ÔN THI CAO HỌC
MÔN TOÁN KINH TẾ
(GV: Trần Ngọc Hội - 2009)

PHẦN III: THỐNG KÊ

§1. CÁC ĐẶC TRƯNG MẪU

1.1. Bảng số liệu

Khi khảo sát đám đông X ta thu thập số liệu của mẫu *cỡ* n: ($X_1, X_2, \ldots, X_n$) và thường lập bảng số liệu theo các dạng sau:

Dạng 1: Liệt kê dưới dạng:

$$x_1, x_2, \ldots, x_n$$

trong đó mỗi số liệu có thể lặp lại nhiều lần.

Dạng 2: Lập bảng có dạng:

X_i	x_1	x_2	………………………..	x_k
n_i	n_1	n_2	………………………….	n_k

trong đó $x_1 < x_2 < \ldots < x_k$ và mỗi số liệu x_i xuất hiện n_i lần.

Dạng 3: Lập bảng có dạng:

X_i	x_1 - x_2	x_2 - x_3	………………………..	x_k - x_{k+1}
n_i	n_1	n_2	………………………….	n_k

trong đó $x_1 < x_2 < \ldots < x_k < x_{k+1}$ và mỗi nửa khoảng $[x_i; x_{i+1})$ (trừ cái cuối cùng là đoạn $[x_k; x_{k+1}]$) chứa n_i số liệu.

Khi xử lý số liệu ta sẽ đưa số liệu về Dạng 2.

Có thể đưa Dạng 1 về Dạng 2 bằng cách thống kê lại.

Dạng 3 được đưa về Dạng 2 bằng cách thay các khoảng x_i-x_{i+1} bằng giá trị trung bình của hai đầu mút $x'_i = \dfrac{x_i + x_{i+1}}{2}$.

Trong các phần sau, ta xét mẫu của đám đông X có dạng 2.

1.2. Kỳ vọng mẫu

1) Định nghĩa. *Kỳ vọng mẫu* hay *Trung bình mẫu* của đám đông X ứng với mẫu ($X_1, X_2, \ldots, X_n$), kí hiệu $\overline{X}_n$ hay $\overline{X}$ là đại lượng ngẫu nhiên định bởi:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{k} X_i n_i}$$

2) Ý nghĩa. Khi $n \to \infty$ kỳ vọng mẫu $\overline{X}_n$ hội tụ về kỳ vọng đám đông $\mu = M(X)$. Do đó khi n khá lớn ta xấp xỉ:

$$\mu = M(X) \approx \overline{X}_n$$

1.3. Phương sai mẫu và độ lệch mẫu

1) **Định nghĩa.** *Phương sai mẫu* của đám đông X ứng với mẫu ($X_1, X_2, \ldots, X_n$), kí hiệu $\hat{S}^2$ (còn kí hiệu là $x\sigma_n^2$ hay σ_n^2) là đại lượng ngẫu nhiên định bởi:

$$\boxed{\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{k} X_i^2 n_i - (\overline{X})^2}$$

Căn bậc hai của phương sai mẫu của X gọi là *độ lệch mẫu,* kí hiệu $\hat{S}$ (còn kí hiệu là $x\sigma_n$ hay σ_n):

$$\boxed{\hat{S} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} X_i^2 n_i - (\overline{X})^2}}$$

2) Phương sai mẫu và độ lệch mẫu hiệu chỉnh

Phương sai mẫu hiệu chỉnh của đám đông X ứng với mẫu ($X_1, X_2, \ldots, X_n$), kí hiệu S^2 (còn kí hiệu là $x\sigma_{n-1}^2$ hay σ_{n-1}^2) là đại lượng ngẫu nhiên định bởi:

$$\boxed{S^2 = \frac{n}{n-1}\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{k} X_i^2 n_i - \frac{n}{n-1}(\overline{X})^2}$$

Căn bậc hai của phương sai mẫu hiệu chỉnh của X gọi là *độ lệch mẫu hiệu chỉnh,* kí hiệu S (còn kí hiệu là $x\sigma_{n-1}$ hay σ_{n-1}):

$$\boxed{S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} X_i^2 n_i - \frac{n}{n-1}(\overline{X})^2}}$$

3) **Ý nghĩa**. Khi $n \to \infty$ phương sai mẫu hiệu chỉnh hội tụ về phương sai đám đông $\sigma^2 = D(X)$. Do đó khi n khá lớn ta xấp xỉ:

$$\sigma^2 = D(X) \approx S^2$$

1.4. Tỉ lệ mẫu

1) **Định nghĩa.** Ta xét đám đông với tỉ lệ các phần tử có tính chất A là p. Dấu hiệu X mà ta quan tâm là các phần tử của đám đông có tính chất A hay không: Nếu có, ta đặt X = 1; nếu không, ta đặt X = 0. Như vậy, đám đông X có phân phối Bernoulli $X \sim B(p)$ như sau:

X	0	1
P	q	p

(q = 1-p). Khi đó một mẫu cỡ n là một bộ gồm n đại lượng ngẫu nhiên ($X_1, X_2, \ldots, X_n$) mà mỗi X_i đều có cùng phân phối Bernoulli với X: $X_i \sim B(p)$, nghĩa là

X_i	0	1
P	q	p

Nói cách khác, mỗi X_i chỉ nhận hai giá trị: 0 (với xác suất q) và 1 (với xác suất p).

Tỉ lệ mẫu của đám đông X ứng với mẫu ($X_1, X_2,\ldots, X_n$), kí hiệu F_n, là đại lượng ngẫu nhiên định bởi:

$$F_n = \frac{1}{n}\sum_{i=1}^{k} X_i n_i$$

2) **Ý nghĩa**. Khi $n \to \infty$ tỉ lệ mẫu F_n hội tụ về tỉ lệ đám đông p. Do đó khi n khá lớn ta xấp xỉ:

$$p \approx F_n$$

3) **Chú ý**. Dưới Dạng 2 của bảng, việc tính giá trị của tỉ lệ mẫu rất đơn giản vì ta chỉ cần xác định số phần tử m thỏa tính chất A của mẫu cỡ n. Khi đó

$$F_n = \frac{m}{n} \quad .$$

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 19 cm trở xuống được xếp vào loại B. Hãy xác định kỳ vọng mẫu, phương sai mẫu, phương sai mẫu hiệu chỉnh, độ lệnh mẫu, độ lệnh mẫu hiệu chỉnh của chỉ tiêu X và tỉ lệ mẫu các sản phẩm loại B.

Giải. Trước hết ta thay các khoảng x_i - x_{i+1} bằng giá trị trung bình của hai đầu mút $x'_i = \dfrac{x_i + x_{i+1}}{2}$.

X_i	13	17	21	25	29	33	37
n_i	8	9	20	16	16	13	18

Ta có:

- Cỡ mẫu n = 100.
- Kỳ vọng mẫu của X là

$$\overline{X} = \frac{1}{n}\sum X_i n_i = 26{,}36\ (cm).$$

- Phương sai mẫu của X là:

$$\hat{S}^2 = \frac{1}{n}\sum X_i^2 n_i - \overline{X}^2 = (7{,}4452)^2 (cm^2).$$

- Độ lệch mẫu của X là: $\hat{S} = 7{,}4452$ (cm)
- Phương sai mẫu hiệu chỉnh của X là:

$$S^2 = \frac{n}{n-1}\hat{S}^2 = (7{,}4827)^2 (cm^2).$$

- Độ lệch mẫu hiệu chỉnh của X là: $S = 7{,}4827$(cm)

- Tỉ lệ mẫu các sản phẩm loại B là:

$$F_n = \frac{m}{n} = \frac{17}{100} = 0{,}17 = 17\%.$$

vì trong n = 100 sản phẩm có m = 8 + 9 = 17 sản phẩm có chỉ tiêu X nhỏ hơn hay bằng 19 cm, nghĩa là có m = 17 sản phẩm loại B.

1.5. Hướng dẫn sử dụng phần mềm thống kê trong các máy tính bỏ túi CASIO 500MS, 570MS, 500ES, 570ES,..) tính các đặc trưng mẫu:

Ví dụ. Xét lại ví dụ trên với bảng số liệu:

X_i	13	17	21	25	29	33	37
n_i	8	9	20	16	16	13	18

a) Đối với loại máy tính CASIO 500 và 570MS:

1) **Vào MODE SD**: Bấm $\boxed{\text{MODE}}$ (vài lần...) và bấm số ứng với SD, trên màn hình sẽ hiện lên chữ SD.
2) **Xóa bộ nhớ thống kê**: Bấm $\boxed{\text{SHIFT}}$ $\boxed{\text{MODE}}$ $\boxed{1}$ (màn hình hiện lên Stat clear) $\boxed{=}$ $\boxed{\text{AC}}$. Kiểm tra lại: Bấm nút tròn $\boxed{\nabla}$ hoặc $\boxed{\Delta}$ thấy n = và ở góc số 0 là đã xóa.
3) **Nhập số liệu**: Trình tự bấm như sau: $\boxed{\text{xi}}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{\text{ni}}$ $\boxed{\text{M}^+}$ (khi bấm $\boxed{\text{SHIFT}}$ $\boxed{,}$ trên màn hình hiện lên dấu ;). Cụ thể, ta bấm:

$\boxed{1}$ $\boxed{3}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{8}$ $\boxed{\text{M}^+}$

$\boxed{1}$ $\boxed{7}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{9}$ $\boxed{\text{M}^+}$

$\boxed{2}$ $\boxed{1}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{2}$ $\boxed{0}$ $\boxed{\text{M}^+}$

$\boxed{2}$ $\boxed{5}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{1}$ $\boxed{6}$ $\boxed{M^+}$

$\boxed{2}$ $\boxed{9}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{1}$ $\boxed{6}$ $\boxed{M^+}$

$\boxed{3}$ $\boxed{3}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{1}$ $\boxed{3}$ $\boxed{M^+}$

$\boxed{3}$ $\boxed{7}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{1}$ $\boxed{8}$ $\boxed{M^+}$

4) **Kiểm tra và sửa số liệu sai**: Bấm nút tròn $\boxed{\nabla}$ để kiểm tra việc nhập số liệu. Thấy số liệu nào sai thì để màn hình ngay số liệu đó, nhập số liệu đúng và bấm $\boxed{=}$ thì số liệu mới sẽ thay cho số liệu cũ.

Ví dụ. Nhập sai $\boxed{1}$ $\boxed{3}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{7}$ $\boxed{M^+}$. Khi kiểm tra ta thấy trên màn hình hiện ra:

- $x_1 = 13$
- Freq1 = 7 (sai)

Sửa như sau: Để màn hình ở Freq1 = 7, bấm $\boxed{8}$ $\boxed{=}$ thì nhận được số liệu đúng Freq1 = 8.

Số liệu nào bị nhập dư thì để màn hình ở số liệu đó và bấm $\boxed{\text{SHIFT}}$ $\boxed{M^+}$ thì toàn bộ số liệu đó (gồm giá trị của X và xác suất tương ứng) sẽ bị xóa. Chẳng hạn, nhập dư $\boxed{4}$ $\boxed{7}$ $\boxed{\text{SHIFT}}$ $\boxed{,}$ $\boxed{1}$ $\boxed{8}$ $\boxed{M^+}$. Khi kiểm tra ta thấy $x_8 = 47$ (dư). Ta để màn hình ở số liệu đó và bấm $\boxed{\text{SHIFT}}$ $\boxed{M^+}$ thì toàn bộ số liệu dư (gồm giá trị của X = 47 và tần số tương ứng 18) sẽ bị xóa.

Chú ý. Sau khi kiểm tra việc nhập số liệu xong, phải bấm $\boxed{\text{AC}}$ để xóa màn hình và thóat khỏi chế độ chỉnh sửa.

5) Đọc kết quả:

Đại lượng cần tìm	**Thao tác**	**Kết quả**	**Ghi chú**
Tổng bình phương $\sum X_i^2 n_i$	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{1}$ $\boxed{=}$	$\sum X^2 = 75028$	$\sum X_i^2 n_i = \sum X^2$
Tổng $\sum X_i n_i$	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{2}$ $\boxed{=}$	$\sum X = 2636$	$\sum X_i n_i = \sum X$
Cỡ mẫu n	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{3}$ $\boxed{=}$	n = 100	
Kỳ vọng mẫu $\overline{X}$	$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{1}$ $\boxed{=}$	$\overline{X} = 26.36$	
Độ lệch mẫu $\hat{S}$	$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{2}$ $\boxed{=}$	$x\sigma_n = 7.4452$	$\hat{S} = x\sigma_n$
Độ lệch mẫu hiệu chỉnh S	$\boxed{\text{SHIFT}}$ $\boxed{2}$ $\boxed{3}$ $\boxed{=}$	$x\sigma_{n-1} = 7.4827$	$S = x\sigma_{n-1}$

- Phương sai mẫu $\hat{S}^2 = (7,4452)^2$
- Phương sai mẫu hiệu chỉnh $S^2 = (7,4827)^2$

b) Đối với loại máy tính CASIO 500 và 570ES

1) Khai báo cột tần số: Bấm $\boxed{\text{SHIFT}}$ $\boxed{\text{SETUP}}$ $\boxed{\nabla}$ $\boxed{4}$ $\boxed{1}$

(Bấm $\boxed{\nabla}$ bằng cách bấm nút tròn xuống)

2) Vào Mode Thống kê: Bấm $\boxed{\text{MODE}}$ $\boxed{3}$ $\boxed{1}$ (hoặc $\boxed{\text{MODE}}$ $\boxed{2}$ $\boxed{1}$)

(Trên màn hình sẽ hiện lên chữ STAT)

3) Nhập số liệu: Như trong bảng sau:

Vị trí con trỏ khi nhập các x_i		X	FREQ	Vị trí con trỏ khi nhập các n_i
Bấm $\boxed{1}$ $\boxed{3}$ $\boxed{=}$	1	13	8	Bấm $\boxed{8}$ $\boxed{=}$
$\boxed{1}$ $\boxed{7}$ $\boxed{=}$	2	17	9	$\boxed{9}$ $\boxed{=}$
$\boxed{2}$ $\boxed{1}$ $\boxed{=}$	3	21	20	$\boxed{2}$ $\boxed{0}$ $\boxed{=}$
	4	...	...	

Chú ý: - Di chuyển con trỏ bằng cách bấm nút tròn
- Để sửa số liệu bị nhập sai, để con trỏ tại đó, nhập lại số liệu đúng và bấm =
- Muốn xoá số liệu nào, để con trỏ tại đó và bấm DEL
- Sau khi nhập xong, bấm AC

4) Kiểm tra và sửa số liệu sai: Bấm nút tròn để kiểm tra việc nhập số liệu. Thấy số liệu nào sai thì để con trỏ ngay số liệu đó, nhập số liệu đúng và bấm $\boxed{=}$ thì số liệu mới sẽ thay cho số liệu cũ.

Số liệu nào bị nhập dư thì để con trỏ ở số liệu đó và bấm $\boxed{\text{DEL}}$ thì toàn bộ số liệu đó (gồm giá trị của X và tần suất tương ứng) sẽ bị xóa.

Chú ý. Sau khi kiểm tra việc nhập số liệu xong, phải bấm $\boxed{\text{AC}}$ để xóa màn hình và thóat khỏi chế độ chỉnh sửa. Trong quá trình xử lý số liệu, muốn xem lại bảng số liệu thì bấm $\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{2}$

5) Đọc kết quả:

Đại lượng cần tìm	**Thao tác**	**Kết quả**	**Ghi chú**
Tổng bình phương $\sum X_i^2 n_i$	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{4}$ $\boxed{1}$ $\boxed{=}$	$\sum X^2 = 75028$	$\sum X_i^2 n_i = \sum X^2$
Tổng $\sum X_i n_i$	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{4}$ $\boxed{2}$ $\boxed{=}$	$\sum X = 2636$	$\sum X_i n_i = \sum X$
Cỡ mẫu n	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{5}$ $\boxed{1}$ $\boxed{=}$	n = 100	
Kỳ vọng mẫu $\overline{X}$	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{5}$ $\boxed{2}$ $\boxed{=}$	$\overline{X} = 26.36$	
Độ lệch mẫu $\hat{S}$	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{5}$ $\boxed{3}$ $\boxed{=}$	$x\sigma_n = 7.4452$	$\hat{S} = x\sigma_n$
Độ lệch mẫu hiệu chỉnh S	$\boxed{\text{SHIFT}}$ $\boxed{1}$ $\boxed{5}$ $\boxed{4}$ $\boxed{=}$	$x\sigma_{n-1} = 7.4827$	$S = x\sigma_{n-1}$

- Phương sai mẫu $\hat{S}^2 = (7,4452)^2$
- Phương sai mẫu hiệu chỉnh $S^2 = (7,4827)^2$

§2. ƯỚC LƯỢNG

2.1. Ước lượng điểm

Xét đám đông X và mẫu ($X_1, X_2,..., X_n$) ta có các ước lượng điểm không chệch sau:

1) Kỳ vọng mẫu $\overline{X}$ là ước lượng không chệch của kỳ vọng đám đông: $\mu = M(X) \approx \overline{X}$.

2) Phương sai mẫu hiệu chỉnh S^2 là ước lượng không chệch của phương sai đám đông: $\sigma^2 = D(X) \approx S^2$.

3) Tỉ lệ mẫu F_n là ước lượng không chệch của tỉ lệ đám đông: $p \approx F_n$.

Ví dụ: Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được xếp vào loại B. Hãy ước lượng giá trị trung bình, phương sai của chỉ tiêu X và tỉ lệ các sản phẩm loại B.

Giải. Trong Ví dụ 1 ở §1, ta đã tìm được:

- Kỳ vọng mẫu của X là $\overline{X} = 26{,}36\ (cm)$.
- Phương sai đã hiệu chỉnh của X là

$$S^2 = \frac{n}{n-1}\hat{S}^2 = (7{,}4827)^2 = 55{,}9903\ (cm^2).$$

- Tỉ lệ mẫu sản phẩm loại B là $F_n = 17\%$.

Ta ước lượng:

- Giá trị trung bình của X là

$$M(X) \approx \overline{X} = 26{,}36\ (cm).$$

- Phương sai của X là

$$D(X) \approx S^2 = 55{,}9903\ (cm^2).$$

- Tỉ lệ các sản phẩm loại B là

$$p \approx F_n = 17\%.$$

2.2. Ước lượng khoảng cho kỳ vọng

1) Ước lượng hai phía: Xét đám đông X và mẫu ($X_1, X_2,..., X_n$), ta có các công thức ước lượng khỏang (hai phía) cho kỳ vọng $\mu = M(X)$ với độ tin cậy $\gamma = 1 - \alpha$ như sau:

BẢNG 1A

ƯỚC LƯỢNG KHOẢNG CHO KỲ VỌNG $\mu = M(X)$ (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)

Trường hợp	Phương sai σ^2	Công thức
$n \geq 30$	Đã biết	$(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}; \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}})$
	Chưa biết	$(\overline{X} - z_\alpha \frac{S}{\sqrt{n}}; \overline{X} + z_\alpha \frac{S}{\sqrt{n}})$
$n < 30$ và X có phân phối chuẩn	Đã biết	$(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}; \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}})$
	Chưa biết	$(\overline{X} - t_\alpha^k \frac{S}{\sqrt{n}}; \overline{X} + t_\alpha^k \frac{S}{\sqrt{n}})$

- z_α thoả $\varphi(z_\alpha) = (1-\alpha)/2 = \gamma/2$ tra từ Bảng giá trị hàm Laplace
- t_α^k với $k = n - 1$ và $\alpha = 1 - \gamma$ tra từ Bảng Phân phối Student

• Tra Bảng hàm Laplace để xác dịnh z_α thỏa $\varphi(z_\alpha) = \frac{1-\alpha}{2} = \frac{\gamma}{2}$ ta được:

$\gamma = 1 - \alpha$	$\varphi(z_\alpha) = \gamma/2$	z_α
90%	0,45	1,65
91%	0,455	1,70
92%	0,46	1,75
93%	0,465	1,81
94%	0,47	1,88
95%	0,475	1,96
96%	0,48	2,06
97%	0,485	2,17
98%	0,49	2,33
99%	0,495	2,58

• Đôi khi giá trị z_α được cho dưới dạng $P(|Z| \leq z_\alpha) = 1 - \alpha = \gamma$ hay $P(Z \leq z_\alpha) = 0{,}5 + \frac{1-\alpha}{2} = 0{,}5 + \frac{\gamma}{2}$, trong đó $Z \sim N(0,1)$.

• Bảng phân phối Student ứng với $k = n - 1$ và $\alpha = 1 - \gamma$ cho ta giá trị t_α^k thỏa $P(|T| > t_\alpha^k) = \alpha = 1 - \gamma$, nghĩa là $P(|T| \leq t_\alpha^k) = 1 - \alpha = \gamma$. **Ví dụ**. Khi $k = 12$, $\alpha = 0{,}01$ ta có $t_\alpha^k = 3{,}055$.

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 19 cm trở xuống được xếp vào loại B.

a) Ước lượng giá trị trung bình của chỉ tiêu X với độ tin cậy 95%.

b) Ước lượng giá trị trung bình của chỉ tiêu X của những sản phẩm loại B với độ tin cậy 99% (Giả sử X có phân phối chuẩn).

Giải. a) Đây là bài toán ước lượng khoảng cho kỳ vọng $\mu = M(X)$ với độ tin cậy $\gamma = 1 - \alpha = 95\% = 0,95$.

Với các số liệu trên, trong §1, ta đã tìm được:

- Cỡ mẫu $n = 100$.
- $\overline{X} = 26,36\ (cm)$.
- $S^2 = (7,4827)^2\ (cm^2)$.

Vì $n \geq 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức ước lượng khoảng cho kỳ vọng:

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

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,95/2 = 0,475$. Tra bảng gia trị hàm Laplace ta được $z_\alpha = 1,96$. Vậy ước lượng khoảng là:

$$(26,36 - 1,96\frac{7,4827}{\sqrt{100}}; 26,36 + 1,96\frac{7,4827}{\sqrt{100}}) = (24,89; 27,83).$$

Nói cách khác, với độ tin cậy 95%, giá trị trung bình của chỉ tiêu X từ 24,89cm đến 27,83 cm.

b) Đây là bài toán ước lượng khoảng cho kỳ vọng $\mu_B = M(X_B)$ của chỉ tiêu $X = X_B$ của những sản phẩm loại B với độ tin cậy $\gamma = 1 - \alpha = 99\% = 0,99$.

Ta lập bảng số liệu của X_B:

X_{Bi}	13	17
n_{Bi}	8	9

Từ bảng trên ta tính được:

$$n_B = 17;\ \sum X_{Bi} n_{Bi} = 257;\ \sum X_{Bi}{}^2 n_{Bi} = 3.953.$$

- Kỳ vọng mẫu của X_B là

$$\overline{X}_B = \frac{1}{n_B}\sum X_{Bi} n_{Bi} = 15,1176\ (cm).$$

- Phương sai mẫu của X_B là:

$$\hat{S}_B^2 = \frac{1}{n_B}\sum X_{Bi}{}^2 n_{Bi} - \overline{X}_B{}^2 = (1,9965)^2 (cm^2).$$

- Phương sai mẫu đã hiệu chỉnh của X_B là:

$$S_B{}^2 = \frac{n_B}{n_B - 1}\hat{S}_B^2 = (2,0580)^2\ (cm^2).$$

Vì $n_B < 30$, X_B có phân phối chuẩn, $\sigma^2{}_B = D(X_B)$ chưa biết, nên ta có công thức ước lượng khoảng cho kỳ vọng:

$$(\overline{X}_B - t_\alpha^k \frac{S_B}{\sqrt{n_B}}; \overline{X}_B + t_\alpha^k \frac{S_B}{\sqrt{n_B}})$$

trong đó t_α^k được xác định từ bảng phân phối Student với $k = n_B - 1 = 16$ và $\alpha = 1 - \gamma = 1 - 0,99 = 0,01$. Tra bảng phân phối Student ta được $t_\alpha^k = 2,921$.

Vậy ước lượng khoảng là:

$$(15,1176 - 2,921\frac{2,0580}{\sqrt{17}}; 15,1176 + 2,921\frac{2,0580}{\sqrt{17}}) = (13,66; 16,58).$$

Nói cách khác, với độ tin cậy 99%, giá trị trung bình của chỉ tiêu X của những sản phẩm loại B từ 13,66cm đến 16,58cm.

2) Ước lượng một phía: Xét đám đông X và mẫu $(X_1, X_2,..., X_n)$, ta có các công thức ước lượng khỏang một phía cho kỳ vọng $\mu = M(X)$ với độ tin cậy $\gamma = 1 - \alpha$ như sau:

BẢNG 1B

ƯỚC LƯỢNG KHOẢNG BÊN TRÁI CHO KỲ VỌNG $\mu = M(X)$ (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)		
Trường hợp	**Phương sai σ^2**	**Công thức**
n ≥ 30	Đã biết	$(-\infty; \overline{X} + z_{2\alpha}\frac{\sigma}{\sqrt{n}})$
	Chưa biết	$(-\infty; \overline{X} + z_{2\alpha}\frac{S}{\sqrt{n}})$
n < 30 và X có phân phối chuẩn	Đã biết	$(-\infty; \overline{X} + z_{2\alpha}\frac{\sigma}{\sqrt{n}})$
	Chưa biết	$(-\infty; \overline{X} + t_{2\alpha}^k\frac{S}{\sqrt{n}})$
• $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ $(\alpha = 1 - \gamma)$ tra từ Bảng giá trị hàm Laplace • $t_{2\alpha}^k$ với $k = n - 1$ và $\alpha = 1 - \gamma$ tra từ Bảng Phân phối Student		

BẢNG 1C

ƯỚC LƯỢNG KHOẢNG BÊN PHẢI CHO KỲ VỌNG $\mu = M(X)$ (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)		
Trường hợp	**Phương sai σ^2**	**Công thức**
n ≥ 30	Đã biết	$(\overline{X} - z_{2\alpha}\frac{\sigma}{\sqrt{n}}; +\infty)$
	Chưa biết	$(\overline{X} - z_{2\alpha}\frac{S}{\sqrt{n}}; +\infty)$
n < 30 và X có phân phối chuẩn	Đã biết	$(\overline{X} - z_{2\alpha}\frac{\sigma}{\sqrt{n}}; +\infty)$
	Chưa biết	$(\overline{X} - t_{2\alpha}^k\frac{S}{\sqrt{n}}; +\infty)$
• $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ $(\alpha = 1 - \gamma)$ tra từ Bảng giá trị hàm Laplace • $t_{2\alpha}^k$ với $k = n - 1$ và $\alpha = 1 - \gamma$ tra từ Bảng Phân phối Student		

Chú ý:

- Khi có ước lượng khoảng bên trái cho kỳ vọng μ với độ tin cậy γ là $(-\infty; \overline{X}+\varepsilon)$, ta nói *giá trị tối đa* của kỳ vọng μ độ tin cậy γ là $\overline{X}+\varepsilon$.
- Khi có ước lượng khoảng bên phải cho kỳ vọng μ với độ tin cậy γ là $(\overline{X}-\varepsilon; +\infty;)$, ta nói *giá trị tối thiểu* của kỳ vọng μ độ tin cậy γ là $\overline{X}-\varepsilon$.

Ví dụ. Tiếp tục xét lại Ví dụ trên.

c) Ước lượng giá trị trung bình tối đa của chỉ tiêu X với độ tin cậy 95%.

d) Ước lượng giá trị trung bình tối thiểu của chỉ tiêu X của những sản phẩm loại B với độ tin cậy 99% (Giả sử X có phân phối chuẩn).

Giải. c) Ta có độ tin cậy $\gamma = 1 - \alpha = 95\% = 0,95$ ($\alpha = 0,05$).

Ta đã tìm được:

- Cỡ mẫu $n = 100$.
- $\overline{X} = 26,36\ (cm)$.
- $S^2 = (7,4827)^2\ (cm^2)$.

Vì $n \geq 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức ước lượng khoảng bên trái cho kỳ vọng:

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

trong đó $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,90/2 = 0,45$. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha} = 1,65$. Suy ra giá trị trung bình tối đa của chỉ tiêu X với độ tin cậy 95% là:

$$\overline{X} + z_{2\alpha}\frac{S}{\sqrt{n}} = 26,36 + 1,65\frac{7,4827}{\sqrt{100}} = 27,5946(\text{cm}).$$

d) Ta có độ tin cậy $\gamma = 1 - \alpha = 99\% = 0,99$ ($\alpha = 0,01$).

Ta đã tìm được:

- Cỡ mẫu $n_B = 17$.
- $\overline{X}_B = 15,1176\ (\text{cm})$.
- $S_B^2 = (2,0580)^2\ (\text{cm}^2)$.

Vì $n_B < 30$, X_B có phân phối chuẩn, $\sigma_B{}^2 = D(X_B)$ chưa biết, nên ta có công thức ước lượng khoảng bên phải cho kỳ vọng với độ tin cậy $\gamma = 1 - \alpha$ là:

$$(\overline{X}_B - t_{2\alpha}^{k}\frac{S_B}{\sqrt{n_B}}; +\infty)$$

trong đó $t_{2\alpha}^{k}$ được xác định từ bảng phân phối Student với $k = n_B - 1 = 16$ và $2\alpha = 0,02$. Tra bảng phân phối Student ta được $t_{2\alpha}^{k} = 2,583$.

Vậy giá trị trung bình tối thiểu của chỉ tiêu X của những sản phẩm loại B với độ tin cậy 99% là:

$$\overline{X}_B - t_{2\alpha}^{k}\frac{S_B}{\sqrt{n_B}} = 15,1176 - 2,583\frac{2,0580}{\sqrt{17}} = 13,8283(\text{cm}).$$

2.3. Ước lượng khoảng cho tỉ lệ

1) Ước lượng hai phía: Xét đám đông X và mẫu ($X_1, X_2,..., X_n$), ta có các công thức ước lượng khỏang (hai phía) cho tỉ lệ $p = P(A)$ với độ tin cậy $\gamma = 1 - \alpha$ như sau:

BẢNG 2A

ƯỚC LƯỢNG KHOẢNG CHO TỈ LỆ $p = P(A)$ (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)
$(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}})$
z_α thoả $\varphi(z_\alpha) = (1 - \alpha)/2 = \gamma/2$ tra từ Bảng giá trị hàm Laplace

(F_n là tỉ lệ mẫu, φ là hàm Laplace). Độ chính xác của ước lượng là $\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}$.

Ví dụ. Để khảo sát trọng lượng của một loại vật nuôi, người ta quan sát một mẫu và có kết quả sau:

X(kg)	110-117	117-124	124-131	131-138	138-145	145-152	152-159
Số con	28	29	35	46	36	7	8

Những con có trọng lượng từ 145kg trở lên được xếp vào loại A. Hãy ước tỉ lệ con vật loại A với độ tin cậy 97%.

Giải. Đây là bài toán ước lượng khoảng cho tỉ lệ p các con loại A với độ tin cậy $\gamma = 1 - \alpha = 97\% = 0,97$.

Ta có công thức ước lượng khoảng :

$$(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}})$$

trong đó $\varphi(z_\alpha) = (1 - \alpha)/2 = \gamma/2 = 0,97/2 = 0,485$.

- Tra bảng giá trị hàm Laplace ta được $z_\alpha = 2,17$.
- Cỡ mẫu $n = 189$.
- Trong $n = 189$ con có $m = 7 + 8 = 15$ con có trọng lượng từ 145kg trở lên nên có $m = 15$ con loại A. Do đó tỉ lệ mẫu các con loại A là:

$$F_n = m/n = 15/189 = 0,0794.$$

Vậy ước lượng khoảng là:

$$(0,0794 - 2,17\sqrt{\frac{0,0794(1-0,0794)}{189}}; 0,0794 + 2,17\sqrt{\frac{0,0794(1-0,0794)}{189}})$$
$$= (0,0367; 0,1221) = (3,67\%; 12,21\%)$$

Nói cách khác, với độ tin cậy 97%, tỉ lệ con loại A từ 3,67% đến 12,21%.

3) Ước lượng một phía: Xét đám đông X và mẫu $(X_1, X_2,..., X_n)$, ta có các công thức ước lượng khoảng một phía cho tỉ lệ p = P(A) với độ tin cậy $\gamma = 1 - \alpha$ như sau:

BẢNG 2B

ƯỚC LƯỢNG KHOẢNG BÊN TRÁI CHO TỈ LỆ P = P(A) (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)
$(-\infty; F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}})$
$z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ $(\alpha = 1 - \gamma)$ tra từ Bảng giá trị hàm Laplace

BẢNG 2C

ƯỚC LƯỢNG KHOẢNG BÊN PHẢI CHO TỈ LỆ P = P(A) (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)
$(F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}}; +\infty)$
$z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ $(\alpha = 1 - \gamma)$ tra từ Bảng giá trị hàm Laplace

Chú ý:

- Khi có ước lượng khoảng bên trái cho tỉ lệ p với độ tin cậy γ là $(-\infty; F_n + \varepsilon)$, ta nói *giá trị tối đa* của tỉ lệ p với độ tin cậy γ là $F_n + \varepsilon$.
- Khi có ước lượng khoảng bên phải cho tỉ lệ p với độ tin cậy γ là $(F_n - \varepsilon; +\infty)$, ta nói *giá trị tối thiểu* của tỉ lệ p với độ tin cậy γ là $F_n - \varepsilon$.

Ví dụ. Tiếp tục xét lại Ví dụ trên.

c) Ước lượng tỉ lệ tối đa con loại A với độ tin cậy 96%.

d) Ước lượng tỉ lệ tối thiểu con loại A với độ tin cậy 98%.

Giải. Ta đã tìm được:

- Cỡ mẫu n = 189.
- Tỉ lệ mẫu con loại A là: $F_n = 0,0794$.

c) Ta có độ tin cậy $\gamma = 1 - \alpha = 96\% = 0,96$ $(\alpha = 0,04)$.

Công thức ước lượng khoảng bên trái cho tỉ lệ p con loại A với độ tin cậy $\gamma = 1 - \alpha = 0,96$ là:

$$(-\infty; F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}})$$

trong đó $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,92/2 = 0,46$. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha} = 1,75$. Suy ra tỉ lệ tối đa con loại A là:

$$F_n + z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}} = 0,0794 + 1,75\sqrt{\frac{0,0794(1-0,0794)}{189}} = 0,1138 = 11,38\%$$

d) Ta có độ tin cậy $\gamma = 1 - \alpha = 98\% = 0,98$ $(\alpha = 0,02)$.

Công thức ước lượng khoảng bên phải cho tỉ lệ p con loại A với độ tin cậy $\gamma = 1 - \alpha = 0,98$ là:

$$(F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}}; +\infty)$$

trong đó $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0,96/2 = 0,48$. Tra bảng giá trị hàm Laplace ta được $z_{2\alpha} = 2,06$. Suy ra tỉ lệ tối thiểu con loại A là:

$$F_n - z_{2\alpha}\sqrt{\frac{F_n(1-F_n)}{n}} = 0,0794 - 2,06\sqrt{\frac{0,0794(1-0,0794)}{189}} = 0,0389 = 3,89\%.$$

2.4. Ước lượng khoảng cho phương sai

1) Ước lượng hai phía: Xét đám đông X có phân phối chuẩn và mẫu $(X_1, X_2,..., X_n)$, ta có các công thức ước lượng khỏang cho phương sai $\sigma^2 = D(X)$ với độ tin cậy $\gamma = 1 - \alpha$ như sau:

BẢNG 3A

ƯỚC LƯỢNG KHOẢNG CHO PHƯƠNG SAI $\sigma^2 = D(X)$ (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)	
X có phân phối chuẩn	**Công thức**
1) $\mu = M(X)$ đã biết	$\left(\sum (X_i - \mu)^2 n_i / \chi^2_{\frac{\alpha}{2}}; \sum (X_i - \mu)^2 n_i / \chi^2_{1-\frac{\alpha}{2}}\right)$ (1)
2) $\mu = M(X)$ chưa biết	$\left((n-1)S^2 / \chi^2_{\frac{\alpha}{2}}; (n-1)S^2 / \chi^2_{1-\frac{\alpha}{2}}\right)$ (2)
(1) Tra $\chi^2_{\frac{\alpha}{2}}; \chi^2_{1-\frac{\alpha}{2}}$ $(\alpha = 1 - \gamma)$ từ Bảng P. phối Chi bình phương χ^2 với n bậc tự do (2) Tra $\chi^2_{\frac{\alpha}{2}}; \chi^2_{1-\frac{\alpha}{2}}$ $(\alpha = 1 - \gamma)$ từ Bảng P. phối Chi bình phương χ^2 với n-1 bậc tự do	

Chú ý:

1) $\sum (X_i - \mu)^2$ là tổng bình phương của mẫu $(X_1 - \mu, X_2 - \mu,..., X_n - \mu)$.

2) Bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(k)$ với k bậc tự do cho ta các giá trị χ^2_α thỏa $P(\chi^2 > \chi^2_\alpha) = \alpha$. **Ví dụ:** Với k = 20; $\alpha = 0,01$ ta có $\chi^2_\alpha = 37,57$ (Trong một số tài liệu khác, kí hiệu χ^2_α chỉ giá trị mà $P(\chi^2 \le \chi^2_\alpha) = \alpha$. Theo nghĩa này thì χ^2_α chính là giá trị $\chi^2_{1-\alpha}$ mà ta đã xét ở trên).

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Giả sử X có phân phối chuẩn. Hãy ước lượng phương sai của X với độ tin cậy 90% trong mỗi trường hợp sau:

a) Biết giá trị trung bình của X là 25cm.

b) Chưa biết giá trị trung bình của X.

Giải. a) Giả thiết cho ta $\mu = M(X) = 25$. Ta có ước lượng khoảng của phương sai với độ tin cậy $\gamma = 1 - \alpha = 90\%$ $(\alpha = 0,1)$ là:

$$\left(\frac{\sum (X_i-\mu)^2 n_i}{\chi^2_{\frac{\alpha}{2}}};\frac{\sum (X_i-\mu)^2 n_i}{\chi^2_{1-\frac{\alpha}{2}}}\right)$$

Ta lập bảng:

$X_i - \mu$	–12	– 8	– 4	0	4	8	12
n_i	8	9	20	16	16	13	18

Từ đó ta tìm được cỡ mẫu n = 100; $\sum (X_i-\mu)^2 n_i = 5728$.

Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(n)$ với n = 100 bậc tự do ta được:

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{0,05} = 124,3 \quad \text{và} \quad \chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0,95} = 77,93$$

Vậy ước lượng khoảng của phương sai là:

$$\left(\frac{5728}{124,3};\frac{5728}{77,93}\right) = (46,08;73,50)$$

Nói cách khác, với độ tin cậy 90%, phương sai của chỉ tiêu X của loại sản phẩm trên từ 46,08(cm^2) đến 73,50(cm^2).

b) Vì μ = M(X) chưa biết, ta có ước lượng khoảng của phương sai với độ tin cậy $\gamma = 1 - \alpha = 90\%$ ($\alpha = 0,1$) là:

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

Các số liệu của bài toán đã được tính trong các ví dụ trước. Nhắc lại rằng :

- Cỡ mẫu n = 100.
- $\overline{X} = 26,36\ (cm)$.
- $S^2 = (7,4827)^2\ (cm^2)$.

Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(n-1)$ với n – 1 = 99 ≈100 bậc tự do ta được:

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{0,05} = 124,3 \quad \text{và} \quad \chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0,95} = 77,93$$

Vậy ước lượng khoảng của phương sai là:

$$\left(\frac{99.(7,4827)^2}{124,3};\frac{99.(7,4827)^2}{77,93}\right) = (44,59;71,13)$$

Nói cách khác, với độ tin cậy 90%, phương sai của chỉ tiêu X của loại sản phẩm trên từ 44,59(cm^2) đến 71,13(cm^2).

2) Ước lượng một phía: Xét đám đông X có phân phối chuẩn và mẫu ($X_1, X_2,..., X_n$), ta có các công thức ước lượng khoảng một phía cho phương sai σ^2 = D(X) với độ tin cậy $\gamma = 1 - \alpha$ như sau:

BẢNG 3B

ƯỚC LƯỢNG KHOẢNG BÊN TRÁI CHO PHƯƠNG SAI σ^2 = D(X) (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)	
X có phân phối chuẩn	**Công thức**
1) μ = M(X) đã biết	$\left(0\,;\sum (X_i-\mu)^2 n_i / \chi^2_{1-\alpha}\right)$ (1)
2) μ = M(X) chưa biết	$\left(0\,;(n-1)S^2 / \chi^2_{1-\alpha}\right)$ (2)
(1) Tra $\chi^2_{1-\alpha}$ ($\alpha = 1 - \gamma$) từ Bảng Phân phối Chi bình phương χ^2 với n bậc tự do (2) Tra $\chi^2_{1-\alpha}$ ($\alpha = 1 - \gamma$) từ Bảng Phân phối chi bình phương χ^2 với n – 1 bậc tự do	

BẢNG 3C

ƯỚC LƯỢNG KHOẢNG BÊN PHẢI CHO PHƯƠNG SAI σ^2 = D(X) (ĐỘ TIN CẬY $\gamma = 1 - \alpha$)	
X có phân phối chuẩn	**Công thức**
1) μ = M(X) đã biết	$\left(\sum (X_i-\mu)^2 n_i / \chi^2_{\alpha}\,;+\infty\right)$ (1)
2) μ = M(X) chưa biết	$\left((n-1)S^2 / \chi^2_{\alpha}\,;+\infty\right)$ (2)
(1) Tra χ^2_{α} ($\alpha = 1 - \gamma$) từ Bảng Phân phối Chi bình phương χ^2 với n bậc tự do (2) Tra χ^2_{α} ($\alpha = 1 - \gamma$) từ Bảng Phân phối chi bình phương χ^2 với n – 1 bậc tự do	

Chú ý:

- Khi có ước lượng khoảng bên trái cho phương sai $\boldsymbol{\sigma^2}$ **= D(X)** với độ tin cậy γ là $(0;D)$, ta nói *giá trị tối đa* của phương sai $\boldsymbol{\sigma^2}$ với độ tin cậy γ là D.
- Khi có ước lượng khoảng bên phải cho phương sai $\boldsymbol{\sigma^2}$ **= D(X)** với độ tin cậy γ là (d; +∞), ta nói *giá trị thiểu* của phương sai $\boldsymbol{\sigma^2}$ với độ tin cậy γ là d.

Ví dụ. Tiếp tục xét lại Ví dụ trên. Với độ tin cậy 99%, hãy ước lượng:

c) Giá trị tối đa của phương sai $\boldsymbol{\sigma^2}$ trong trường hợp biết giá trị trung bình của X là 25cm.

d) Giá trị tối thiểu của phương sai $\boldsymbol{\sigma^2}$ trong trường hợp chưa biết giá trị trung bình của X.

Giải. c) Giả thiết cho ta μ = M(X) = 25. Ta có ước lượng khoảng bên trái của phương sai với độ tin cậy $\gamma = 1 - \alpha = 99\%$ ($\alpha = 0,01$) là:

$$\left(0\,;\frac{\sum (X_i-\mu)^2 n_i}{\chi^2_{1-\alpha}}\right)$$

Tương tự câu a), ta tìm được cỡ mẫu n = 100; $\sum (X_i-\mu)^2 n_i = 5728$.

Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(n)$ với n = 100 bậc tự do ta được:

$$\chi^2_{1-\alpha} = \chi^2_{0,99} = 70,065$$

Suy ra giá trị tối đa của phương sai $\boldsymbol{\sigma^2}$ là:

$$\frac{\sum (X_i-\mu)^2 n_i}{\chi^2_{1-\alpha}} = \frac{5728}{70,065} = 81,7527.$$

d) Vì $\mu = M(X)$ chưa biết, ta có ước lượng khoảng bên phải của phương sai với độ tin cậy $\gamma = 1 - \alpha = 99\%$ ($\alpha = 0,01$) là:

$$\left(\frac{(n-1)S^2}{\chi^2_\alpha}; +\infty\right)$$

Ta đã biết:

- Cỡ mẫu $n = 100$.
- $\overline{X} = 26,36\ (cm)$.
- $S^2 = (7,4827)^2\ (cm^2)$.

Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(n\text{-}1)$ với $n - 1 = 99 \approx 100$ bậc tự do ta được:

$$\chi^2_\alpha = \chi^2_{0,01} = 135,8.$$

Suy ra giá trị tối thiểu của phương sai $\boldsymbol{\sigma^2}$ là:

$$\frac{(n-1)S^2}{\chi^2_\alpha} = \frac{99.(7,4827)^2}{135,8} = 40,8180.$$

2.5. Các chỉ tiêu chính của bài toán ước lượng khoảng cho kỳ vọng và tỉ lệ

Trong bài toán ước lượng khoảng cho kỳ vọng và tỉ lệ có 3 chỉ tiêu chính là:

- Cỡ mẫu n.
- Độ chính xác ε.
- Độ tin cậy $\gamma = 1 - \alpha$.

Nếu biết được 2 trong 3 chỉ tiêu trên thì có thể suy ra chỉ tiêu còn lại.

1) Trương hợp ước lượng khoảng cho kỳ vọng

Ta xét trường hợp phổ biến nhất là $n \ge 30$; $\sigma^2 = D(X)$ chưa biết. Khi đó, ta có công thức ước lượng khoảng cho kỳ vọng $\mu = M(X)$ với độ tin cậy γ:

$$(\overline{X} - z_\alpha\frac{S}{\sqrt{n}}; \overline{X} + z_\alpha\frac{S}{\sqrt{n}}) \quad \text{với} \quad \varphi(z_\alpha) = \frac{\gamma}{2}.$$

Do đó ta có công thức độ chính xác của ước lượng là:

$$\varepsilon = z_\alpha\frac{S}{\sqrt{n}} \qquad (1)$$

- Nếu biết cỡ mẫu n và độ tin cậy γ thì ta tra bảng giá trị hàm Laplace để tìm z_α thoả $\varphi(z_\alpha) = \gamma/2$. Từ đó ta tìm được độ chính xác ε theo (1).

- Nếu biết cỡ mẫu n và độ chính xác ε thì từ (1) ta suy ra

$$z_\alpha = \frac{\varepsilon\sqrt{n}}{S}$$

Tra bảng giá trị hàm Laplace ta tìm được $\varphi(z_\alpha)$. Từ đó suy ra độ tin cậy $\gamma = 2\varphi(z_\alpha)$.

- Nếu biết độ chính xác ε và độ tin cậy γ thì từ (1) ta suy ra:

$$n = \left(\frac{z_\alpha S}{\varepsilon}\right)^2$$

Chú ý rằng $\left(\frac{z_\alpha S}{\varepsilon}\right)^2$ có thể không là số nguyên, hơn nữa, ta đã biết trong ước lượng, cỡ mẫu càng lớn thì ước lượng càng chính xác. Do đó trong thực tế ta có yêu cầu:

$$n \ge n_1 \qquad (2)$$

trong đó $n_1 = \left\lceil (z_\alpha S/\varepsilon)^2 \right\rceil$ là số nguyên nhỏ nhất lớn hơn hay bằng $(z_\alpha S/\varepsilon)^2$.

Gọi n_0 là cỡ mẫu đang xét, ta có:

Nếu $n_1 \le n_0$ thì ta không cần điều tra thêm vì cỡ mẫu đang có đã thỏa (2).

Nếu $n_1 > n_0$ thì ta cần điều tra thêm ít nhất là $n_1 - n_0$ số liệu nữa để đảm bảo tổng số liệu là n_1 thoả (2).

Tóm lại, ta có qui tắc xác định các chỉ tiêu chính khi ước lượng khoảng cho kỳ vọng như sau:

BẢNG 4A

XÁC ĐỊNH CÁC CHỈ TIÊU CHÍNH TRƯỜNG HỢP ƯỚC LƯỢNG KHOẢNG CHO KỲ VỌNG $\mu = M(X)$		
Chỉ tiêu đã biết	**Chỉ tiêu cần tìm**	**Công thức**
- Cỡ mẫu n **- Độ tin cậy $\gamma = 1 - \alpha$**	**Độ chính xác ε**	$\varepsilon = z_\alpha \frac{S}{\sqrt{n}}$
- Cỡ mẫu n **- Độ chính xác ε**	**Độ tin cậy $\gamma = 1 - \alpha$**	$\gamma = 2\varphi(\frac{\varepsilon\sqrt{n}}{S})$
- Độ tin cậy $\gamma = 1 - \alpha$ **- Độ chính xác ε**	**Cỡ mẫu n**	$n \ge \left\lceil (z_\alpha S/\varepsilon)^2 \right\rceil$
• z_α thoả $\varphi(z_\alpha) = (1-\alpha)/2 = \gamma/2$ tra từ Bảng giá trị hàm Laplace $\varphi(x)$ • $\left\lceil (z_\alpha S/\varepsilon)^2 \right\rceil$ là số nguyên nhỏ nhất $\ge (z_\alpha S/\varepsilon)^2$		

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

a) Nếu muốn ước lượng giá trị trung bình của chỉ tiêu X của loại sản phẩm trên với độ chính xác 1,8cm thì sẽ đạt được độ tin cậy là bao nhiêu?

b) Nếu muốn ước lượng giá trị trung bình của chỉ tiêu X của loại sản phẩm trên với độ chính xác 1,5cm và độ tin cậy 97% thì phải điều tra thêm ít nhất bao nhiêu sản phẩm nữa?

Giải. Các số liệu của bài toán đã được tính trong các ví dụ trước. Nhắc lại rằng :

- Cỡ mẫu $n = 100$.
- $\overline{X} = 26{,}36\ (cm)$.
- $S^2 = (7{,}4827)^2\ (cm^2)$.

a) Đây là bài toán xác định độ tin cậy $\gamma = 1 - \alpha$ khi ước lượng kỳ vọng của chỉ tiêu X với độ chính xác $\varepsilon = 1{,}8$cm.

Vì $n \geq 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha \frac{S}{\sqrt{n}}$$

trong đó $\varphi(z_\alpha) = \gamma/2$. Suy ra

$$z_\alpha = \frac{\varepsilon\sqrt{n}}{S} = \frac{1,8.\sqrt{100}}{7,4827} = 2,41$$

Tra bảng giá trị hàm Laplace ta được độ tin cậy là:

$$\gamma = 2\varphi(z_\alpha) = 2\varphi(2,41) = 2.0,4920 = 98,40\%.$$

Vậy độ tin cậy đạt được là 98,40%.

b) Đây là bài toán xác định cỡ mẫu khi ước lượng kỳ vọng của chỉ tiêu X với độ chính xác $\varepsilon = 1{,}5$cm và độ tin cậy $\gamma = 1 - \alpha = 97\% = 0{,}97$.

Vì $n \geq 30$, $\sigma^2 = D(X)$ chưa biết nên ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha \frac{S}{\sqrt{n}}$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0{,}97/2 = 0{,}485$.

Tra bảng giá trị hàm Laplace ta được $z_\alpha = 2{,}17$. Suy ra

$$n = \left(\frac{z_\alpha S}{\varepsilon}\right)^2 = \left(\frac{2,17.7,4827}{1,5}\right)^2 \approx 117,18$$

Thực tế yêu cầu: $n \geq \lceil 117{,}18 \rceil = 118$. Vì $n_1 = 118 > 100$ (100 là cỡ mẫu đang có) nên ta cần điều tra thêm ít nhất là $118 - 100 = 18$ sản phẩm nữa.

2) Trường hợp ước lượng khoảng cho tỉ lệ

Ta xét trường hợp cỡ mẫu khá lớn. Khi đó, ta có công thức ước lượng khoảng cho tỉ lệ p với độ tin cậy γ:

$$\left(F_n - z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}; F_n + z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}\right) \quad \text{với} \quad \varphi(z_\alpha) = \frac{1-\alpha}{2} = \frac{\gamma}{2}.$$

Do đó ta có công thức độ chính xác của ước lượng là:

$$\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}} \qquad (1)$$

- Nếu biết cỡ mẫu n và độ tin cậy γ thì ta tra bảng giá trị hàm Laplace để tìm z_α thoả $\varphi(z_\alpha) = \gamma/2$. Từ đó ta tìm được độ chính xác ε theo (1).

- Nếu biết cỡ mẫu n và độ chính xác ε thì từ (1) ta suy ra

$$z_\alpha = \varepsilon\sqrt{\frac{n}{F_n(1-F_n)}}$$

Tra bảng giá trị hàm Laplace ta tìm được $\varphi(z_\alpha)$. Từ đó suy ra độ tin cậy $\gamma = 2\varphi(z_\alpha)$.

- Nếu biết độ chính xác ε và độ tin cậy γ thì từ (1) ta suy ra:

$$n = \frac{z_\alpha^2 F_n(1-F_n)}{\varepsilon^2}$$

Chú ý rằng $\frac{z_\alpha^2 F_n(1-F_n)}{\varepsilon^2}$ có thể không là số nguyên, hơn nữa, ta đã biết trong ước lượng, cỡ mẫu càng lớn thì ước lượng càng chính xác. Do đó trong thực tế ta có yêu cầu:

$$n \geq n_1 \qquad (2)$$

trong đó $n_1 = \lceil z_\alpha^2 F_n(1-F_n)/\varepsilon^2 \rceil$ là số nguyên nhỏ nhất lớn hơn hay bằng $z_\alpha^2 F_n(1-F_n)/\varepsilon^2$. Gọi n_0 là cỡ mẫu đang xét, ta có:

Nếu $n_1 \leq n_0$ thì ta không cần điều tra thêm vì cỡ mẫu đang có đã thỏa (2).

Nếu $n_1 > n_0$ thì ta cần điều tra thêm ít nhất là $n_1 - n_0$ số liệu nữa để đảm bảo tổng số liệu là n_1 thoả (2).

Tóm lại, ta có qui tắc xác định các chỉ tiêu chính khi ước lượng khoảng cho tỉ lệ như sau:

BẢNG 4B

XÁC ĐỊNH CÁC CHỈ TIÊU CHÍNH TRƯỜNG HỢP ƯỚC LƯỢNG KHOẢNG CHO TỈ LỆ p = P(A)		
Chỉ tiêu đã biết	**Chỉ tiêu cần tìm**	**Công thức**
- Cỡ mẫu n - Độ tin cậy $\gamma = 1 - \alpha$	Độ chính xác ε	$\varepsilon = z_\alpha\sqrt{\frac{F_n(1-F_n)}{n}}$
- Cỡ mẫu n - Độ chính xác ε	Độ tin cậy $\gamma = 1 - \alpha$	$\gamma = 2\varphi(\varepsilon\sqrt{\frac{n}{F_n(1-F_n)}})$
- Độ tin cậy $\gamma = 1 - \alpha$ - Độ chính xác ε	Cỡ mẫu n	$n \geq \lceil z_\alpha^2 F_n(1-F_n)/\varepsilon^2 \rceil$
• z_α thoả $\varphi(z_\alpha) = (1-\alpha)/2 = \gamma/2$ tra từ Bảng giá trị hàm Laplace $\varphi(x)$ • $\lceil z_\alpha^2 F_n(1-F_n)/\varepsilon^2 \rceil$ là số nguyên nhỏ nhất $\geq z_\alpha^2 F_n(1-F_n)/\varepsilon^2$		

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được xếp vào loại B.

a) Nếu muốn ước lượng tỉ lệ các sản phẩm loại B với độ chính xác 8% thì sẽ đạt được độ tin cậy là bao nhiêu?

b) Nếu muốn ước lượng tỉ lệ các sản phẩm loại B với độ chính xác 9% và độ tin cậy 96% thì phải điều tra thêm ít nhất bao nhiêu sản phẩm nữa?

Giải. Các số liệu của bài toán đã được xét nhiều lần. Nhắc lại rằng:

- Cỡ mẫu $n = 100$.
- Tỉ lệ mẫu các sản phẩm loại B là $F_n = 0,17$.

a) Đây là bài toán xác định độ tin cậy $\gamma = 1 - \alpha$ khi lượng tỉ lệ các sản phẩm loại B với độ chính xác $\varepsilon = 8\% = 0,08$.

Ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha \sqrt{\frac{F_n(1-F_n)}{n}}$$

trong đó $\varphi(z_\alpha) = \gamma/2$. Suy ra

$$z_\alpha = \varepsilon\sqrt{\frac{n}{F_n(1-F_n)}} = 0,08.\sqrt{\frac{100}{0,17(1-0,17)}} = 2,13$$

Tra bảng giá trị hàm Laplace ta được độ tin cậy là

$$\gamma = 2\varphi(z_\alpha) = 2\varphi(2,13) = 2.0,4834 = 96,68\%.$$

Vậy độ tin cậy đạt được là 96,68%.

b) Đây là bài toán xác định cỡ mẫu khi ước lượng tỉ lệ các sản phẩm loại B với độ chính xác $\varepsilon = 9\% = 0,09$ và độ tin cậy $\gamma = 1 - \alpha = 96\% = 0,96$.

Ta có công thức tính độ chính xác của ước lượng:

$$\varepsilon = z_\alpha \sqrt{\frac{F_n(1-F_n)}{n}}$$

trong đó $\varphi(z_\alpha) = \gamma/2 = 0,96/2 = 0,48$.

Tra bảng giá trị hàm Laplace ta được $z_\alpha = 2,06$. Suy ra

$$n = \frac{z_\alpha^2 F_n(1-F_n)}{\varepsilon^2} = \frac{2,06^2.0,17(1-0,17)}{0,09^2} \approx 73,92.$$

Thực tế yêu cầu: $n \geq \lceil 73,92 \rceil = 74$. Vì $n_1 = 74 < 100$ (100 là cỡ mẫu đang có) nên ta không cần điều tra thêm sản phẩm nữa.

§3. KIỂM ĐỊNH GIẢ THIẾT

3.1. Kiểm định giả thiết về kỳ vọng

1) Kiểm định hai phía: Xét đám đông X có kỳ vọng $\mu = M(X)$ chưa biết. Với mỗi số α ($0 < \alpha < 1$) khá bé, dựa vào mẫu ($X_1, X_2,..., X_n$) ta có qui tắc kiểm định giả thiết hai phía về kỳ vọng $\mu = M(X)$ với mức ý nghĩa α như sau:

BẢNG 5A

KIỂM ĐỊNH GIẢ THIẾT VỀ KỲ VỌNG $\mu = M(X)$
H_0: $\mu = \mu_0$ với giả thiết đối H_1: $\mu \neq \mu_0$ (mức ý nghĩa α)

Trường hợp / Bước	$n \geq 30$		$n < 30$	
	σ^2 đã biết	σ^2 chưa biết	σ^2 đã biết	σ^2 chưa biết
1) Tính z	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S}$	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S}$
2) Tra Bảng	z_α	z_α	z_α	t_α^k
3a) Chấp nhận H_0	$\|z\| \leq z_\alpha$	$\|z\| \leq z_\alpha$	$\|z\| \leq z_\alpha$	$\|z\| \leq t_\alpha^k$
3b) Bác bỏ H_0	$\|z\| > z_\alpha$	$\|z\| > z_\alpha$	$\|z\| > z_\alpha$	$\|z\| > t_\alpha^k$

- z_α thoả $\varphi(z_\alpha) = (1-\alpha)/2$ tra từ Bảng giá trị hàm Laplace
- t_α^k với $k = n-1$ tra từ Bảng Phân phối Student

2) Kiểm định một phía: Xét đám đông X có kỳ vọng $\mu = M(X)$ chưa biết. Với mỗi số α ($0 < \alpha < 1$) khá bé, dựa vào mẫu ($X_1, X_2,..., X_n$) ta có qui tắc kiểm định giả thiết một phía về kỳ vọng $\mu = M(X)$ với mức ý nghĩa α như sau:

BẢNG 5B

KIỂM ĐỊNH GIẢ THIẾT VỀ KỲ VỌNG $\mu = M(X)$
H_0: $\mu = \mu_0$ với giả thiết đối H_1: $\mu > \mu_0$ (mức ý nghĩa α)

Trường hợp / Bước	$n \geq 30$		$n < 30$	
	σ^2 đã biết	σ^2 chưa biết	σ^2 đã biết	σ^2 chưa biết
1) Tính z	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S}$	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$	$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S}$
2) Tra Bảng	$z_{2\alpha}$	$z_{2\alpha}$	$z_{2\alpha}$	$t_{2\alpha}^k$
3a) Chấp nhận H_0	$z \leq z_{2\alpha}$	$z \leq z_{2\alpha}$	$z \leq z_{2\alpha}$	$z \leq t_{2\alpha}^k$
3b) Bác bỏ H_0	$z > z_{2\alpha}$	$z > z_{2\alpha}$	$z > z_{2\alpha}$	$z > t_{2\alpha}^k$

- $z_{2\alpha}$ thoa $\varphi(z_{2\alpha}) = (1-2\alpha)/2$ tra từ Bảng giá trị hàm Laplace
- $t_{2\alpha}^k$ với $k = n-1$ tra từ Bảng Phân phối Student

BẢNG 5C

KIỂM ĐỊNH GIẢ THIẾT VỀ KỲ VỌNG $\mu = M(X)$

H_0: $\mu = \mu_0$ với giả thiết đối H_1: $\mu < \mu_0$ (mức ý nghĩa α)

Trường hợp / Bước	$n \ge 30$		$n < 30$	
	σ^2 đã biết	σ^2 chưa biết	σ^2 đã biết	σ^2 chưa biết
1) Tính z	$z = \dfrac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$	$z = \dfrac{(\overline{X}-\mu_0)\sqrt{n}}{S}$	$z = \dfrac{(\overline{X}-\mu_0)\sqrt{n}}{\sigma}$	$z = \dfrac{(\overline{X}-\mu_0)\sqrt{n}}{S}$
2) Tra Bảng	$z_{2\alpha}$	$z_{2\alpha}$	$z_{2\alpha}$	$t^k_{2\alpha}$
3a) Chấp nhận H_0	$-z \le z_{2\alpha}$	$-z \le z_{2\alpha}$	$-z \le z_{2\alpha}$	$-z \le t^k_{2\alpha}$
3b) Bác bỏ H_0	$-z > z_{2\alpha}$	$-z > z_{2\alpha}$	$-z > z_{2\alpha}$	$-z > t^k_{2\alpha}$

- $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ tra từ Bảng giá trị hàm Laplace
- $t^k_{2\alpha}$ với $k = n - 1$ tra từ Bảng Phân phối Student

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được xếp vào loại B.

a) Giả sử trung bình tiêu chuẩn của chỉ tiêu X là 29cm. Hãy nhận định về tình hình sản xuất với mức ý nghĩa 1%.

b) Theo qui định, gía trị trung bình của chỉ tiêu X là 25cm. Các số liệu trên thu thập được từ các sản phẩm do một máy sản xuất. Với mức ý nghĩa 2% có thể kết luận rằng các sản phẩm do máy sản suất có chỉ tiêu X cao hơn qui định hay không?

c) Bằng phương pháp sản xuất mới, sau một thời gian, người ta thấy giá trị trung bình của chỉ tiêu X của những sản phẩm loại B là 16cm. Hãy cho kết luận về phuong pháp mới với mức ý nghĩa 2% (Giả sử X có phân phối chuẩn).

d) Theo số liệu thống kê cũ, gía trị trung bình của chỉ tiêu X của những sản phẩm loại B là 16,5cm. Các số liệu trên thu thập được sau khi đã áp dụng một phương pháp sản xuất mới. Hãy cho kết luận về nhận định cho rằng phương pháp mới có tác dụng làm giảm chỉ tiêu X của những sản phẩm loại B với mức ý nghĩa 2% (Giả sử X có phân phối chuẩn).

Giải. Các số liệu của bài toán đã tính được:

- Cỡ mẫu $n = 100$.
- Kỳ vọng mẫu của X: $\overline{X} = 26{,}36\,(cm)$.
- Phương sai mẫu hiệu chỉnh của X: $S^2 = (7{,}4827)^2\ (cm^2)$.
- Cỡ mẫu loại B: $n_B = 17$.
- Kỳ vọng mẫu của X_B: $\overline{X}_B = 15{,}1176\,(cm)$.

- Phương sai mẫu hiệu chỉnh của X_B: $S_B^{\,2} = (2{,}0580)^2\ (cm^2)$.

a) Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu = M(X)$ với mức ý nghĩa $\alpha = 1\% = 0{,}01$:

H_0: $\mu = 29$ với giả thiết đối **H_1**: $\mu \ne 29$.

Vì $n \ge 30$; $\sigma^2 = D(X)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} = \frac{(26,36-29)\sqrt{100}}{7,4827} = -3,5281.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$\varphi(z_\alpha) = (1 - \alpha)/2 = 0{,}99/2 = 0{,}495$

ta được $z_\alpha = 2{,}58$.

Bước 3: Kiểm định.

Vì $|z| = 3{,}5281 > 2{,}58 = z_\alpha$ nên ta bác bỏ giả thiết H_0: $\mu = 29$, nghĩa là chấp nhận H_1: $\mu \ne 29$.

Kết luận: Với mức ý nghĩa 1%, tình hình sản xuất không bình thường vì giá trị trung bình của chỉ tiêu X không đúng tiêu chuẩn.

b) Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu = M(X)$ với mức ý nghĩa $\alpha = 2\% = 0{,}02$:

H_0: $\mu = 25$ với giả thiết đối **H_1**: $\mu > 25$.

Vì $n \ge 30$; $\sigma^2 = D(X)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\overline{X}-\mu_0)\sqrt{n}}{S} = \frac{(26,36-25)\sqrt{100}}{7,4827} = 1,8175.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2 = 0{,}96/2 = 0{,}48$ ta được $z_{2\alpha} = 2{,}06$.

Bước 3: Kiểm định.

Vì $z = 1{,}18175 < 2{,}06 = z_{2\alpha}$ nên ta chấp nhận gia thiết H_0: $\mu = 25$.

Kết luận: Với mức ý nghĩa 2%, không thể kết luận rằng các sản phẩm do máy trên sản suất có chỉ tiêu X cao hơn qui định.

c) Đây là bài toán kiểm định giả thiết về kỳ vọng $\mu_B = M(X_B)$ của chỉ tiêu $X = X_B$ của các sản phẩm loại B với mức ý nghĩa $\alpha = 2\% = 0{,}02$:

H_0: $\mu_B = 16$ với giả thiết đối **H_1**: $\mu_B \ne 16$

Vì $n_B < 30$, X_B có phân phối chuẩn, $\sigma^2_B = D(X_B)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\overline{X}_B-\mu_0)\sqrt{n_B}}{S_B} = \frac{(15,1176-16)\sqrt{17}}{2,0580} = -1,7678.$$

Bước 2: Đặt $k = n_B - 1 = 16$. Tra bảng phân phối Student ứng với $k = 16$ và $\alpha = 0,02$ ta được $t_\alpha^k = 2,583$.

Bước 3: Kiểm định.

Vì $|z| = 1,7678 < 2,583 = t_\alpha^k$ nên ta chấp nhận giả thiết H_0: $\mu_B = 16$.

Kết luận: Với mức ý nghĩa 2%, phương pháp mới không có tác dụng làm thay đổi giá trị trung bình của chỉ tiêu X_B của các sản phẩm loại B.

d) Đây là bài toán kiểm định giả thiếtvề kỳ vọng $\mu_B = M(X_B)$ của chỉ tiêu $X = X_B$ của các sản phẩm loại B với mức ý nghĩa $\alpha = 2\% = 0,02$:

$$\mathbf{H_0}: \mu_B = 16,5 \text{ với giả thiết đối } \mathbf{H_1}: \mu_B < 16,5$$

Vì $n_B < 30$, X_B có phân phối chuẩn, $\sigma^2_B = D(X_B)$ chưa biết, nên ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(\overline{X}_B - \mu_0)\sqrt{n_B}}{S_B} = \frac{(15,1176 - 16,5)\sqrt{17}}{2,0580} = -2,7696.$$

Bước 2: Đặt $k = n_B - 1 = 16$. Tra bảng phân phối Student ứng với $k = 16$ và $2\alpha = 0,04$ ta được $t_{2\alpha}^k = 2,2354$.

Bước 3: Kiểm định.

Vì $-z = 2,7696 > 2,2354 = t_{2\alpha}^k$ nên ta bác bỏ giả thiết H_0: $\mu_B = 16,5$, nghĩa là chấp nhận H_1: $\mu_B < 16,5$.

Kết luận: Với mức ý nghĩa 2%, phương pháp mới có tác dụng làm giảm giá trị trung bình của chỉ tiêu X_B của các sản phẩm loại B.

3.2. Kiểm định giả thiết về tỉ lệ

1) Kiểm định hai phía: Xét đám đông X có tỉ lệ $p = P(A)$ chưa biết. Với mỗi số α $(0 < \alpha < 1)$ khá bé, dựa vào mẫu $(X_1, X_2,..., X_n)$ ta có qui tắc kiểm định giả thiết hai phía về tỉ lệ $p = P(A)$ với mức ý nghĩa α như sau:

BẢNG 6A

KIỂM ĐỊNH GIẢ THIẾT VỀ TỈ LỆ p = P(A) **H_0: p = p_0 với giả thiết đối H_1: p ≠ p_0 (mức ý nghĩa α)**	
Bước 1: Tính z	$z = \dfrac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$
Bước 2: Tra Bảng	z_α
Bước 3a: Chấp nhận H_0	$\lvert z\rvert \le z_\alpha$
Bước 3b: Bác bỏ H_0	$\lvert z\rvert > z_\alpha$
z_α thoả $\varphi(z_\alpha) = (1-\alpha)/2$ tra từ Bảng giá trị hàm Laplace	

2) Kiểm định một phía: Xét đám đông X có tỉ lệ $p = P(A)$ chưa biết. Với mỗi số α $(0 < \alpha < 1)$ khá bé, dựa vào mẫu $(X_1, X_2,..., X_n)$ ta có qui tắc kiểm định giả thiết một phía về tỉ lệ $p = P(A)$ với mức ý nghĩa α như sau:

BẢNG 6B

KIỂM ĐỊNH GIẢ THIẾT VỀ TỈ LỆ p = P(A) **H_0: p = p_0 với giả thiết đối H_1: p > p_0 (mức ý nghĩa α)**	
Bước 1: Tính z	$z = \dfrac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$
Bước 2: Tra Bảng	$z_{2\alpha}$
Bước 3a: Chấp nhận H_0	$z \le z_{2\alpha}$
Bước 3b: Bác bỏ H_0	$z > z_{2\alpha}$
$z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1-2\alpha)/2$ tra từ Bảng giá trị hàm Laplace	

BẢNG 6C

KIỂM ĐỊNH GIẢ THIẾT VỀ TỈ LỆ p = P(A) **H_0: p = p_0 với giả thiết đối H_1: p < p_0 (mức ý nghĩa α)**	
Bước 1: Tính z	$z = \dfrac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$
Bước 2: Tra Bảng	$z_{2\alpha}$
Bước 3a: Chấp nhận H_0	$-z \le z_{2\alpha}$
Bước 3b: Bác bỏ H_0	$-z > z_{2\alpha}$
$z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1-2\alpha)/2$ tra từ Bảng giá trị hàm Laplace	

Ví dụ. Để khảo sát chỉ tiêu X của một loại sản phẩm, người ta quan sát một mẫu và có kết quả sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	8	9	20	16	16	13	18

Những sản phẩm có chỉ tiêu X từ 27cm trở lên được xếp vào loại A.

a) Một tài liệu cũ cho rằng tỉ lệ sản phẩm loại A là 60%. Hãy nhận định về tài liệu cũ với mức ý nghĩa 1%.

b) Tỉ lệ sản phẩm loại A trước đây là 40%. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Với mức ý nghĩa 3%, có thể nói rằng kỹ thuật mới làm tăng tỉ lệ sản phẩm loại A hay không?

Giải. Ta tính được:

- Cỡ mẫu $n = 100$.
- Tỉ lệ mẫu các sản phẩm loại A là $F_n = 47/100 = 0,47$.

a) Đây là bài toán kiểm định giả thiết về tỉ lệ p các sản phẩm loại A với mức ý nghĩa $\alpha = 1\% = 0,01$:

$$H_0: p = 60\% = 0,6 \text{ với giả thiết đối } H_1: p \neq 0,6$$

Ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0 q_0}} = \frac{(0,47 - 0,6)\sqrt{100}}{\sqrt{0,6(1-0,6)}} = -2,6536.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1-\alpha)/2 = 0,99/2 = 0,495$$

ta được $z_\alpha = 2,58$.

Bước 3: Kiểm định. Vì $|z| = 2,6536 > 2,58 = z_\alpha$ nên ta bác bỏ giả thiết H_0: p = 0,6, nghĩa là chấp nhận H_1: p ≠ 0,6.

Kết luận: Với mức ý nghĩa 1%, tài liệu thống kê cũ dã lạc hậu, không còn phù hợp với thực tế.

b) Đây là bài toán kiểm định giả thiết về tỉ lệ p các sản phẩm loại A với mức ý nghĩa α = 3% = 0,03:

H_0: p = 40% = 0,4 với giả thiết đối H_1: p > 0,4

Ta kiểm định như sau:

Bước 1: Ta có

$$z = \frac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0q_0}} = \frac{(0,47-0,4)\sqrt{100}}{\sqrt{0,4(1-0,4)}} = 1,4289.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1-2\alpha)/2 = 0,94/2 = 0,47$$

ta được $z_{2\alpha} = 1,88$.

Bước 3: Kiểm định. Vì $z = 1,4289 < 1,88 = z_{2\alpha}$ nên ta chấp nhận giả thiết H_0: p = 0,6.

Kết luận: Với mức ý nghĩa 3%, kỹ thuật mới không làm tăng tỉ lệ sản phẩm loại A.

3.3. Kiểm định giả thiết về phương sai

1) Kiểm định hai phía: Xét đám đông X có phân phối chuẩn với phương sai $\sigma^2 = D(X)$ chưa biết. Với mỗi số α (0 < α < 1) khá bé, dựa vào mẫu $(X_1, X_2,..., X_n)$ ta có qui tắc kiểm định giả thiết hai phía về phương sai $\sigma^2 = D(X)$ với mức ý nghĩa α như sau:

BẢNG 7A

KIỂM ĐỊNH GIẢ THIẾT VỀ PHƯƠNG SAI $\sigma^2 = D(X)$ **H_0: $\sigma^2 = \sigma_0^2$ với giả thiết đối H_1: $\sigma^2 \neq \sigma_0^2$ (mức ý nghĩa α)**	
Bước 1: Tính z	$z = \dfrac{(n-1)S^2}{\sigma_0^2}$
Bước 2: Tra Bảng	$\chi^2_{\frac{\alpha}{2}}$ và $\chi^2_{1-\frac{\alpha}{2}}$
Bước 3a: Chấp nhận H_0	$\chi^2_{1-\frac{\alpha}{2}} \le z \le \chi^2_{\frac{\alpha}{2}}$
Bước 3b: Bác bỏ H_0	$z < \chi^2_{1-\frac{\alpha}{2}}$ hoặc $z > \chi^2_{\frac{\alpha}{2}}$
$\chi^2_{\frac{\alpha}{2}}$ và $\chi^2_{1-\frac{\alpha}{2}}$ tra từ Bảng Phân phối Chi bình phương χ^2 với n–1 bậc tự do	

2) Kiểm định một phía: Xét đám đông X có phân phối chuẩn với phương sai $\sigma^2 = D(X)$ chưa biết. Với mỗi số α (0 < α < 1) khá bé, dựa vào mẫu $(X_1, X_2,..., X_n)$ ta có qui tắc kiểm định giả thiết một phía về phương sai $\sigma^2 = D(X)$ với mức ý nghĩa α như sau:

BẢNG 7B

KIỂM ĐỊNH GIẢ THIẾT VỀ PHƯƠNG SAI $\sigma^2 = D(X)$ **H_0: $\sigma^2 = \sigma_0^2$ với giả thiết đối H_1: $\sigma^2 > \sigma_0^2$ (mức ý nghĩa α)**	
Bước 1: Tính z	$z = \dfrac{(n-1)S^2}{\sigma_0^2}$
Bước 2: Tra Bảng	χ^2_α
Bước 3a: Chấp nhận H_0	$z \le \chi^2_\alpha$
Bước 3b: Bác bỏ H_0	$z > \chi^2_\alpha$
χ^2_α tra từ Bảng Phân phối Chi bình phương χ^2 với n–1 bậc tự do	

BẢNG 7C

KIỂM ĐỊNH GIẢ THIẾT VỀ PHƯƠNG SAI $\sigma^2 = D(X)$ **H_0: $\sigma^2 = \sigma_0^2$ với giả thiết đối H_1: $\sigma^2 < \sigma_0^2$ (mức ý nghĩa α)**	
Bước 1: Tính z	$z = \dfrac{(n-1)S^2}{\sigma_0^2}$
Bước 2: Tra Bảng	$\chi^2_{1-\alpha}$
Bước 3a: Chấp nhận H_0	$z \ge \chi^2_{1-\alpha}$
Bước 3b: Bác bỏ H_0	$z < \chi^2_{1-\alpha}$
$\chi^2_{1-\alpha}$ tra từ Bảng Phân phối Chi bình phương χ^2 với n–1 bậc tự do	

Ví dụ. Đường kính của một chi tiết máy là đại lượng ngẫu nhiên X có phân phối chuẩn. Người ta đo thử 28 chi tiết máy do một máy sản xuất và tìm được phương sai mẫu hiệu chỉnh là $S^2 = (2,0853)^2$ (cm²).

a) Khi máy hoạt động bình thường thì độ lệch chuẩn của X của các chi tiết máy do máy sản xuất là 1,8cm. Với mức ý nghĩa 1%, hãy xét xem máy có hoạt động bình thường không.

b) Theo qui định mới, nếu độ lệch chuẩn của X lớn hơn 1,6cm thì phải điều chỉnh lại máy. Với mức ý nghĩa 5%, có phải điều chỉnh lại máy không?

Giải. Ta có:

- Cỡ mẫu n = 28.
- Phương sai mẫu hiệu chỉnh của X: $S^2 = (2,0853)^2$ (cm²).

a) Đây là bài toán kiểm định giả thiết về phương sai $\sigma^2 = D(X)$ với mức ý nghĩa α = 1% = 0,01:

H_0: $\sigma^2 = (1,8)^2$ với giả thiết đối H_1: $\sigma^2 \neq (1,8)^2$

Bước 1: Ta có:

$$z = \frac{(n-1)S^2}{\sigma_0^2} = \frac{27.(2,0853)^2}{(1,8)^2} = 36,2373$$

Bước 2: Tra bảng Phân phối Chi bình phương χ^2 với k= n – 1 = 27 bậc tự do, ta tìm được $\chi^2_{\frac{\alpha}{2}} = \chi^2_{0,005} = 49,65$ và $\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0,995} = 11,80765$.

Bước 3: Kiểm định. Vì $\chi^2_{1-\frac{\alpha}{2}} = 11,80765 \le z = 36,2373 \le 49,65 = \chi^2_{\frac{\alpha}{2}}$ nên ta chấp nhận giả thiết H_0: $\sigma^2 = (1,8)^2$.

Kết luận: Với mức ý nghĩa 1%, máy hoạt động bình thường.

b) Đây là bài toán kiểm định giả thiết về phương sai $\sigma^2 = D(X)$ với mức ý nghĩa $\alpha = 5\% = 0,05$:

H_0: $\sigma^2 = (1,6)^2$ với giả thiết đối H_1: $\sigma^2 > (1,6)^2$

Bước 1: Ta có:

$$z = \frac{(n-1)S^2}{\sigma_0^2} = \frac{27.(2,0853)^2}{(1,6)^2} = 45,8628$$

Bước 2: Tra bảng phân phối Chi bình phương χ^2 với k = n – 1 = 27 bậc tự do, ta tìm được $\chi^2_{\alpha} = \chi^2_{0,05} = 40,11$.

Bước 3: Kiểm định. Vì z = 45,8628 > 40,11 = χ^2_{α} nên ta bác bỏ giả thiết H_0: $\sigma^2 = (1,6)^2$, nghĩa là chấp nhận H_1: $\sigma^2 > (1,6)^2$.

Kết luận: Với mức ý nghĩa 5%, phải điều chỉnh lại máy.

3.4. Kiểm định giả thiết về so sánh hai kỳ vọng

1) Kiểm định hai phía: Xét hai đám đông X, Y với các kỳ vọng $\mu_X = M(X)$ và $\mu_Y = M(Y)$ đều chưa biết. Với mỗi số α ($0 < \alpha < 1$) khá bé, dựa vào các mẫu $(X_1, X_2,..., X_{n_X})$ và $(Y_1, Y_2,..., Y_{n_Y})$ ta có qui tắc kiểm định giả thiết hai phía về so sánh hai kỳ vọng như sau:

BẢNG 8A

KIỂM ĐỊNH GIẢ THIẾT VỀ SO SÁNH HAI KỲ VỌNG
H_0: $\mu_X = \mu_Y$ với giả thiết đối H_1: $\mu_X \ne \mu_Y$ (mức ý nghĩa α)

Trường hợp / Bước	$n_X \ge 30$ và $n_Y \ge 30$	$n_X < 30$ hoặc $n_Y < 30$
1) Tính z	$z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}$	$z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}$
2) Tra Bảng	z_{α}	t^k_{α}
3a) Chấp nhận H_0	$\lvert z\rvert \le z_{\alpha}$	$\lvert z\rvert \le t^k_{\alpha}$
3b) Bác bỏ H_0	$\lvert z\rvert > z_{\alpha}$	$\lvert z\rvert > t^k_{\alpha}$

- z_{α} thoả $\varphi(z_{\alpha}) = (1 - \alpha)/2$ tra từ Bảng giá trị hàm Laplace
- t^k_{α} với $k = n_X + n_Y - 2$ tra từ Bảng Phân phối Student

2) Kiểm định một phía: Xét hai đám đông X, Y với các kỳ vọng $\mu_X = M(X)$ và $\mu_Y = M(Y)$ đều chưa biết. Với mỗi số α ($0 < \alpha < 1$) khá bé, dựa vào

các mẫu $(X_1, X_2,..., X_{n_X})$ và $(Y_1, Y_2,..., Y_{n_Y})$ ta có qui tắc kiểm định giả thiết một phía về so sánh hai kỳ vọng như sau:

BẢNG 8B

KIỂM ĐỊNH GIẢ THIẾT VỀ SO SÁNH HAI KỲ VỌNG
H_0: $\mu_X = \mu_Y$ với giả thiết đối H_1: $\mu_X > \mu_Y$ (mức ý nghĩa α)

Trường hợp / Bước	$n_X \ge 30$ và $n_Y \ge 30$	$n_X < 30$ hoặc $n_Y < 30$
1) Tính z	$z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}$	$z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}$
2) Tra Bảng	$z_{2\alpha}$	$t^k_{2\alpha}$
3a) Chấp nhận H_0	$z \le z_{2\alpha}$	$z \le t^k_{2\alpha}$
3b) Bác bỏ H_0	$z > z_{2\alpha}$	$z > t^k_{2\alpha}$

- $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ tra từ Bảng giá trị hàm Laplace
- $t^k_{2\alpha}$ với k = n – 1 tra từ Bảng Phân phối Student

BẢNG 8C

KIỂM ĐỊNH GIẢ THIẾT VỀ SO SÁNH HAI KỲ VỌNG
H_0: $\mu_X = \mu_Y$ với giả thiết đối H_1: $\mu_X < \mu_Y$ (mức ý nghĩa α)

Trường hợp / Bước	$n_X \ge 30$ và $n_Y \ge 30$	$n_X < 30$ hoặc $n_Y < 30$
1) Tính z	$z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}$	$z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}$
2) Tra Bảng	$z_{2\alpha}$	$t^k_{2\alpha}$
3a) Chấp nhận H_0	$-z \le z_{2\alpha}$	$-z \le t^k_{2\alpha}$
3b) Bác bỏ H_0	$-z > z_{2\alpha}$	$-z > t^k_{2\alpha}$

- $z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ tra từ Bảng giá trị hàm Laplace
- $t^k_{2\alpha}$ với k = n –1 tra từ Bảng Phân phối Student

Ví dụ. Theo dõi giá cổ phiếu của hai công ty A và B trong một số ngày, người ta tính được các số liệu sau:

	Kỳ vọng mẫu	Độ lệch mẫu hiệu chỉnh
Công ty A	38,24	2,2
Công ty B	37,10	1,5

a) Cho biết số liệu trên có được từ 31 ngày theo dõi giá trị cổ phiếu (mỗi ngày một giá trị cho mỗi công ty). Vậy với mức ý nghĩa 1%, có thể nói rằng có sự khác biệt thực sự về giá cổ phiếu trung bình của hai công ty A và B hay không?

b) Cho biết số liệu trên có được từ 20 ngày theo dõi giá trị cổ phiếu (mỗi ngày một giá trị cho mỗi công ty).Với mức ý nghĩa 4%, có thể nói rằng giá cổ phiếu trung bình của công ty A thực sự cao hơn của công ty B hay không (Giả sử các giá trị cổ phiếu có phân phối chuẩn)?

Giải. a) Đây là bài toán kiểm định về so sánh hai kỳ vọng với mức ý nghĩa α = 1% = 0,01:

$$\mathbf{H_0}: \mu_A = \mu_B \text{ với giả thiết đối } \mathbf{H_1}: \mu_A \neq \mu_B$$

Vì $n_A = n_B = 31 > 30$ nên ta kiểm định như sau:

Bước 1: Ta có:

$$z = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{38,24 - 37,1}{\sqrt{\frac{(2,2)^2}{31} + \frac{(1,5)^2}{31}}} = 2,3838.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1 - \alpha)/2 = 0,99/2 = 0,495$$

ta được $z_\alpha = 2,58$.

Bước 3: Kiểm định.

Vì $|z| = 2,3838 < 2,58 = z_\alpha$ nên ta chấp nhận giả thiết $H_0: \mu_A = \mu_B$.

Kết luận: Với mức ý nghĩa 1%, giá trị cổ phiếu trung bình của hai công ty A và B có thể xem là như nhau, nghĩa là không có sự khác biệt thực sự về giá cổ phiếu trung bình của hai công ty này.

b) Đây là bài toán kiểm định về so sánh hai kỳ vọng với mức ý nghĩa α = 4% = 0,04:

$$\mathbf{H_0}: \mu_A = \mu_B \text{ với giả thiết đối } \mathbf{H_1}: \mu_A > \mu_B$$

Vì $n_A = n_B = 20 < 30$ và các giá trị cổ phiếu X_A, X_B đều có phân phối chuẩn nên ta kiểm định như sau:

Bước 1: Ta có:

$$z = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{38,24 - 37,1}{\sqrt{\frac{(2,2)^2}{20} + \frac{(1,5)^2}{20}}} = 1,9147.$$

Bước 2: Đặt $k = n_A + n_B - 2 = 38$. Tra bảng phân phối Student ứng với $k = 38$ và $2\alpha = 0,08$ ta được $t_{2\alpha}^k = 1,799$.

Bước 3: Kiểm định:

Vì $z = 1,9147 > 1,799 = t_{2\alpha}^k$ nên ta bác bỏ $H_0: \mu_A = \mu_B$, nghĩa là chấp nhận $\mu_A > \mu_B$.

Kết luận: Với mức ý nghĩa 4%, có thể xem giá trị cổ phiếu trung bình của công ty A thực sự cao hơn của công ty B.

3.5. Kiểm định giả thiết về so sánh hai tỉ lệ

1) Kiểm định hai phía: Xét hai đám đông X, Y trong đó X có tỉ lệ p_X; Y có tỉ lệ p_Y đều chưa biết. Với mỗi số α ($0 < \alpha < 1$) khá bé, dựa vào các mẫu $(X_1, X_2,..., X_{n_X})$ và $(Y_1, Y_2,..., Y_{n_Y})$ ta có qui tắc kiểm định giả thiết hai phía về so sánh hai tỉ lệ như sau:

BẢNG 9A

KIỂM ĐỊNH GIẢ THIẾT VỀ SO SÁNH HAI TỈ LỆ
H_0: $p_X = p_Y$ (= p_0)với giả thiết đối H_1: $p_X \neq p_Y$ (mức ý nghĩa α)

Trường hợp / Bước	p_0 đã biết	p_0 chưa biết
1) Tính z	$z = \dfrac{F_{n_X} - F_{n_Y}}{\sqrt{p_0(1-p_0)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$	$z = \dfrac{F_{n_X} - F_{n_Y}}{\sqrt{p_0'(1-p_0')\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$ với $p_0' = \dfrac{n_X F_{nX} + n_Y F_{nY}}{n_X + n_Y}$
2) Tra Bảng	z_α	z_α
3a) Chấp nhận H_0	$\lvert z\rvert \le z_\alpha$	$\lvert z\rvert \le z_\alpha$
3b) Bác bỏ H_0	$\lvert z\rvert > z_\alpha$	$\lvert z\rvert > z_\alpha$
z_α thoả $\varphi(z_\alpha) = (1 - \alpha)/2$ tra từ Bảng giá trị hàm Laplace		

2) Kiểm định một phía: Xét hai đám đông X, Y trong đó X có tỉ lệ p_X; Y có tỉ lệ p_Y đều chưa biết. Với mỗi số α ($0 < \alpha < 1$) khá bé, dựa vào các mẫu $(X_1, X_2,..., X_{n_X})$ và $(Y_1, Y_2,..., Y_{n_Y})$ ta có qui tắc kiểm định giả thiết một phía về so sánh hai tỉ lệ như sau:

BẢNG 9B

KIỂM ĐỊNH GIẢ THIẾT VỀ SO SÁNH HAI TỈ LỆ
H_0: $p_X = p_Y$ (= p_0)với giả thiết đối H_1: $p_X > p_Y$ (mức ý nghĩa α)

Trường hợp / Bước	p_0 đã biết	p_0 chưa biết
1) Tính z	$z = \dfrac{F_{n_X} - F_{n_Y}}{\sqrt{p_0(1-p_0)\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$	$z = \dfrac{F_{n_X} - F_{n_Y}}{\sqrt{p_0'(1-p_0')\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$ với $p_0' = \dfrac{n_X F_{nX} + n_Y F_{nY}}{n_X + n_Y}$
2) Tra Bảng tìm	$z_{2\alpha}$	$z_{2\alpha}$
3a) Chấp nhận H_0	$z \le z_{2\alpha}$	$z \le z_{2\alpha}$
3b) Bác bỏ H_0	$z > z_{2\alpha}$	$z > z_{2\alpha}$
$z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1 - 2\alpha)/2$ tra từ Bảng giá trị hàm Laplace		

BẢNG 9C

KIỂM ĐỊNH GIẢ THIẾT VỀ SO SÁNH HAI TỈ LỆ

H_0: $p_X = p_Y$ ($= p_0$) với giả thiết đối H_1: $p_X < p_Y$ (mức ý nghĩa α)

Trường hợp / Bước	p_0 đã biết	p_0 chưa biết
1) Tính z	$z = \dfrac{F_{n_X} - F_{n_Y}}{\sqrt{p_0(1-p_0)\left(\dfrac{1}{n_X}+\dfrac{1}{n_Y}\right)}}$	$z = \dfrac{F_{n_X} - F_{n_Y}}{\sqrt{p'_0(1-p'_0)\left(\dfrac{1}{n_X}+\dfrac{1}{n_Y}\right)}}$ với $p'_0 = \dfrac{n_X F_{nX} + n_Y F_{nY}}{n_X + n_Y}$
2) Tra Bảng	$z_{2\alpha}$	$z_{2\alpha}$
3a) Chấp nhận H_0	$-z \le z_{2\alpha}$	$-z \le z_{2\alpha}$
3b) Bác bỏ H_0	$-z > z_{2\alpha}$	$-z > z_{2\alpha}$
$z_{2\alpha}$ thoả $\varphi(z_{2\alpha}) = (1-2\alpha)/2$ tra từ Bảng giá trị hàm Laplace		

Ví dụ. Khảo sát một số sản phẩm cùng loại ở hai kho I và II, ta thu được các số liệu sau:

	Số sản phẩm	Số phế phẩm
Kho I	100	4
Kho II	200	24

a) Với mức ý nghĩa 5%, có thể nói rằng chất lượng hàng ở hai kho là như nhau hay không?

b) Với mức ý nghĩa 1%, có thể nói rằng chất lượng hàng ở kho I tốt hơn kho II không?

Giải. Từ các giả thiết của bài toán ta suy ra:

- Đối với kho I: Cỡ mẫu $n_1 = 100$; tỉ lệ mẫu phế phẩm $F_{n1} = 0{,}04$.
- Đối với kho II: Cỡ mẫu $n_2 = 200$; tỉ lệ mẫu phế phẩm $F_{n2} = 0{,}12$.
- $p'_0 = \dfrac{n_1F_{n1} + n_2F_{n2}}{n_1 + n_2} = \dfrac{100.0{,}4 + 200.0{,}12}{100 + 200} = \dfrac{7}{75}$.

a) Đây là bài toán kiểm định giả thiết về so sánh hai tỉ lệ với mức ý nghĩa $\alpha = 5\% = 0{,}05$:

$$\mathbf{H_0}: p_1 = p_2 \text{ với giả thiết đối } \mathbf{H_1}: p_1 \neq p_2$$

Ta kiểm định như sau:

Bước 1: Ta có:

$$z = \frac{F_{n1} - F_{n2}}{\sqrt{p'_0(1-p'_0)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} = \frac{0{,}04 - 0{,}12}{\sqrt{\dfrac{7}{75}\left(1-\dfrac{7}{75}\right)\left(\dfrac{1}{100}+\dfrac{1}{200}\right)}} = -2{,}2454.$$

Bước 2: Tra bảng giá trị hàm Laplace để tìm z_α thoả

$$\varphi(z_\alpha) = (1-\alpha)/2 = 0{,}95/2 = 0{,}475$$

ta được $z_\alpha = 1{,}96$.

Bước 3: Kiểm định:

Vì $|z| = 2{,}2454 > 1{,}96 = z_\alpha$ nên ta bác bỏ giả thiết H_0: $p_1 = p_2$, nghĩa là chấp nhận H_1: $p_1 \neq p_2$.

Kết luận: Với mức ý nghĩa 5%, chất lượng hàng ở hai kho không như nhau.

b) Đây là bài toán kiểm định giả thiết về so sánh hai tỉ lệ với mức ý nghĩa $\alpha = 1\% = 0{,}01$:

$$\mathbf{H_0}: p_1 = p_2 \text{ với giả thiết đối } \mathbf{H_1}: p_1 < p_2$$

Ta kiểm định như sau:

Bước 1: Tính z như trong Bước 1 ở câu a) ta được $z = -2{,}2454$.

Bước 2: Tra bảng giá trị hàm Laplace để tìm $z_{2\alpha}$ thoả

$$\varphi(z_{2\alpha}) = (1-2\alpha)/2 = 0{,}98/2 = 0{,}49$$

ta được $z_{2\alpha} = 2{,}33$.

Bước 3: Kiểm định:

Vì $-z = 2{,}2454 < 2{,}33 = z_{2\alpha}$ nên ta chấp nhận giả thiết H_0: $p_1 = p_2$.

Kết luận: Với mức ý nghĩa 1%, chưa thể nói rằng chất lượng hàng ở kho I tốt hơn kho II.

3.6. Kiểm định giả thiết về phân phối

1) Bài toán. Xét đám đông X chưa biết luật phân phối. Vơi mỗi số α ($0 < \alpha < 1$) khá bé, hãy dựa vào một mẫu thu được của X để kiểm định giả thiết:

H_0: X có phân phối theo qui luật đã cho

với giả thiết đối:

H_1: X không có phân phối theo qui luật đã cho

với mức ý nghĩa α.

2) Qui tắc kiểm định: Giả sử mẫu thu được gồm k nhóm có dạng:

X_i	x_0-x_1	x_1-x_2	....	x_{i-1}-x_i	....	x_{k-1}-x_k
n_i	n_1	n_2	....	n_i	....	n_k

trong đó các giá trị n_i (ngoại trừ n_1 và n_k ứng với các khoảng đầu và cuối) không quá bé ($n_i \ge 5$).

Đối với trường hợp rời rạc, ta thay khoảng x_{i-1}-x_i bởi $x'_i = \dfrac{x_{i-1} + x_i}{2}$, hơn nữa, khi X có thể lấy vô hạn giá trị, ta còn phải thay khoảng cuối x_{k-1}-x_k bằng $(x_{k-1}, +\infty)$ (hoặc khoảng đầu x_0-x_1 bằng $(-\infty, x_1)$, nếu cần). Dựa vào phân phối đã cho trong H_0 để tính các xác suất $p_i = P(X = x_i')$.

Đối với trường hợp X liên tục, ta thay khoảng đầu x_0-x_1 bằng $(-\infty, x_1)$; thay khoảng cuối x_{k-1}-x_k bằng $(x_{k-1}, +\infty)$ và dựa vào phân phối đã cho trong H_0 để tính các xác suất $p_i = P(x_{i-1} \le X \le x_i)$.

Chú ý. Khi tính các p_i, nếu chưa biết tham số nào của phân phối đã cho thì ta thay bằng ước lượng không chệch từ mẫu đang xét.

Ta có qui tắc kiểm định như sau:

BẢNG 10

KIỂM ĐỊNH GIẢ THIẾT VỀ PHÂN PHỐI **H_0: X có phân phối theo qui luật đã cho (mức ý nghĩa α)**	
Bước 1: Tính χ^2	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$
Bước 2: Tra Bảng	χ^2_α
Bước 3a: Chấp nhận H_0	$\chi^2 \le \chi^2_\alpha$
Bước 3b: Bác bỏ H_0	$\chi^2 > \chi^2_\alpha$
χ^2_α tra từ Bảng Phân phối Chi bình phương χ^2 với k – r –1 bậc tự do, trong đó r là số tham số chưa biết của phân phối.	

Ví dụ 1. Điều tra 160 gia đình 4 con ở một vùng dân cư người ta thu được bảng số liệu sau:

Số con gái	0	1	2	3	4
Số gia đình	16	48	62	30	4

Với mức ý nghĩa 5%, có thể cho rằng số con gái trong một gia đình 4 con có phân phối nhị thức hay không?

Giải. Gọi X là số con gái trong một gia đình 4 con. Bài toán yêu cầu kiểm định giả thiết sau với mức ý nghĩa α = 5% = 0,05:

H_0: X có phân phối nhị thức X ~ B(4,p) với p chưa biết

với giả thiết đối:

H_1 : X không có phân phối nhị thức như trên.

Trước hết ta thay p bằng tỉ lệ mẫu số con gái trong một gia đình:

$$p \approx F_n = \frac{1.48 + 2.62 + 3.30 + 4.4}{160.4} = 0,4344.$$

Ta tính các $p_i = P(X = i)$ theo công thức Bernoulli:

$$p_i = C_4^i (0,4344)^i (0,5656)^{4-i}$$

Cụ thể ta tính được:

$p_0 = 0,1023$; $p_1 = 0,3144$; $p_2 = 0,3622$; $p_3 = 0,1855$; $p_4 = 0,0356$.

Ta lập bảng:

X_i	n_i	p_i	np_i	$(n_i-np_i)^2/np_i$
0	16	0,1023	16,368	0,0083
1	48	0,3144	50,304	0,1055
2	62	0,3622	57,952	0,2828
3	30	0,1855	29,68	0,0035
4	4	0,0356	5,696	0,5050
Tổng	n = 160			$\chi^2 = 0,9051$

Bước 1: Ta có $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = 0,9051.$

Bước 2: Số tham số chưa biết là r = 1 (do p chưa biết). Ta có k – r – 1 = 5 – 1 – 1 = 3. Tra bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(3)$ với 3 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,05} = 7,815$.

Bước 3: Kiểm định:

Vì $\chi^2 = 0,9051 < 7,815 = \chi^2_\alpha$ nên ta chấp nhận giả thiết H_0.

Kết luận: Với mức ý nghĩa 5%, có thể cho rằng số con gái trong một gia đình 4 con là X có phân phối nhị thức: X ~ B(4, 0,4344).

Ví dụ 2. Quan sát một số người đến một trung tâm bưu điện trong 110 khoảng (mỗi khoảng 5 phút) ta thu được kết quả sau:

Số người	0	1	2	3	4	5
Số khoảng	19	34	19	15	12	11

Gọi X là số người đến trung tâm này trong một khoảng thời gian 5 phút. Với mức ý nghĩa 3%, có thể cho rằng X có phân phối Poisson hay không?

Giải. Bài toán yêu cầu kiểm định giả thiết sau với mức ý nghĩa α = 3% = 0,03:

H_0: X có phân phối Poisson X ~ P(a) (a chưa biết)

với giả thiết đối:

H_1 : X không có phân phối Poisson.

Trước hết ta thay a bằng kỳ vọng mẫu

$$a \approx \bar{X} = \frac{1}{n}\sum X_i n_i = 2$$

Ta tính các $p_i = P(X = i)$ theo công thức:

$$p_i = \frac{e^{-2} 2^i}{i!}$$

và lập bảng:

X_i	n_i	p_i	np_i	$(n_i-np_i)^2/np_i$
0	19	0,135335	14,8869	1,136408
1	34	0,270671	29,7738	0,599882
2	19	0,270671	29,7738	3,898554
3	15	0,180447	19,8492	1,184669
4	12	0,090224	9,92464	0,434982
(5;+∞)	11	0,052652	5,79172	4,683614
Tổng	n = 110			χ^2 =11,9381

Bước 1: Ta có $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = 11,9381.$

Bước 2: Số tham số chưa biết là r = 1 (do a chưa biết). Ta có k – r – 1 = 6 – 1 – 1 = 4. Tra bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(4)$ với 4 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,03} = 10,7119$.

Bước 3: Kiểm định:

Vì $\chi^2 = 11,9381 > 10,7119 = \chi^2_\alpha$ nên ta bác bỏ giả thiết H_0.

Kết luận: Với mức ý nghĩa 3%, X không có phân phối Poisson.

Ví dụ 3. Khảo sát chỉ tiêu X của một loại sản phẩm ta thu được kết quả sau:

X_i	20-22	22-24	24-26	26-28	28-30
Số sản phẩm	7	14	33	27	19

Kiểm định giả thiết X có phân phối chuẩn với mức ý nghĩa 2%.

Giải. Bài toán yêu cầu kiểm định giả thiết với mức ý nghĩa $\alpha = 2\% = 0{,}02$:

H_0: X có phân phối chuẩn $X \sim N(\mu,\sigma^2)$ (μ, σ^2 chưa biết)

với giả thiết đối:

H_1 : X không có phân phối chuẩn.

Trước hết xấp xỉ:

$$\mu \approx \overline{X} = \frac{1}{n}\sum X_i n_i = 25{,}74;$$

$$\sigma^2 \approx S^2 = \frac{1}{n}\sum X_i^2 n_i - (\overline{X})^2 = (2{,}3034)^2.$$

Ta tính các $p_i = P(x_{i-1} \le X \le x_i)$ theo công thức:

$$p_i = \varphi(\frac{x_i-\mu}{\sigma}) - \varphi(\frac{x_{i-1}-\mu}{\sigma}) = \varphi(\frac{x_i-25{,}74}{2{,}3034}) - \varphi(\frac{x_{i-1}-25{,}74}{2{,}3034})$$

trong đó φ là hàm Laplace, và lập bảng:

X_i	n_i	p_i	np_i	$(n_i-np_i)^2/np_i$
$(-\infty, 22)$	7	0,0516	5,16	0,6561
22-24	14	0,1720	17,20	0,5953
24-26	33	0,3203	32,03	0,0294
26-28	27	0,2927	29,27	0,1760
$(28,+\infty)$	19	0,1634	16,34	0,4330
Tổng	n = 100			χ^2 =1,8898

Bước 1: Ta có $\chi^2 = \sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i} = 1{,}8898$.

Bước 2: Số tham số chưa biết là r = 2 (do μ, σ^2 chưa biết). Ta có k – r – 1 = 5 – 2 – 1 = 2. Tra bảng phân phối Chi bình phương $\chi^2 \sim \chi^2(2)$ với 2 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,02} = 7{,}824$.

Bước 3: Kiểm định:

Vì $\chi^2 = 1{,}8898 < 7{,}824 = \chi^2_\alpha$ nên ta chấp nhận giả thiết H_0.

Kết luận: Với mức ý nghĩa 2%, X có phân phối chuẩn: $X \sim N(\mu,\sigma^2)$ với $\mu = 25{,}74$; $\sigma^2 = (2{,}3034)^2$.

3.7. Kiểm định giả thiết về tính độc lập

1) Bài toán. Từ hai đám đông X và Y ta tiến hành quan sát và được kết quả trong bảng sau:

Y \ X	y_1	...	y_j	...	y_k	$\mathbf{m_X}$
x_1	n_{11}	...	n_{1j}		n_{1k}	$\mathbf{m_1}$
...	...	...	...	...	...	**...**
x_i	n_{i1}	...	n_{ij}		n_{ik}	$\mathbf{m_i}$
...	...	...	...	...	...	**...**
x_h	n_{h1}	...	n_{hj}	...	n_{hk}	$\mathbf{m_h}$
$\mathbf{n_Y}$	$\mathbf{n_1}$	**...**	$\mathbf{n_j}$	**...**	$\mathbf{n_k}$	**n**

trong đó

- n_{ij} là số lần $(X,Y) = (x_i,y_j)$ với $1 \le i \le h$; $1 \le j \le k$;
- $m_i = \sum_{j=1}^{k} n_{ij}$ là số lần $X = x_i$ với $1 \le i \le h$;
- $n_j = \sum_{i=1}^{h} n_{ij}$ là số lần $Y = y_j$ với $1 \le j \le k$;
- $n = \sum_{i=1}^{h}\sum_{j=1}^{k} n_{ij}$ là cỡ mẫu (X,Y).

Với mỗi số α ($0 < \alpha < 1$) khá bé, hãy dựa vào mẫu trên để kiểm định giả thiết: **H_0: X và Y độc lập**

với giả thiết đối **H_1: X và Y không độc lập**

với mức ý nghĩa α.

2) Qui tắc kiểm định:

BẢNG 11

KIỂM ĐỊNH GIẢ THIẾT VỀ TÍNH ĐỘC LẬP **H_0: X và Y độc lập (mức ý nghĩa α)**	
Bước 1: Tính χ^2	$\chi^2 = n\left(\sum_{i=1}^{h}\sum_{j=1}^{k}\alpha_{ij} - 1\right)$ với $\alpha_{ij} = \frac{(n_{ij})^2}{m_i n_j}$
Bước 2: Tra Bảng	χ^2_α
Bước 3a: Chấp nhận H_0	$\chi^2 \le \chi^2_\alpha$
Bước 3b: Bác bỏ H_0	$\chi^2 > \chi^2_\alpha$
χ^2_α tra từ Bảng Phân phối Chi bình phương χ^2 với (h–1)(k–1) bậc tự do	

Ví dụ. Một công ty điều tra sở thích của khách hàng về 3 loại mẫu khác nhau của cùng một mặt hàng. Kết quả thu được như sau:

Mẫu hàng \ Ý kiến	A	B	C
Thích	43	30	42
Không thích	35	53	39
Không có ý kiến	22	17	19

Hỏi đối với mặt hàng trên, có sự phân biệt về sở thích của khách hàng đối với 3 loại mẫu hàng A, B, C hay không với mức ý nghĩa 3%?

Giải. Gọi

- X là ý kiến của khách hàng;
- Y là mẫu hàng.

Bài toán yêu cầu kiểm định giả thiết sau với mức ý nghĩa $\alpha = 3\% = 0,03$:

H₀: X độc lập với Y

với giả thiết đối: **H₁**: X không độc lập với Y

Ta lập bảng:

Y / X	A	B	C	Tổng
Thích	43 $\alpha_{11} = 0,160783$	30 $\alpha_{12} = 0,078261$	42 $\alpha_{13} = 0,153391$	115
Không thích	35 $\alpha_{21} = 0,096457$	53 $\alpha_{22} = 0,221181$	39 $\alpha_{23} = 0,119764$	127
Không ý kiến	22 $\alpha_{31} = 0,083448$	17 $\alpha_{32} = 0,049828$	19 $\alpha_{33} = 0,062241$	58
Tổng	100	100	100	n=300

trong đó α_{ij} được tính theo công thức: $\alpha_{ij} = \dfrac{(n_{ij})^2}{m_i n_j}$. Cụ thể:

$\alpha_{11} = \dfrac{43^2}{115 \times 100} = 0,160783$,... (kết quả được ghi chi tiết trong bảng).

Bước 1: Ta có $\chi^2 = n\left(\sum\sum \alpha_{ij} - 1\right) = 7,6062.$

Bước 2: Ta có (h-1)(k-1) = 4 (do h = k = 3). Tra bảng phân phối chi bình phương $\chi^2 \sim \chi^2(4)$ với 4 bậc tự do, ta được: $\chi^2_\alpha = \chi^2_{0,03} = 10,7119.$

Bước 3: Kiểm định:

Vì $\chi^2 = 7,6062 < 10,7119 = \chi^2_\alpha$ nên ta chấp nhận giả thiết H_0.

Kết luận: Với mức ý nghĩa 3%, không có sự phân biệt về sở thích của khách hàng đối với các loại mẫu hàng.

BÀI TẬP

Bài 1. Để khảo sát chiều cao X của một giống cây trồng, người ta quan sát một mẫu và có kết qủa sau:

X(cm)	95-105	105-115	115-125	125-135	135-145	145-155	155-165
Số cây	10	10	15	30	10	10	15

a) Ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 96%. Với độ tin cậy đó, chiều cao trung bình tối đa của giống cây trồng trên là bao nhiêu? Tối thiểu là bao nhiêu?

b) Nếu muốn ước lượng chiều cao trung bình của giống cây trồng trên với độ tin cậy 99% và độ chính xác 4 cm thì cần phải điều tra thêm bao nhiêu cây nữa?

c) Nếu ước lượng chiều cao trung bình của giống cây trồng trên với độ chính xác 4,58cm thì sẽ đạt được độ tin cậy là bao nhiêu?

d) Một tài liệu thống kê cũ cho rằng chiều cao trung bình của giống cây trồng trên là 127cm. Hãy cho kết luận về tài liệu đó với mức ý nghĩa 1%.

e) Những cây trồng có chiều cao từ 135cm trở lên được gọi là những cây "cao". Hãy ước lượng tỉ lệ cây cao với độ tin cậy 95%. Với độ tin cậy đó, tỉ lệ cây cao đạt giá trị tối đa là bao nhiêu? Tối thiểu là bao nhiêu?

f) Nếu ước lượng tỉ lệ cây cao với độ chính xác 10% thì sẽ đạt được độ tin cậy là bao nhiêu?

g) Nếu ước lượng tỉ lệ cây cao với độ tin cậy 95% và độ chính xác 11% thì cần phải điều tra thêm bao nhiêu cây nữa?

h) Trước đây, tỉ lệ cây cao của loại cây trồng trên là 40%. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Hãy cho kết luận về kỹ thuật mới với mức ý nghĩa 5%.

i) Những cây trồng có chiều cao từ 105cm đến 125cm được gọi là những cây loại A. Hãy ước lượng chiều cao trung bình của những cây loại A với độ tin cậy 95% (GS X có phân phối chuẩn).

j) Bằng phương pháp mới, sau một thời gian người ta thấy chiều cao trung bình của những cây loại A là 119,5cm. Hãy cho kết luận về phương pháp mới với mức ý nghĩa 1% (GS X có phân phối chuẩn).

k) Giả sử X có phân phối chuẩn. Với độ tin cậy 95%, hãy ước lượng phương sai của X trong hai trường hợp :

α) Biết kỳ vọng của X là 130 cm.

β) Chưa biết kỳ vọng của X.

l) Khi canh tác bình thường thì phương sai của chiều cao X là 300cm². Hãy nhận định về tình hình canh tác với mức ý nghĩa 5% (GS X có phân phối chuẩn).

Bài 2. Để nghiên cứu nhu cầu của một loại hàng ở một khu vực, người ta khảo sát 400 hộ gia đình. Kết quả như sau:

Nhu cầu (kg/tháng/hộ)	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8
Số hộ	10	35	86	132	78	31	18	10

Cho biết trong khu vực có 4000 hộ.

a) Ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm với độ tin cậy 95%.

b) Khi ước lượng nhu cầu trung bình về mặt hàng này của toàn khu vực trong một năm, nếu ta muốn đạt được độ tin cậy 99% và độ chính xác là 4,8tấn thì cần khảo sát ở ít nhất bao nhiêu hộ gia đình?

Bài 3. Để khảo sát chỉ tiêu X của một loại sản phẩm của xí nghiệp I, người ta quan sát một mẫu trong kho và có kết qủa sau:

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sphẩm	8	9	20	16	16	13	18

a) Những sản phẩm có chỉ tiêu X từ 19cm trở xuống được gọi là những sản phẩm loại B. Hãy ước lượng tỉ lệ sản phẩm loại B với độ tin cậy 92%.

b) Giả sử trong kho có 1000 sản phẩm loại B. Hãy ước lượng số sản phẩm trong kho với độ tin cậy 92%.

c) Giả sử trong kho có 10.000 sản phẩm. Hãy ước lượng số sản phẩm loại B có trong kho với độ tin cậy 92%.

d) Giả sử trong kho để lẫn 1000 sản phẩm của xí nghiệp II và trong 100 sản phẩm lấy từ kho có 9 sản phẩm của xí nghiệp II. Hãy ước lượng số sản phẩm của xí nghiệp I có trong kho với độ tin cậy 82%.

Bài 4. Trái cây của một chủ hàng được đựng trong các sọt, mỗi sọt 100 trái. Người ta kiểm tra 50 sọt thì thấy có 450 trái không đạt tiêu chuẩn.

a) Ước lượng tỉ lệ trái không đạt tiêu chuẩn của lô hàng trên với độ tin cậy 95%.

b) Nếu muốn ước lượng tỉ lệ trái không đạt tiêu chuẩn với độ chính xác 0,5% thì sẽ đạt được độ tin cậy là bao nhiêu?

c) Nếu muốn ước lượng tỉ lệ trái không đạt tiêu chuẩn với độ chính xác 1% và độ tin cậy 99% thì phải điều tra thêm ít nhất bao nhiêu sọt nữa?

Bài 5. Để biết số lượng cá trong hồ lớn người ta bắt lên 2000 con đánh dấu xong rồi thả chúng xuống hồ. Sau đó người ta bắt lên 400 con và thấy có 80 con được đánh dấu.

a) Với độ tin cậy 95%, hãy ước lượng số cá có trong hồ.

b) Ước lượng số cá tối đa có trong hồ với độ tin cậy 96%.

c) Ước lượng số cá tối thiểu có trong hồ với độ tin cậy 94%.

Bài 6. Cho các số liệu như Bài 1.

a) Giả sử trung bình tiêu chuẩn của chiều cao X là 125cm. Có thể khẳng định rằng việc canh tác làm tăng chiều cao trung bình của giống cây trồng trên với mức ý nghĩa 1% hay không?

b) Giả sử trung bình tiêu chuẩn của chiều cao X là 134cm. Có thể khẳng định rằng việc canh tác làm giảm chiều cao trung bình của giống cây trồng trên với mức ý nghĩa 2% hay không?

c) Sau khi áp dụng phương pháp canh tác mơi, người ta thấy chiều cao trung bình của các cây loại A là 114cm. Hãy kết luận xem phương pháp mới có làm giảm chiều cao trung bình của các cây loại A hay không với mức ý nghĩa 3% (Giả sử X có phân phối chuẩn) .

d) Trước đây, chiều cao trung bình của các cây loại A là 120cm. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Hãy kết luận xem kỹ thuật mới có làm giảm chiều cao trung bình của các cây loại A hay không với mức ý nghĩa 2% (Giả sử X có phân phối chuẩn).

e) Sau khi áp dụng một phương pháp sản xuất, người ta thấy tỉ lệ các cây loại A là 35%. Hãy kết luận xem phương pháp mới có làm tăng tỉ lệ các cây loại A lên hay không với mức ý nghĩa 2%.

f) Theo tài liệu thống kê, tỉ lệ cây loại A là 20%. Hãy xét xem hiện nay việc canh tác có làm tăng tỉ lệ các cây loại A hay không với mức ý nghĩa 5%?

g) Theo tài liệu cũ, phương sai của chiều cao X là 250cm^2. Với mức ý nghĩa 5%, xét xem hiện tại chiều cao của cây trồng có biến động hơn so với trước đây hay không (GS X có phân phối chuẩn)?

h) Trước đây, phương sai của chiều cao X là 350cm^2. Các số liệu trên thu thập được sau khi đã áp dụng một kỹ thuật mới. Với mức ý nghĩa 5%, hãy xét xem kỹ thuật mới có làm chiều cao của giống cây trồng trên ít biến động hơn hay không (GS X có phân phối chuẩn)?

Bài 7. Để khảo sát đường kính của một chi tiết máy người ta kiểm tra một số sản phẩm của hai nhà máy. Trong kết quả sau đây, X là đường kính của chi tiết máy do nhà máy I sản xuất còn Y là đường kính của chi tiết máy do nhà máy II sản xuất. Những sản phẩm có chi tiết máy nhỏ hơn 19cm được xếp vào loại C.

X(cm)	11-15	15-19	19-23	23-27	27-31	31-35	35-39
Số sản phẩm	9	19	20	26	16	13	18

Y(cm)	13-16	16-19	19-22	22-25	25-28	28-31	31-34
Số sản phẩm	7	9	25	26	18	15	11

a) Có thể kết luận rằng đường kính trung bình của một chi tiết máy do hai nhà máy sản xuất bằng nhau hay không với mức ý nghĩa 1%?

b) Có thể cho rằng đường kính trung bình của một chi tiết máy do nhà máy I san xuất lớn hơn đường kính trung bình của một chi tiết máy do nhà máy II sản xuất hay không với mức ý nghĩa 5%?

c) Xét xem đường kính trung bình của một chi tiết máy do nhà máy II sản xuất có nhỏ hơn đường kính trung bình của một chi tiết máy do nhà máy I sản xuất hay không với mức ý nghĩa 2%?

d) Với mức ý nghĩa 4%, tỉ lệ sản phẩm loại C do hai nhà máy sản xuất có như nhau không?

e) Với mức ý nghĩa 3%, có thể cho rằng tỉ lệ sản phẩm loại C do nhà máy I sản xuất lớn hơn tỉ lệ sản phẩm loại C do nhà máy II sản xuất hay không?

f) Hãy nhận xét về ý kiến cho rằng tỉ lệ sản phẩm loại C do nhà máy II sản xuất nhỏ hơn tỉ lệ sản phẩm loại C do nhà máy thứ I sản xuất với mức ý nghĩa 5%?

Bài 8. Sản phẩm sản xuất ra trên một dây chuyền tự động được đóng gói theo qui cách 3 sản phẩm/hộp. Với mức ý nghĩa 1%, hãy xét xem số sản phẩm loại I có trong mỗi hộp có phải là ĐLNN có phân phối nhị thức hay không. Biết rằng khi kiểm tra 100 hộp người ta thấy có 75 hộp có 3 sản phẩm loại I, 20 hộp có 2 sản phẩm loại I; 5 hộp có 1 sản phẩm loại I.

Bài 9. Qua sát trong một số ngày về số tai nạn giao thông X xảy mỗi ngày ở một thành phố ta được số liệu sau:

Số tai nạn X	0	1	2	3	4	≥ 5
Số ngày	10	32	46	35	20	13

Với mức ý nghĩa 1%, có thể xem số tai nạn giao thông xảy mỗi ngày ở thành phố trên là ĐLNN có phân phối Poisson được hay không?

Bài 10. Quan sát năng suất X của một giống lúa thử nghiệm trên 100 thửa ruộng ta có kết quả sau:

Năng suất (tấn/ha)	8-9	9-10	10-11	11-12	12-13	13-14	14-15
Số thửa	8	15	21	23	16	9	8

Với mức ý nghĩa 5% có thể xem năng suất X là ĐLNN có phân phối chuẩn hay không?

Bài 11. Bảng số liệu điều tra về tình hình học tập của 10.000 sinh viên của một trường đại học như sau:

	Giỏi	Khá	Trung bình và kém
Nam	1620	2680	2500
Nữ	880	1320	1000

Với mức ý nghĩa 5%, xét xem có sự khác biệt về chất lượng học tập của nam và nữ hay không.

ĐÁP SỐ

Bài 1. 1a) 127,2447cm - 134,7553cm; 134,1902cm; 127,8098cm
1b) 39 **1c)** 98,82%
1d) $|z| = 2,1942 < 2,58 = z_\alpha$: Tài liệu cũ còn phù hợp.
1e) 25,65% - 44,35%; 42,87%; 27,13%
1f) 96,42% **1g)** 0
1h) $|z| = 1,0206 < 1,96 = z_\alpha$: Làm thay đổi tỉ lệ cây cao.
1i) 113,936cm - 118,064cm
1j) $|z| = 3,5 > 2,492 = t^k_\alpha$: Thay đổi chiều cao trung bình cây loại A, theo hướng tăng.
1kα) 254,7082(cm^2) - 444,6121(cm^2)
1kβ) 253,9354(cm^2) - 443,2631(cm^2)
1l) $\chi^2_{1-\frac{\alpha}{2}} = 74,222 \le \chi^2 = 109,6662 \le 129,56 = \chi^2_{\frac{\alpha}{2}}$: Bình thường.
Bài 2. 2a) 166,9584tấn - 180,5616tấn **2b)** 1392.
Bài 3. 3a) 10,43% - 23,57% **3b)** 4261-9587**3c)** 1043- 2357
3d) 6795 - 18342.
Bài 4. 4a) 8,21% - 9,79% **4b)** 79,5% **4c)** 5.
Bài 5. 5a) 8362 - 12437 **5b)** 12121 **5c)** 8651.
Bài 6. 6a) $z = 3,2913 > 2,33 = z_{2\alpha}$: Tăng chiều cao trung bình.
6b) $-z = 1,6457 < 2,06 = z_{2\alpha}$: Không làm giảm chiều cao TB.
6c) $z = 2 > 1,974 = t_{2\alpha}$: Làm giảm chiều cao trung bình cây loại A.

6d) $-z = 4 > 2,1715 = t_{2\alpha}$: Làm giảm chiều cao TB cây loại A.
6e) $-z = 2,0966 > 2,06 = z_{2\alpha}$. Làm tăng tỉ lệ cây loại A.
6f) $z = 1,25 < 1,65 = z_{2\alpha}$. Không làm tăng tỉ lệ cây loại A.
6g) $\chi^2 = 131,5995 > 124,3 = \chi^2_\alpha$: Chiều cao biến động hơn.
6h) $\chi^2 = 93,9996 > 77,93 = \chi^2_{1-\alpha}$: Chiều cao không ít biến động hơn.
Bài 7. 7a) $|z| = 1,7188 < 2,58 = z_\alpha$: Bằng nhau.
7b) $z = 1,7188 > 1,65 = z_{2\alpha}$: Của nhà máy I sản xuất lớn hơn.
7c) $z = 1,7188 < 2,06 = z_{2\alpha}$: Không thể xem của nhà máy II nhỏ hơn.
7d) $|z| = 1,6942 < 2,06 = z_\alpha$: Như nhau.
7e) $z = 1,6942 < 1,88 = z_{2\alpha}$: Của nhà máy I lớn hơn.
7f) $z = 1,6942 > 1,65 = z_{2\alpha}$: Của nhà máy II nhỏ hơn.
Bài 8. $\chi^2 = 2,881 < \chi^2_\alpha = 9,21$: X có PP nhị thức X ~ B(3 ; 0,9).
Bài 9. $\chi^2 = 2,4592 < 13,28 = \chi^2_\alpha$: X có phân phối Poisson.
Bài 10. $\chi^2 = 1,9883 < 9,488 = \chi^2_\alpha$: X có phân phối chuẩn.
Bài 11. $\chi^2 = 32,52 > \chi^2_\alpha = 5,99$: Có sự khác biệt.
