.

Xét bài toán  $H_0 : EX = 17, H_1 : EX < 17, \alpha = 1\%.$



$$t_{qs} = \frac{\bar{x} - a_0}{s} \sqrt{n-1} = -13.1806, t_0 = t_{2\alpha}^{n-1} = 2.4922.$$

$t_{qs} < -t_0 \Rightarrow$  bác bỏ  $H_0$ , tức là, định mức thời gian chạy 17 giờ là nhiều.

**7.4** a)  $\bar{x} = 169.3333, s'^2 = 87.4713$

b)  $\left( \frac{ns^2}{\chi^2(\frac{1-\gamma}{2}, n-1)}, \frac{ns^2}{\chi^2(\frac{1+\gamma}{2}, n-1)} \right) = (52.9413, 169.089)$

c) Xét bài toán  $H_0 : EX = 167, H_1 : EX > 167, \alpha = 2\%$ .

$$t_{qs} = 1.3665, t_0 = 2.1503.$$

$t_{qs} < t_0 \Rightarrow$  chấp nhận  $H_0$ , (bác bỏ  $H_1$ ), tức là chưa thể khẳng định lượng điện sử dụng trung bình vượt quá 167 kWh.

**7.5**

$X$	2.275	2.325	2.375	2.425
$n_i$	4	11	13	3

a)  $\bar{x} = 2.3492, s'^2 = 0.001812$

b)  $\left( \bar{x} - t_0 \frac{s}{\sqrt{n-1}}, \bar{x} + t_0 \frac{s}{\sqrt{n-1}} \right) = (2.3282, 2.3702)$

c)  $t_{qs} = -6.6457, t_0 = 2.1470. t_{qs} < -t_0 \Rightarrow$  bác bỏ  $H_0$ .

**7.6** a)  $\bar{x} = 4, s' = 0.5$

b)  $\left( \bar{x} - t_0 \frac{s}{\sqrt{n-1}}, \bar{x} + t_0 \frac{s}{\sqrt{n-1}} \right) = (3.6443, 4.3557)$

c)  $t_{qs} = 1.99, t_0 = 1.9481. t_{qs} > t_0 \Rightarrow$  bác bỏ  $H_0$ .

**7.7** Bài toán  $H_0 : EX = EY, H_1 : EX \neq EY, \alpha = 10\%$ .

$$t_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{ns_X^2 + ms_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}} = -1.7321, t_0 = t_{\alpha}^{n+m-2} = 1.6772.$$

$|t_{qs}| > t_0 \Rightarrow$  bác bỏ  $H_0$ .

**7.8** a)  $\left( \bar{x} - t_0 \frac{s}{\sqrt{n-1}}, \bar{x} + t_0 \frac{s}{\sqrt{n-1}} \right) = (3585.4599, 3596.2642)$

b)

thép chưa đạt yêu cầu kỹ thuật

$\Leftrightarrow$  cường độ chịu kéo trung bình  $< 3600 \text{ kg/cm}^2$

$\Leftrightarrow EX < 3600.$

Xét bài toán  $H_0 : EX = 3600, H_1 : EX < 3600, \alpha = 1\%$ .

$$t_{qs} = -4.6742, t_0 = 2.4671.$$

$t_{qs} < -t_0 \Rightarrow$  bác bỏ  $H_0$ , tức là thép chưa đạt yêu cầu.

### 7.9

sinh viên năm nay khá hơn năm ngoái

$\Leftrightarrow$  tỷ lệ sinh viên phải học lại môn xác suất < năm ngoái

$\Leftrightarrow p < 20\%$ .

Xét bài toán  $H_0 : p = 0.2, H_1 : p < 0.2, \alpha = 5\%$ .

$$z_{qs} = \frac{\frac{m}{n} - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} = -1.75, z_0 = 1.6449.$$

$z_{qs} < -z_0 \Rightarrow$  bác bỏ  $H_0$ , tức là học sinh năm nay khá hơn.

$$7.10 \quad z_{qs} = \frac{49.27 - 50}{0.1} \sqrt{25} = -36.5, z_0 = 1.6449.$$

$z_{qs} < -z_0 \Rightarrow$  bác bỏ  $H_0$ , tức là phản ánh của khách hàng về việc khối lượng bao bị thiếu là có cơ sở.

$$7.11 \quad z_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{DX}{n} + \frac{DY}{m}}} = 2.5036, z_0 = 1.4051.$$

$z_{qs} > z_0 \Rightarrow$  bác bỏ  $H_0$ .

$$7.12 \quad a) \left( \bar{x} - t_0 \frac{s}{\sqrt{n-1}}, \bar{x} + t_0 \frac{s}{\sqrt{n-1}} \right) = (0.1072, 72.1428)$$

b)  $H_0 : EX = EY, H_1 : EX > EY, \alpha = 0.01$ .

$$t_{qs} = \frac{\bar{x} - \bar{y}}{\sqrt{ns_X^2 + ms_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}} = 3.0729, t_0 = 2.6810.$$

$t_{qs} > t_0 \Rightarrow$  bác bỏ  $H_0$ .

7.13 Kiểm định  $X \sim U(10, 16)$ , tức có hàm mật độ

$$f(x) = \begin{cases} \frac{1}{16-10}, & x \in [10, 16] \\ 0, & x \notin [10, 16]. \end{cases}$$

$S_i$	$n_i$	$p_{i0}$	$e_i$	$\frac{(n_i - e_i)^2}{e_i}$	Cách tính
$-\infty \rightarrow 11$	95	1/6	100	0.25	$\int_{-\infty}^{11} f(x) dx$
$11 \rightarrow 12$	104	1/6	100	0.16	$600 \cdot 1/6$
$12 \rightarrow 13$	96	1/6	100	0.16	$\frac{(96 - 100)^2}{100}$
$13 \rightarrow 14$	106	1/6	100	0.36	
$14 \rightarrow 15$	93	1/6	100	0.49	
$15 \rightarrow \infty$	106	1/6	100	0.36	
$\Sigma$				1.78	

$$\chi_{qs}^2 = 1.78.$$

$$\chi_0^2 = \chi^2(\alpha, h - 1) = \chi^2(3\%, 6 - 1) = 12.3746.$$

$$\chi_{qs}^2 > \chi_0^2 \Rightarrow \text{chấp nhận } X \sim U(10, 16).$$

**7.14**  $\chi_{qs}^2 = 7.5333, \chi_0^2 = 12.3746. \chi_{qs}^2 < \chi_0^2 \Rightarrow \text{chấp nhận } X \in U[10, 22].$

**7.15** Kiểm định thời gian gia công một chi tiết máy có phân bố chuẩn  $N(16, 3^2)$  :  $\mu = 16, \sigma = 3$ , tức là có hàm mật độ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$S_i$	$n_i$	$p_{i0}$	$e_i$	$\frac{(n_i - e_i)^2}{e_i}$
$-\infty \rightarrow 12$	16	0.09121	18.2422	0.2756
$12 \rightarrow 14$	35	0.1613	32.2563	0.2334
$14 \rightarrow 16$	49	0.2475	49.5015	0.005081
$16 \rightarrow 18$	50	0.2475	49.5015	0.005020
$18 \rightarrow 20$	30	0.1613	32.2563	0.1578
$20 \rightarrow \infty$	20	0.09121	18.2422	0.1694
$\Sigma$				0.8463

$$\chi_{qs}^2 = 0.8463, \chi_0^2 = \chi^2(\alpha, h - 1) = \chi^2(10\%, 6 - 1) = 9.2364.$$

$$\chi_{qs}^2 < \chi_0^2 \Rightarrow \text{chấp nhận thời gian gia công có phân bố } N(16, 3^2).$$

**7.16** •  $H_0 : X \in N(\mu, \sigma^2); H_1 : X \notin N(\mu, \sigma^2); \alpha = 0.05.$

Thay  $\mu, \sigma^2$  bởi ước lượng theo phương pháp bình phương nhỏ nhất  $\mu^* = \bar{x} = 11.98, \sigma^{*2} = s^2 \Rightarrow \sigma^* = s = 0.8886.$

• Xét bài toán  $H_0^* : X \in N(\mu^*, \sigma^{*2}); H_1^* : X \notin N(\mu^*, \sigma^{*2}); \alpha = 0.05.$

$$\chi_{qs}^2 = 2.6603, \chi_0^2 = \chi^2(\alpha, h - r - 1) = \chi^2(0.04, 4 - 2 - 1) = 12.7062.$$

$$\chi_{qs}^2 < \chi_0^2 \Rightarrow \text{chấp nhận } H_0^*, \text{ nên chấp nhận } H_0, \text{ tức là } X \text{ có phân bố chuẩn.}$$

