

# Bài giảng 23: Ước tính cỡ mẫu: Nghiên cứu 1 nhóm

**Nguyễn Văn Tuấn**

Viện nghiên cứu y khoa Garvan (Úc)

Đại học New South Wales và Đại học Công nghệ Sydney

Đại học Tôn Đức Thắng, Việt Nam

# Ước tính cỡ mẫu cho nghiên cứu 1 nhóm

- Biến outcome là biến phân loại
- Biến outcome là biến liên tục

**Cỡ mẫu cho tỉ lệ**

# Giới thiệu vấn đề

- Quần thể (population)
- Gọi  $\pi$  là tỉ lệ mắc bệnh
- Mẫu  $n$  đối tượng
- Gọi  $p$  là tỉ lệ quan sát
- Chúng ta muốn dùng  $p$  để ước tính  $\pi$

# Ước tính khoảng tin cậy 95%

- Quần thể (population)
- Gọi  $\pi$  là tỉ lệ mắc bệnh
- Mẫu  $n$  đối tượng
- Gọi  $p$  là tỉ lệ quan sát
- Chúng ta muốn dùng  $p$  để ước tính  $\pi$

**Lí thuyết:**  $p$  là ước số khách quan nhất của  $\pi$ .

Khoảng tin cậy 95% của  $\pi$ :

$$p - z_{\alpha/2}e < \pi < p + z_{\alpha/2}e$$

Nếu chọn KTC95%, tức  $z_{\alpha/2} = 1.96$

$$p - 2e < \pi < p + 2e$$

# Ước tính khoảng tin cậy 95%

**Lí thuyết:**  $p$  là ước số khách quan nhất cho  $\pi$ .

Khoảng tin cậy 95% của  $\pi$ :

$$p - 2e < \pi < p + 2e$$

Trong đó,  $e$  là sai số chuẩn:

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

Nói cách khác,

$$\left( p - 2\sqrt{\frac{p(1-p)}{n}} \right) < \pi < \left( p + 2\sqrt{\frac{p(1-p)}{n}} \right)$$

# Đặt ngược vấn đề

Tìm  $n$  để ước tính  $\pi$  với một sai số  $e$

$$n = \left( \frac{2}{e} \right)^2 p(1-p)$$

Hay nói chung:

$$n = \left( \frac{z_{\alpha/2}}{e} \right)^2 p(1-p)$$

# Cỡ mẫu cho một tỉ lệ

Do đó, cỡ mẫu cho một tỉ lệ tùy thuộc vào:

- Sai số I (alpha):  $\alpha$
- Tỉ lệ bệnh (hay biến cố):  $p$
- Sai số (marginal error):  $e$



# Ví dụ

- Mục tiêu: xác định tỉ lệ bệnh tiểu đường
- Theo y văn, tỉ lệ bệnh tiểu đường trong cộng đồng người trung niên là 10% ( $p = 0.10$ )
- Chấp nhận xác suất 95% là tỉ lệ có thể dao động trong khoảng 8% đến 12%
- Nói cách khác,  $e = (0.12 - 0.08) / 4 = 0.01$

# Dùng R

- Cách tính cỡ mẫu bằng R rất đơn giản!
- Có thể viết vài mã R, hoặc dùng hàm *n.for.survey* trong package *epicalc*

```
alpha=0.05; z = qnorm(0.05/2)
```

```
p=0.10; e = 0.01
```

```
n = (z/e)^2*p*(1-p)
```

# Package epicalc

```
library(epicalc)
```

```
n = n.for.survey(p=0.10, delta=0.01,  
alpha=0.05)
```

e



```
> n
```

Sample size for survey.

Assumptions:

Proportion = 0.1

Confidence limit = 95 %

Delta = 0.01 from the estimate.

Sample size = 3457

**Cỡ mẫu cho số trung bình**

# Giới thiệu vấn đề

- Quần thể (population)
- Gọi  $\mu$  là trung bình
- Mẫu  $n$  đối tượng
- Gọi  $m$  là tỉ lệ quan sát
- Chúng ta muốn dùng  $m$  để ước tính  $\mu$

# Ước tính khoảng tin cậy 95%

- KTC95% của  $\mu$

$$m - z_{\alpha/2}e < \mu < m + z_{\alpha/2}e$$

- Nếu  $\alpha = 5\%$ , KTC95% của  $\mu$

$$m - 2e < \mu < m + 2e$$

- Trong đó,

$$e = \frac{s}{\sqrt{n}}$$

# Cỡ mẫu để ước tính $\mu$ cho trung bình

- KTC95% của  $\mu$

$$\left( m - z_{\alpha/2} \frac{s}{\sqrt{n}} \right) < \mu < \left( m + z_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

- Do đó, cỡ mẫu để ước tính  $\mu$  với sai số  $e$  là:

$$n = \left( \frac{z_{\alpha/2} s}{e} \right)^2$$

# Ví dụ

- Mục tiêu: ước tính tuổi bị ung thư vú
- Độ lệch chuẩn của độ tuổi ung thư là 12 tuổi
- Muốn ước tính độ tuổi ung thư chính xác trong khoảng  $\pm 2.5$  tuổi (tức độ rộng của khoảng tin cậy 95% là 5 tuổi).
- Nói cách khác  $e = 2.5$  và  $s = 12$



# Dùng R

```
alpha=0.05; z = qnorm(0.05/2)
```

```
sigma=12; e = 2.5
```

```
n = (z/e*sigma)^2 ; n
```

```
> n = (z/e*sigma)^2 ; n
```

```
[1] 88.50721
```