ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

TIỂU LUẬN

NHẬP MÔN SUY DIỄN THỐNG KÊ
Kiểm định giả thiết một biến số

Đại học Quốc gia Hà Nội
Đại học Khoa học Tự nhiên
Khoa Toán cơ tin

Giảng viên: Hoàng Thị Phương Thảo

Thành viên nhóm: NGUYỄN MẠNH LINH, LÊ THỊ VÂN LY

Mục lục

Tóm tắt

Tiểu luận này trình bày một số phương pháp lặp giải hệ phương trình tuyến tính lớn bằng cách sử dụng kĩ thuật tối thiểu hóa phần dư tại mỗi bước lặp (minimal residual (MR)). Ma trận của hệ tuyến tính được xem xét cả dạng tổng quát (sử dụng phương pháp GMRES (Generalized minimal residual method)) và dạng đối xứng xác định dương (sử dụng phương pháp MINRES (Minimal residual method)). Không gian con Krylov được xem xét để tìm nghiệm gần đúng của hệ. Bài này cũng giới thiệu về phương pháp lặp Arnoldi và Lanczos. Cuối cùng ta sử dụng ngôn ngữ lập trình Python để minh họa các thuật giải và so sánh hiệu năng của chúng.

1 Phương pháp lặp đơn giản

Ma trận M hệ phương trình tuyến tính $Ax = b$, ta cùng tìm một ý tưởng tự nhiên để giải gần đúng nghiệm của phương trình này. Ma trận M được chọn sao cho $M^{-1}A$ theo một nghĩa nào đó gần đúng với ma trận đơn vị. $M^{-1}(b - Ax_k)$ có thể được dùng làm xấp xỉ lỗi $A^{-1}b - x_k$ với nghiệm gần đúng x_k

2 Tổng quan

Xem xét hệ phương trình tuyến tính $Ax = b$ trong đó $A \in \mathbb{C}^{m \times m}$ và $b \in \mathbb{C}^m$.

Nghiệm chính xác của hệ $x_* = A^{-1}b$. Đương nhiên việc tính ma trận nghịch đảo A^{-1} hoặc dùng phương pháp khử Gauss là rất tốn chi phí. Ta biết rằng độ phức tạp của thuật toán khi sử dụng phương pháp khử Gauss là $O(m^3)$. Để có cái nhìn trực quan về độ phức tạp này, ta có thể xem qua bảng sau

Năm	m	m^3
1950	20	2×10^3
1965	200	2×10^6
1980	2000	2×10^9

Khi kích thước của ma trận tăng lên, số phép tính cũng tăng lên rất nhanh. Cùng với sự phát triển của phần cứng chúng ta cũng tính toán được với những ma trận với kích thước lớn hơn rất nhiều so với những năm 1950. Nhưng ta cũng thấy rằng, khi kích thước của ma trận là hàng nghìn thì số phép tính đã là hàng tỉ. Với những bài toán lớn như mô phỏng thời tiết hay thiên hà trong thiên văn học, phần cứng máy tính không thể đuổi kịp để tính toán cũng như lưu trữ.

Điều này đòi hỏi chúng ta cần tìm ra những phương pháp tốt hơn nhằm giảm độ phức tạp của thuật toán. Tiểu luận này giới thiệu 1 vài phương pháp với độ phức tạp $O(m^2)$.

Phương pháp lặp cho phép chúng ta tiến gần đến nghiệm chính xác của phương trình qua mỗi bước lặp. Đến một lúc nào đó phần dư là đủ chấp nhận được, ta có nghiệm gần đúng của phương trình.

Gọi x_n là nghiệm gần đúng của phương trình tại bước lặp thứ n. Ta định nghĩa phần dư

$$r_n = b - Ax_n \tag{2.1}$$

Khi $r_n < \epsilon$ (là một sai số nhỏ chấp nhận được) ta dừng lại và thu được nghiệm gần đúng x_n

Trong bài này chúng ta sẽ tìm nghiệm $x_n \in \kappa_n$, trong đó κ_n là không gian con Krylov

Không gian con Krylov

Cho ma trận $A \in \mathbb{C}^{m \times m}$ và vector $b \in \mathbb{C}^m$.

Dãy Krylov được định nghĩa là tập các vectors $b, Ab, A^2b, ...$

Không gian con Krylov là không gian được sinh bởi các vectors này

Ma trận Krylov

$$K_n = \left[b \quad | \quad Ab \quad | \quad A^2b \quad ... \quad | \quad A^{n-1}b\right] \tag{2.2}$$

Trở lại bài toán, ta muốn tìm nghiệm x_n trong không gian κ_n hay nói các khác $x_n \in range(K_n)$. Ta có thể sử dụng phân tích QR cho ma trận Krylov và đặt $x_n = Q_n y$ trong đó Q_n là ma trận unitary.

Bài toán cực tiểu hóa r_n đưa về việc cực tiểu hóa $||AQ_ny - b||$

3 Kiểm định giả thiết cho trung bình

3.1 Kiểm định giả thiết cho trung bình với mẫu có phân bố chuẩn và phương sai đã biết

Giả thiết null $H_0 : \mu = \mu_0$
Thực hiện test với mẫu ngẫu nhiên kích thước n, phân bố của $\bar{X}$ là phân bố chuẩn $N(\mu_0, \sigma/\sqrt{n})$ (Giải sử H_0 là đúng).
Định lý giới hạn trung tâm cũng chỉ ra rằng $\bar{X}$ có phân bố chuẩn với cỡ mẫu lớn.
Tham số chuẩn hóa

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \tag{3.1}$$

Ta có

Đối thiết	$H_1 : \mu < \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu \neq \mu_0$
Miền bác bỏ	$z < z_\alpha$	$z > z_{1-\alpha}$	$\lvert z \rvert > z_{1-\alpha/2}$

Trong đó $\Phi(z_\alpha) = \alpha$
Nhắc lại, trong ngôn ngữ R

```
z.x = qnorm(x)
```

demo code:

```
sigma <- 6
mu <- 40
pnorm(-2)
pnorm(2)
```

3.2 Kiểm định cho trung bình với mẫu có phân bố chuẩn và phương sai chưa biết

Giả thiết null $H_0 : \mu = \mu_0$
Tham số chuẩn hóa được xem xét trong trường hợp này khác với trường hợp đã biết phương sai của tổng thể :

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \tag{3.2}$$

Ta có:

Đối thiết	$H_1 : \mu < \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu \neq \mu_0$
Miền bác bỏ	$t < t_{\alpha;n-1}$	$t > t_{1-\alpha;n-1}$	$\lvert t\rvert > t_{1-\alpha/2;n-1}$

Để tính $t_{x;y}$ trong R ta dùng hàm:

```
t.x.y = qt(x, y)
```

3.3 Bài tập

pr3: Mức ý nghĩa 5%, tính power với giả thiết $H_0 : \mu = 100$ với đối thiết $H_1 : \mu \neq 100$ với cỡ mẫu 36 có phân phối chuẩn $N(120, 50)$

Tham số chuẩn hóa:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Hàm power:

$$1 - \beta = P(|Z| > Z_{1-\alpha/2}) = P(Z < Z_{\alpha/2}) + P(Z > Z_{1-\alpha/2})$$

Code R:

```
mu = 100
mu0 = 120
sigmaX = 50 / 6
alpha = 0.05
z1 = qnorm(alpha / 2, mu, sigmaX)
z2 = qnorm(1 - alpha / 2, mu, sigmaX)
pnorm(z1, mu0, sigmaX) + 1 - pnorm(z2, mu0, sigmaX)
```

Kết quả $1 - \beta = 0.67$

```
[1] 0.670051
```

pr4: Kiểm định cho giả thiết $H_0 : \mu = 170$ và đối thiết $H_1 : \mu < 170$ cho đại lượng loss trong tập dữ liệu Rubber.

Cách làm đơn giản với code R:

```
library(MASS)
data("Rubber")
loss <- Rubber$loss
t.test(loss, alternative = c("less"), mu = 170)
```

Kết quả:

```
        One Sample t-test

data:  loss
t = 0.33785, df = 29, p-value = 0.631
alternative hypothesis: true mean is less than 170
95 percent confidence interval:
    -Inf 202.7589
sample estimates:
mean of x
175.4333
```

Kết luận: với độ tin cậy 95% ta có thể bác bỏ giả thiết H_0, chấp nhận $H_1 : \mu < 170$

pr12: Xem xét data frame Fertilize, chiều cao của cây được cho qua biến self

(a) Kiểm định giả thiết cho chiều cao trung bình của cây lớn hơn 17 inchs với mức ý nghĩa $\alpha = 5\%$

(b) Ước lượng khoảng một phía với độ tin cậy 95% cho chiều cao trung bình ($H_1 : \mu > 17$)

(c) Tính cỡ mẫu cần thiết để có giá trị power là 0.9 với $\mu_1 = 18$ inchs giả thiết $\sigma = s$

(d) Tính power cho kiểm định trong phần a với $\sigma = s$ và $\mu_1 = 18$

Lời giải:

(a) Giả thiết $H_0 : \mu = 17$ và đối thiết $H_1 : \mu > 17$

```
library(PASWR)
data("Fertilize")
self <- Fertilize$self
t.test(self, mu = 17, alternative = c("greater"))
```

Kết quả thu được:

```
        One Sample t-test

data:  self
t = 1.0854, df = 14, p-value = 0.148
alternative hypothesis: true mean is greater than 17
95 percent confidence interval:
16.64196      Inf
sample estimates:
mean of x
17.575
```

Ta thấy $p - value = 0.148 > \alpha = 0.05$ nên chưa đủ cơ sở để bác bỏ H_0.

Kết luận: chưa đủ cơ sở để kết luận chiều cao trung bình $\mu > 17$ inchs với mức ý nghĩa $\alpha = 5\%$.

(b) Ước lượng khoảng với độ tin cậy 95% cho chiều cao trung bình của cây là $\mu > 16.642$ inchs

(d)

$$
\begin{aligned}
power(\mu_1) = P(\text{chấp nhận } H_1|H_1 \text{ đúng}) &= P\left(\mu_{obs} > \mu_0 | N\left(\mu_1, \frac{s}{\sqrt{n}}\right)\right) \\
&= P\left(Z > Z_{1-\alpha}\right) = 0.596
\end{aligned}
$$

```
n <- length(self)
z <- qnorm(0.95, 17, sqrt(s2/n))
round(1 - pnorm(z, 18, sqrt(s2/n)), 3) # = 0.596
```

(c) Thử cho giá trị n khác nhau vào đoạn code `R` trên và tính giá trị power ta thấy giá trị nhỏ nhất để có $power = 0.9$ là $n = 36$

pr18: Tìm power của giả thiết $H_0 : \mu = 65$, đối thiết $H_1 : \mu > 65$ nếu $\mu_1 = 70$, mức ý nghĩa $\alpha = 0.01$ cho data `hard`, data frame `Rubber`. Giả sử $\sigma = s$

Lời giải:

$$
\begin{aligned}
power(\mu_1) = P(\text{chấp nhận } H_1|H_1 \text{ đúng}) &= P\left(\mu_{obs} > \mu_0 | N\left(\mu_1, \frac{s}{\sqrt{n}}\right)\right) \\
&= P(Z > Z_{1-\alpha}) = 0.496
\end{aligned}
$$

`R` code:

```
library(MASS); data("Rubber")
hard <- Rubber$hard
n <- length(hard)
s2 <- var(hard)
z <- qnorm(0.99, 65, sqrt(s2/n))
round(1 - pnorm(z, 70, sqrt(s2/n)), 3) # = 0.469
```

4 Kiểm định cho phương sai

4.1 Kiểm định cho phương sai với mẫu có phân bố chuẩn

Cho n mẫu ngẫu nhiên $X_1, X_2, .., X_n$ có phân bố chuẩn $N(\mu, \sigma)$, đại lượng chuẩn hóa sau có phân bố χ^2:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \tag{4.1}$$

Giả thiết null cho phương sai $H_0 : \sigma^2 = \sigma_0^2$, giá trị test thống kê:

$$\chi^2_{obs} = \frac{(n-1)S^2}{\sigma_0^2}$$

Ba đối thiết và miền bác bỏ H_0 tương ứng cho dưới bảng sau:

Đối thiết	$H_1 : \sigma^2 < \sigma_0^2$	$H_1 : \sigma^2 > \sigma_0^2$	$H_1 : \sigma^2 \neq \sigma_0^2$
Miền bác bỏ	$\chi^2_{obs} < \chi^2_{\alpha,n-1}$	$\chi^2_{obs} > \chi^2_{1-\alpha,n-1}$	$\chi^2_{obs} < \chi^2_{\alpha/2,n-1}$ $\cup\chi^2_{obs} > \chi^2_{1-\alpha/2,n-1}$

4.2 Bài tập

pr5: Xem xét biến totalprice trong tập dữ liệu VT2005 trong package PASWR. Người thẩm định nói rằng với những căn hộ có diện tích lớn hơn hoặc bằng $90m^2$ thì độ phương sai giá nhỏ hơn $60000^2€^2$. Kiểm định giả thiết xem phương sai có lớn hơn $60000^2€^2$.

Đầu tiên ta lọc ra từ dataframe gốc những căn hộ có diện tích lớn hơn hoặc bằng $90m^2$ và tính được cỡ mẫu.

```
greater90 <- subset(x = vit2005, subset = area >= 90)
n <- length(greater90$totalprice) # n = 94
```

Sử dụng quy trình 5 bước:

Bước 1: **Giả thuyết**: Giả thuyết null, $H_0 : \sigma^2 = 60000^2$ với đối thiết $H_1 : \sigma^2 > 60000^2$

Bước 2: **Test thống kê**: Sử dụng test thống kê cho S^2. Ta tính lần lượt S^2 và tham số chuẩn hóa $\frac{(n-1)S^2}{\sigma_0^2}$ với giả định H_0 là đúng

```
v <- var(greater90$totalprice)
x <- (n - 1) * v / (60000^2)
```

so sánh tham số chuẩn hóa với $\chi^2_{\alpha,n-1}$

```
chi <- qchisq(0.95, n - 1)
```

Kết quả thu được $\chi^2_{obs} = 98.7603$ và $\chi_{0.95,94-1} = 116.511$. Miền bác bỏ: $\chi^2_{obs} > \chi^2_{1-\alpha,n-1}$.

Bước 4: **Kết luận về mặt thống kê**: Từ miền bác bỏ, không thể bác bỏ H_0 do $\chi^2_{obs} = 98.7603$ không lớn hơn $\chi_{0.95,94-1} = 116.511$

Bước 5: **Kết luận**: Không đủ bằng chứng để khẳng định phương sai của giá cắn hộ có diện tích lớn hơn hoặc bằng $90m^2$ thì lớn hơn 60000^2€^{2}

5 Kiểm định giả thiết cho tỉ lệ

5.1 Kiểm định cho tỉ lệ với trong test nhị thức

Ta xem xét biến ngẫu nhiên có phân bố nhị thức $Y \sim Bin(n, \pi)$, giả thiết null $H_0 : \pi = \pi_0$. Test thống kê cho y_{obs} = Số quan sát thành công

Ta có 3 đối thiết ứng với công thức p-value tương ứng như sau:

Đối thiết	p-value
$H_1 : \pi < \pi_0$	$P(Y \leqslant y_{obs} \| H_0) = \sum_{i=0}^{y_{obs}} \binom{n}{i} \pi_0^i (1-\pi_0)^{n-i}$
$H_1 : \pi > \pi_0$	$P(Y \geqslant y_{obs} \| H_0) = \sum_{i=y_{obs}}^{n} \binom{n}{i} \pi_0^i (1-\pi_0)^{n-i}$
$H_1 : \pi \neq \pi_0$	$\sum_{i=0}^{n} I(P(Y=i) \leqslant P(Y=y_{obs})) \binom{n}{i} \pi_0^i (1-\pi_0)^{n-i}$

5.2 Kiểm định cho tỉ lệ trong test nhị thức (xấp xỉ chuẩn)

Theo tính chất tiệm cận của ước lượng MLE ta có:

$$P = \frac{Y}{n} \sim N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right) \text{ với } n \to \infty \tag{5.1}$$

Trong thực tế Với $n\pi$ và $n(1-\pi)$ đều lớn hơn hoặc bằng 10, ta có thể sử dụng xấp xỉ trên.

Đại lượng chuẩn hóa:

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1) \tag{5.2}$$

Khi $|p - \pi_0| > \frac{1}{2n}$ nhiều nhà thống kê cho rằng cần phải thêm vào 1 hiệu chỉnh. Hàm `prop.test()` trong `R` mặc định thêm vào hiệu chỉnh $\pm\frac{1}{2n}$.

Ta có bảng tổng quan như sau

Giải thiết null $H_0 : \pi = \pi_0$

Đối thiết	miền bác bỏ
$H_1 : \pi < \pi_0$	$z_{obs} < z_\alpha$
$H_1 : \pi > \pi_0$	$z_{obs} > z_{1-\alpha}$
$H_1 : \pi \neq \pi_0$	$\|z_{obs}\| > z_{1-\alpha/2}$

Đối với $|p - \pi_0| > \frac{1}{2n}$ ta có bảng hiệu chỉnh như sau

Điều kiện	Hiểu chỉnh	Tham số chuẩn hóa
$p - \pi_0 > 0$	$-\frac{1}{2n}$	$z_{obs} = \frac{p-\pi_0-\frac{1}{2n}}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$
$p - \pi_0 < 0$	$\frac{1}{2n}$	$z_{obs} = \frac{p-\pi_0+\frac{1}{2n}}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$

5.3 Bài tập

pr17: Theo Pamplona, Tây Ban Nha, 0.4% người nhập cư năm 2002 đến từ Bolivia. Tháng 6 năm 2005, một nhóm gồm 3740 người nước ngoài được điều tra ngẫu nhiên thì có 87 người Bolivia. Có thể phòng đoán số người nhập cư Boliva đã tăng lên hay không với mức ý nghĩa $\alpha = 0.05$.

Giả thiết null: $H_0 : P = \pi_0 = 0.004$, đối thiết $P > \pi_0 = 0.004$ Với $n = 3740$, ta có $n\pi_0$ và $n(1 - \pi_0)$ đều lớn hơn 10 nên ta có thể sử dụng xấp xỉ chuẩn để kiểm định cho tỉ lệ trong bài toán này.

Ta có $P = 87/3740$ nên $P - \pi_0 = 0.01926203 > 1/2n = 0.0001336898$, do đó ta phải sử dụng hiệu chỉnh $1/2n$

tham số chuẩn hóa

$$Z = \frac{P - \pi_0 - \frac{1}{2n}}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = 18.53333 > 1.644854 = Z_{0.95}$$

Do đó ta có thể kết bác bỏ giả thiết H_0

Kết luận: với mức ý nghĩa 5% ta có thể nói rằng tỉ lệ người nhập cư vào Tây Ban Nha đến từ Bolivia (điều tra tháng 6 năm 2005) đã tăng lên so với thời điểm năm 2002.

Ngoài ra, chúng ta có thể sử dụng phương pháp tính giá trị `p-value` được đề cập trong phần 4.1. Hoặc có thể sử dụng test dựng sẵn trong `R` đơn giản như sau:

```
 binom.test(x = 87, n = 3740, p = 0.004, alternative = "
greater")
```

Kết quả thu được:

```
        Exact binomial test

data:  87 and 3740
```

```
number of successes = 87, number of trials = 3740, p-value <
 2.2e-16
alternative hypothesis: true probability of success is
greater than 0.004
95 percent confidence interval:
0.01935316 1.00000000
sample estimates:
probability of success
             0.02326203
```

`p-value` $< \alpha = 0.05$, ta có kết luận tương tự như trên.

pr19: Giám đóc nhà đô thị ở Vitoria, Tây Ban Nha nói rằng ít nhất 50% các căn hộ có nhiều hơn 1 nhà tắm và ít nhất 70% các căn hộ có một thang máy.

(a) Kiểm chứng tuyên bố của giám đốc. Test thống kê với mức ý nghĩa $\alpha = 0.1$. Dữ liệu trong cột `toilets`, data frame `vit2005`.

(b) Kiểm chứng tuyên bố về thang máy với mức ý nghĩa $\alpha = 0.1$ sử dụng phương pháp xấp xỉ và phương pháp chính xác.

(c) Kiểm tra xem tỷ lệ các căn hộ được xây dựng trước năm 1980 không có nhà để xe có tỷ lệ có thang máy cao hơn không có thang máy?

Lời giải: (a) Bài toán kiểm định: Giả thiết $H_0 : \pi = 0.5$ với đối thiết $H_1 : \pi > 0.5$

Ta sử dụng hàm dựng sẵn trong `R` `binom.test()` như sau:

```
data("vit2005")
toilets <- vit2005$toilets
n <- length(toilets)
ns <- length(toilets[toilets > 1])
binom.test(ns, n, 0.5, alternative = c("less"), conf.level =
 0.1)
```

Thu được kết quả:

```
    Exact binomial test

data:  ns and n
number of successes = 102, number of trials = 218, p-value =
 0.8452
alternative hypothesis: true probability of success is
greater than 0.5
10 percent confidence interval:
0.5089864 1.0000000
sample estimates:
probability of success
             0.4678899
```

Ta thấy giá trị `p-value` $= 0.8452 > 0.1$ nên chưa đủ cơ sở để bác bỏ H_0

Kết luận: Chưa đủ cơ sở để bác bỏ tuyên bố có ít nhất 50% số căn hộ có nhiều hơn 1 nhà tắm với mức ý nghĩa $\alpha = 0.1$.

(b) Bài toán kiểm định: Giả thiết: $H_0 : p = 0.75$ với đối thiết $H_1 : p > 0.75$

Phương pháp xấp xỉ chuẩn

Đầu tiên ta khởi tạo data cần tính toán:

```
data("vit2005")
elevator <- vit2005$elevator
n <- length(elevator) # = 218
ns <- length(elevator[elevator == 1]) # = 174
```

Kiểm tra hiệu chỉnh:

```
pi0 <- 0.75
p <- ns / n # = 174/218 > 0.75
p - pi0 - 1/(2*n) # > 0
```

Ta có tham số chuẩn hóa

$$Z = \frac{p - \pi_0 - \frac{1}{2n}}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0, 1)$$

Thay số vào phương trình trên ta tính được giá trị của $Z = 1.564124 > Z_{0.9} = 1.281552$ nên có thể bác bỏ giả thiết H_0

Kết luận: Chấp nhận khẳng định có ít nhất 75% số căn hộ có thang máy với mức ý nghĩa $\alpha = 0.1$.

Phương pháp chính xác

Tính giá trị `p-value`:

$$\begin{aligned}\text{p-value} = P(Y \geqslant y_{obs}|H_0) &= \sum_{i=y_{obs}}^{n} \binom{n}{i} \pi_0^i (1-\pi_0)^{n-i} \\ &= \sum_{i=174}^{218} \binom{218}{i} 0.75^i (0.25)^{218-i}\end{aligned}$$

Ta có thể tính trong `R` như sau:

```
pvalue <- sum(dbinom(ns:n, n, 0.75)) # = 0.05643458
```

Ta có `p-value` < 0.1 nên có thể bác bỏ H_0.

Ngoài ra cũng có thể dùng hàm `binom.test()` như sau:

```
binom.test(x = ns, n = n, p = 0.75, alternative = c("greater
"), conf.level = 0.1)
```

Ta thu được kết quả tương tự:

```
     Exact binomial test

data:  ns and n
number of successes = 174, number of trials = 218, p-value =
 0.05643
alternative hypothesis: true probability of success is
greater than 0.75
10 percent confidence interval:
0.8288551 1.0000000
sample estimates:
probability of success
             0.7981651
```

Kết luận: Chấp nhận khẳng định có ít nhất 75% số căn hộ có thang máy với mức ý nghĩa $\alpha = 0.1$.