

Lâm sàng thống kê

Ước tính cỡ mẫu để đo ... dương vật

Xin nói ngay rằng tiêu đề trên không phải là chuyện đùa, mà là chuyện khoa học nghiêm chỉnh! Một [nghiên cứu](#) quan trọng và mới ở nước ta vừa được một đồng nghiệp ở Bệnh viện Bình Dân tiến hành để tìm hiểu kích thước dương vật ở đàn ông Việt Nam. Trả lời phóng viên Vietnamnet về độ khác biệt giữa chiều dài của dương vật người Việt Nam và các dân tộc khác, vị đồng nghiệp bình luận: *“Để kết luận chính xác thì cần phải đo tất cả đàn ông Việt Nam. Điều này phi thực tế! Trong thống kê, người ta cho phép chọn ngẫu nhiên một số người để đo, rồi qui ra con số trung bình cho cả một tập thể lớn. Số người tôi chọn là 100, cũng tương đương với các nghiên cứu khác, ngoại trừ nghiên cứu của Mỹ có số lượng người được đo trên 1000. Do đó, có thể nói kết luận của nghiên cứu của chúng tôi là chấp nhận được.”*

Có lẽ chúng ta đều đồng ý với anh đồng nghiệp là nếu muốn biết chính xác kích thước dương vật của đàn ông Việt Nam, cần phải đo tất cả đàn ông người Việt Nam, nhưng việc làm này chẳng những không thực tế, mà còn không cần thiết. Tuy nhiên, chọn 100 người để đo và kết luận rằng kết quả “có thể chấp nhận được” thì có lẽ cần phải bàn thêm, bởi vì vấn đề mấu chốt là: thế nào để có thể xem là “chấp nhận được”. Nói cách khác, dựa vào tiêu chí khoa học nào để kết quả có thể xem là chấp nhận được. Nếu mục tiêu là ước tính chiều dài của dương vật đàn ông Việt Nam, thì tiêu chí ở đây phải là *độ chính xác* của ước số. Do đó, câu hỏi quan trọng cần được đặt ra là: cần phải đo bao nhiêu đàn ông để có thể có một độ chính xác có thể chấp nhận được. Trong bài này tôi sẽ bàn ngắn gọn về câu trả lời mà tôi tin rằng rất nhiều bạn đọc cũng muốn biết.

Theo một suy nghĩ thông thường, câu trả lời cho câu hỏi “cần đo bao nhiêu” là “càng nhiều càng tốt”. Nhưng có thật là càng nhiều càng tốt hay không? Số cỡ mẫu bao nhiêu thì gọi là “nhiều”? Thật ra, đứng trên quan điểm nghiên cứu và xác suất, không hẳn đo càng nhiều càng tốt, vì đến một số cỡ mẫu nào đó, độ chính xác sẽ không tăng hơn nữa. (Bạn đọc có thể chứng minh câu phát biểu này một cách dễ dàng bằng cách sử dụng phần mềm R như tôi cung cấp trong phần ghi chú phía dưới bài viết).

Gọi M là số trung bình của chiều dài dương vật ở đàn ông Việt Nam. Chúng ta không biết giá trị của M (vì không thể đo toàn bộ dân số), cho nên chúng ta phải chọn ngẫu nhiên một nhóm đàn ông (gọi là “mẫu” - sample) từ một quần thể (tức là population) đàn ông người Việt để ước tính M .

Giả dụ chúng ta chọn ngẫu nhiên một nhóm gồm N đàn ông, và đo chiều dài dương vật của họ. Gọi m là số trung bình, và s là độ lệch chuẩn chiều dài dương vật của N đối tượng. Chúng ta muốn tìm N sao cho m càng gần với giá trị của M càng tốt.

Nếu chiều dài dương vật của các đối tượng tuân theo luật phân phối chuẩn (Normal distribution), chúng ta biết rằng xác suất 95% là M nằm trong khoảng $m - \frac{1.96 \times s}{\sqrt{N}}$ đến khoảng $m + \frac{1.96 \times s}{\sqrt{N}}$. Cũng có thể nói rằng xác suất 99% là M nằm trong khoảng $m \pm \frac{3 \times s}{\sqrt{N}}$ (xem bài “[Độ lệch chuẩn hay sai số chuẩn](#)” trong mục *lâm sàng thống kê*).

Do đó, sai số đo lường (tạm kí hiệu là h) có thể định nghĩa bằng: $h = \frac{z \times s}{\sqrt{N}}$, trong đó z là hằng số từ phân phối chuẩn cho xác suất 95% hay 99%. Nếu xác suất khoảng tin cậy là 95%, $z = 1.96$; và xác suất 99%, $z = 3.0$. Nói cách khác, nếu chúng ta muốn có một sai số h với xác suất khoảng tin cậy 95% thì công thức ước tính cỡ mẫu là:

$$N = \frac{(1.96)^2 s^2}{h^2}$$

hay với khoảng tin cậy 99%:

$$N = \frac{9s^2}{h^2}$$

Trong công thức trên đây, chúng ta cần phải biết một thông số quan trọng: đó là độ lệch chuẩn s của chiều dài dương vật. Ngoài ra, chúng ta còn phải xác định sai số h là bao nhiêu. **Không có hai thông số này thì không thể ước tính được số lượng cỡ mẫu cần thiết.**

Quay lại với nghiên cứu của vị đồng nghiệp, rất tiếc là chúng ta chỉ biết chiều dài trung bình của dương vật trong nhóm đàn ông Việt Nam là 6.4 cm (khi chưa cương) và 11.1 cm (khi cương), mà không biết độ lệch chuẩn là bao nhiêu. Chúng ta có thể tìm thông số này trong y văn. Thật vậy, chỉ cần không đầy 5 phút, tôi đã có được vài bài báo, mà trong đó độ lệch chuẩn chiều dài của dương vật là khoảng 1.4 cm [1].

Kể đến, chúng ta vẫn cần phải quyết định chọn sai số h . Vì đây là nghiên cứu nhằm xác định chiều dài dương vật của đàn ông Việt Nam (có ý nghĩa lớn về nhân chủng học và thương mại), nên chúng ta cần có độ chính xác cao. Thế nào là cao? Chú ý rằng độ lệch chuẩn là 1.4 cm, do đó, sai số mà chúng ta chấp nhận phải 10% (hay thấp hơn) độ lệch chuẩn. Nói cách khác, chúng ta muốn có m (số trung bình mẫu) xê xích với M (số trung bình của quần thể) khoảng $h = 1.4 \times 0.1 = 0.14$ cm. (Tức là m có thể cao hơn hay thấp hơn M khoảng 0.14 cm). Với $s = 1.4$ cm và $h = 0.14$ cm, áp dụng công thức trên, chúng ta có:

$$N = \frac{(1.96)^2 (1.4)^2}{(0.14)^2} = 384$$

Như vậy, chúng ta cần chọn (ngẫu nhiên) tối thiểu là 384 đối tượng để ước tính chiều dài dương vật đàn ông Việt Nam với sai số 0.14 cm và khoảng tin cậy 95%. Nếu chúng ta nâng khoảng tin cậy lên 99% (cũng cần thiết), thì số cỡ mẫu cần thiết lên đến 900 người trở lên:

$$N = \frac{(3)^2 (1.4)^2}{(0.14)^2} = 900$$

Có lẽ chính vì các tính toán này mà các nhà nghiên cứu Mĩ phải đo đến hơn 1000 dương vật để ước tính cho đàn ông người Mĩ. Do đó, với số cỡ mẫu 100 đàn ông mà vị đồng nghiệp đã đo lường có thể chưa đủ để ước tính chính xác chiều dài dương vật ở đàn ông Việt Nam.

Tất nhiên, để có một ước tính chính xác, không chỉ số cỡ mẫu phải đầy đủ, mà cách chọn mẫu phải có hệ thống và ngẫu nhiên. Cụm từ “ngẫu nhiên” được lặp đi lặp lại nhiều lần trong bài viết ngắn này chỉ để nói lên tầm quan trọng của cách chọn đối tượng. Nếu đối tượng được chọn ngẫu nhiên, thì tính đại diện cho quần thể sẽ cao; nhưng nếu không ngẫu nhiên thì các ước số thống kê sẽ thiếu tính khách quan. Định nghĩa “ngẫu nhiên” không phải đơn giản, nhưng có thể nói rằng một mẫu dựa vào một cơ sở y tế không thể xem là đại diện tốt cho quần thể đàn ông Việt Nam. Nhưng đây là một chủ đề quan trọng mà tôi sẽ có dịp bàn tiếp trong một dịp khác.

Vài hàng lạc đề

Nhân đọc bài báo trên vietnamnet về nghiên cứu này, tôi phải có vài hàng bàn thêm ngoài vấn đề phương pháp nghiên cứu. Có lẽ không có một cơ phận nào của con người được chú ý và bàn thảo (truyền miệng) nhiều nhất như cái dương vật. Chỉ cần tò mò “dạo” một vòng y văn thế giới đã thấy hàng trăm công trình nghiên cứu khoa học về đề tài này, từ y học lâm sàng, sinh lí học, dịch tễ học, nhân chủng học, văn hóa học, đến nghệ thuật. Đàn ông, dù là người Mĩ, Âu, hay Á châu, đều có xu hướng quan tâm đến kích thước dương vật vì họ cho rằng dương vật của họ ... ngắn. Bài báo trên Vietnamnet cung cấp vài con số thống kê thú vị về chiều dài của dương vật ở các sắc dân như sau:

Sắc dân	Chiều dài khi chưa cương (cm)	Chiều dài khi cương (cm)
Mĩ	8.8	12.9
Á Rập	9.3	13.5

Hàn quốc	6.9	9.6
Mã Lai	8.3	12.1
Việt Nam	6.4	11.1

Rất khó suy luận một cách khoa học từ các con số trên vì chúng ta không biết độ lệch chuẩn là bao nhiêu và số cỡ mẫu là bao nhiêu. Nhưng chỉ nhìn qua số trung bình, bảng thống kê trên cho thấy so với các sắc dân khác, chiều dài của dương vật đàn ông Việt Nam ngắn nhất khi chưa cương; nhưng khi cương thì cũng ... không đến nỗi ngắn. Tuy nhiên, bảng số liệu còn cho ra một câu hỏi quan trọng khác: tại sao chiều dài lúc cương và lúc chưa cương quá khác nhau giữa các sắc dân? Chẳng hạn như tại sao khi cương, dương vật Việt Nam tăng 4.1 cm (tức 73%), mà trong khi đó dương vật Hàn quốc chỉ tăng 2.7 cm (39%)? Trong khi đó, các sắc dân khác, độ tăng là 3.8 đến 4.2 cm.

Khó tìm một lí do sinh học nào để giải thích những khác biệt đó, nếu không muốn nói là lí do đặc điểm nhân trắc của các đối tượng nghiên cứu và phương pháp chọn mẫu. Tôi cho rằng các con số trên đây có vấn đề, nhất là con số của Hàn quốc và có thể cả của Việt Nam nữa.

Người ta quan tâm đến kích thước của dương vật vì cho rằng dương vật càng dài và càng dày là những đặc tính tốt. Nhưng “tốt” cho cái gì? Người ta cho rằng tốt cho quan hệ sex. Nếu giả thiết này đúng, thì đối tượng cần được tham vấn phải là phụ nữ, chứ không phải đàn ông!

Một nhóm nghiên cứu đã tiến hành một nghiên cứu ở 170 phụ nữ Âu châu để thăm dò ý kiến của họ về chiều kích của dương vật [2]. Các nhà nghiên cứu hỏi họ chiều dài và độ dày của dương vật có quan trọng với họ hay không, và kết quả cho thấy: khoảng 70% cho rằng cả chiều dài và độ dày của dương vật chẳng quan trọng hay hoàn toàn không quan trọng. Tuy nhiên, có 30% phụ nữ cho rằng chiều dài quan trọng và 20% cho rằng độ dày quan trọng [2].

Có giả thuyết cho rằng chiều dài của dương vật có tương quan thuận với chiều cao hay kích thước bàn chân (hay nói cách khác, có thể nhìn chiều cao và kích thước bàn chân để biết dương vật dài hay ngắn). Sự thật ra sao? Một nghiên cứu cho thấy hệ số tương quan giữa chiều dài dương vật và chiều cao là 0.26, và giữa chiều dài dương vật và kích thước bàn chân là 0.27 [1]. Nói cách khác, hai yếu tố này chỉ giải thích khoảng 10% độ khác biệt về chiều dài dương vật giữa các đàn ông. Giả thuyết vừa đề cập trên không chính xác!

Quay lại công trình nghiên cứu của vị đồng nghiệp ở Bệnh viện Bình Dân, tôi thấy đây là một nghiên cứu độc đáo (dù hơi tế nhị!), có ý nghĩa nhân chủng học và thương mại. Các công ti sản xuất bao cao su có thể nghiên cứu kĩ các số liệu này để thiết kế bao cao su cho thích hợp với người (à không, đàn ông) Việt Nam. Tuy nhiên, số

lượng cỡ mẫu cần được tăng thêm để nâng cao tính khoa học và độ chính xác của ước số. Số liệu cần phải được phân tích kĩ hơn và rất đáng được công bố trên một tập san y học quốc tế.

Nguyễn Văn Tuấn

10/5/07

Tài liệu tham khảo:

[1] Siminoski K, Bain J. The relationships among height, penile length and foot size. Annals of Sex Research 1993; 6:231-235.

[2] Frankcken AB, et al. What importance do women attribute to the size of the penis? European Urology 2002; 42:426-431.

Bạn đọc muốn biết thêm về các phương pháp ước tính cỡ mẫu và nguyên lí đằng sau các phương pháp này có thể tìm đọc cuốn sách “[Phân tích số liệu và tạo biểu đồ bằng R](#)” của tôi do Nhà xuất bản Khoa học Kỹ thuật phát hành đầu năm 2007.

Ngoài ra, tôi cũng có viết một bài dài về những phương pháp ước tính cỡ mẫu cho nghiên cứu y khoa đăng trên Tạp chí Thời sự Y học số tháng 4 và 5 / 2007 thuộc Hội Y học Thành phố Hồ Chí Minh (liên lạc bác sĩ Lê Anh Bá leaba@hcm.vnn.vn).

Sau đây là những giải thích về cách tính cỡ mẫu qua kỹ thuật mô phỏng với ngôn ngữ R.

Giả dụ chúng ta biết rằng đàn ông Việt Nam có chiều dài trung bình là 11.0 cm và độ lệch chuẩn là 1.45 cm. Nói theo kí hiệu trong bài viết, chúng ta có $M = 11.0$ và $s = 1.45$. Có thể sử dụng R để mô phỏng “dân số” này dễ dàng:

```
length <- rnorm(n=1000000, mean=11.0, sd=1.45)
```

Chú ý lệnh trên, chúng ta tạo ra một dân số gồm 1 triệu đàn ông với trung bình 11 cm và độ lệch chuẩn 1.45. Sau khi chạy lệnh này, chúng ta đã có một biến `length` gồm 1 triệu số.

Bây giờ, chúng ta thử lấy mẫu từ “dân số” 1 triệu người đó, và cứ mỗi lần lấy mẫu, chúng ta tính số trung bình.

(a) Chúng ta bắt đầu lấy mẫu chỉ 10 người từ quần thể trên với lệnh `sample` và lưu trữ số liệu 10 người này trong biến gọi là `sample.10`, sau đó ước tính số trung bình từ mẫu này:

```
sample.10 <- sample(length, 10)
mean(sample.10)
10.12
```

Số trung bình là 10.12 cm. Như vậy, sai số khoảng 1.12 cm (vì chúng ta biết rằng số trung bình của toàn quần thể là 11 cm). Chúng ta thử lấy 10 mẫu một lần nữa và xem số trung bình ra sao:

```
sample.10 <- sample(length, 10)
mean(sample.10)
10.57
```

Vẫn sai số hơn 1 cm. Thật ra, chúng ta có thể sử dụng lệnh sau đây để lấy mẫu 100 lần, mỗi lần 10 người, và xem độ dao động của số trung bình trong 100 lần ra sao:

```
B = 100
k = 10
mean = numeric(B)
for (i in 1:B)
{
  sample <- sample(length, k, replace=F)
  mean[i] = mean(sample)
}
quantile(mean, probs=c(0.025, 0.50, 0.975))

2.5%      50%      97.5%
```

10.06309 11.01132 11.84446

Sau khi chạy các lệnh trên, chúng ta sẽ thấy số trung bình của 100 số trung bình mẫu là 11.0, nhưng độ dao động có thể từ 10.06 cm đến 11.84 cm (tức sai số có thể khoảng 1 cm).

(b) Bây giờ chúng ta tăng số cỡ mẫu lên 100 người:

```
sample.100 <- sample(length, 100)
mean(sample.100)
10.85
```

một lần nữa:

```
sample.100 <- sample(length, 100)
mean(sample.100)
11.27
```

Và thử chọn 100 lần, mỗi lần 100 người:

```
B = 100
k = 100
mean = numeric(B)
for (i in 1:B)
{
  sample <- sample(length, k, replace=F)
  mean[i] = mean(sample)
}
quantile(mean, probs=c(0.025, 0.50, 0.975))

      2.5%      50%      97.5%
10.73081 11.00330 11.25547
```

Với cỡ mẫu 100, chúng ta vẫn có sai số khoảng 0.25 đến 0.73 cm, tức là chính xác hơn cỡ mẫu 10 người, nhưng vẫn còn khá thiếu chính xác!

(c) Bây giờ chúng ta tăng số cỡ mẫu lên 400 người:

```
B = 100
k = 400
mean = numeric(B)
for (i in 1:B)
{
  sample <- sample(length, k, replace=F)
  mean[i] = mean(sample)
}
quantile(mean, probs=c(0.025, 0.50, 0.975))
```

```
      2.5%      50%      97.5%
10.84646 11.00394 11.14335
```

Trong các lệnh trên (bạn đọc có thể cắt và dán vào R để chạy) chúng ta chọn 100 lần, mỗi lần 400 người, và cứ mỗi lần chọn mẫu, chúng ta tính số trung bình. Độ dao động số trung bình mẫu bây giờ thấp hơn so với mẫu 100 người: từ 0.14 đến 0.16 cm.

(d) Bây giờ chúng ta tăng số cỡ mẫu lên 1000 người:

```
B = 100
k = 1000
mean = numeric(B)
for (i in 1:B)
{
  sample <- sample(length, k, replace=F)
  mean[i] = mean(sample)
}
quantile(mean, probs=c(0.025, 0.50, 0.975))
```

```
      2.5%      50%      97.5%
10.92698 10.99714 11.09589
```

Sai số bây giờ chỉ 0.1 cm!

(e) Bây giờ chúng ta tăng số cỡ mẫu lên 2000 người:

```
B = 100
k = 2000
mean = numeric(B)
for (i in 1:B)
{
  sample <- sample(length, k, replace=F)
  mean[i] = mean(sample)
}
quantile(mean, probs=c(0.025, 0.50, 0.975))
```

```
      2.5%      50%      97.5%
10.93211 10.99557 11.05812
```

Qua các ví dụ mô phỏng trên, chúng ta dễ dàng thấy sai số giữa cỡ mẫu 1000 và 2000 không khác nhau mấy. Vì thế, vấn đề là chúng ta chấp nhận một sai số chấp nhận được, và công thức trong phần trên của bài viết cung cấp cho chúng ta một cách tính hữu hiệu.

