Lâm sàng thống kê

Đánh giá độ tin cậy của đo lường

Nguyễn Văn Tuấn

*Bạn -- người đang đọc bài này -- thử làm một thí nghiệm nhỏ: tìm vài đồng nghiệp (hay sinh viên), và cứ mỗi "tình nguyện viên", đo huyết áp hai lần, và khoảng cách hai lần có thể là 10 phút. Cố nhiên, bạn chẳng ngạc nhiên nếu thấy huyết áp rất khác nhau giữa các tình nguyện viên. Nhưng điều quan trọng hơn, có lẽ bạn sẽ thấy ở mỗi tình nguyện viên, huyết áp lần thứ nhất không giống như huyết áp đo lần thứ hai. Ở đây chúng ta (bây giờ tôi nói "chúng ta" – bạn và tôi) có hai nguồn dao động: nguồn thứ nhất là độ khác biệt về huyết áp giữa các đối tượng (thuật ngữ thống kê học gọi là **between-subject variation**), và nguồn thứ hai là độ khác biệt ở mỗi đối tượng (còn gọi là **within-subject variation**).*

*Tại sao có sự khác biệt giữa các đối tượng? Lí do chắc có nhiều, chẳng hạn như di truyền, lối sống, độ tuổi, trọng lượng, chiều cao, v.v… Trong bài này tôi sẽ không bàn đến các lí do này. Nhưng câu hỏi đáng quan tâm hơn là tại sao có sự khác biệt trong mỗi đối tượng? Nên nhớ ở đây, chúng ta không có can thiệp gì cả và thời gian đo lường chỉ cách nhau có 10 phút hay ngắn hơn. Vì thế, lí do cho sự khác biệt giữa hai đo lường ở mỗi đối tượng phản ảnh độ tin cậy (thuật ngữ tiếng Anh là **reliability** hay **reproducibility**) của phương tiện đo lường. Vấn đề đặt ra là làm sao đánh giá, hay nói chính xác hơn, là định lượng độ tin cậy? Bài viết này sẽ hướng dẫn các bạn một vài phương pháp để phân tích độ tin cậy của một phương tiện đo lường.*

Đo lường, hay nói nôm na là "cân, đo, đong, đếm", đóng một vai trò cực kì quan trọng trong nghiên cứu khoa học nói chung và y khoa nói riêng. Trước một vấn đề y học, chúng ta cần phải định lượng để biết được qui mô của vấn đề, biết được mối liên hệ giữa các yếu tố trong vấn đề. Nếu không có đo lường và không có thông tin về qui mô cũng như mối liên hệ, cái "khoa học tính" của nghiên cứu sẽ rất thấp. Do đó, có thể nói không ngoa rằng không có đo lường cũng có nghĩa là không có khoa học.

Nhưng bất cứ phương pháp đo lường nào cũng có một số sai sót. Nếu chiều cao của cơ thể hay cân nặng của một đối tượng được đo nhiều lần (cách nhau khoảng vài phút) do một nhân viên y tế, chúng ta sẽ thấy sự dao động về đo lường của đối tượng đó, và đó chính là sai số đo lường (measurement error). Sai số này có thể xuất phát từ những yếu tố khách quan và ngẫu nhiên mà chúng ta không kiểm soát được: chẳng hạn như người đo lường đo hai vị trí khác, hay đối tượng thay đổi thế đứng, hay đơn giản là phương tiện đo lường có sai số, v.v… Tóm lại, đó là những sai số đo lường hoàn toàn ngẫu nhiên. Các phương pháp đo lường sinh hóa như cholesterol, hormones, lượng glucose, v.v… cũng có nhiều sai số, không chỉ do phương pháp assay mà còn do dao động cách phân tích máu.

Do đó, một nhu cầu hết sức quan trọng trước khi nghiên cứu là phải xác định được độ tin cậy của đo lường bao nhiêu. Nếu bệnh viện mới thiết lập một máy quang tuyến, mới lập một hệ thống phân tích sinh hóa, v.v… thì việc đầu tiên là phải làm nghiên cứu để đánh giá độ tin cậy của máy. Không biết độ tin cậy, chúng ta rất dễ dàng chẩn đoán sai, và thậm chí có thể dẫn đến điều trị sai.

I. Một ví dụ cụ thể: IGF-I

Để cụ thể hóa vấn đề, tôi sẽ lấy một ví dụ như sau. Phân tích nồng độ IGF-I (insulin-like growth factor-I) trong máu có thể cho chúng ta biết vận động viên có "ăn gian" (tức có sử dụng các chất hóa học bất hợp pháp) trong thi đấu hay không. Do đó, ước tính độ tin cậy của đo lường rất quan trọng vì nó có thể ảnh hưởng đến sự nghiệp của vận động viên. Tôi và đồng nghiệp đo nồng độ hormone IGF-I ở 20 vận động viên; mỗi người được lấy máu 2 lần, cách nhau khoảng 12 giờ, và đo IGF-I hai lần. Kết quả đo lường có thể trình bày trong bảng số liệu sau đây:

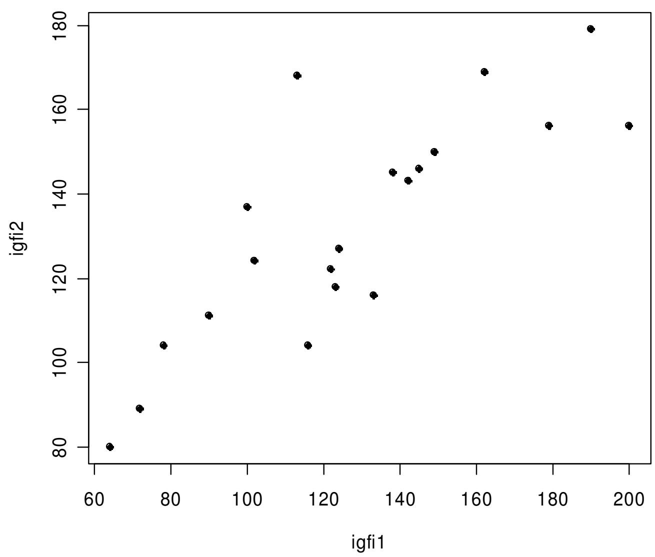
Bảng 1. Nồng độ IGF-I (đo hai lần) ở 20 đối tượng khỏe mạnh

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
179	142	200	145	162	138	190	100	90	102	123	113	124	122	78	64	116	149	133	73
156	143	156	146	169	145	179	137	111	124	118	168	127	122	104	80	104	150	116	89

Số trung bình của lần đo lường thứ nhất là 127.1 và lần thứ hai là 132.2, tức không khác nhau bao nhiêu, nhưng độ lệch chuẩn lần thứ nhất là 37.6 và giảm xuống 27.2 trong lần đo lường thứ 2, một hiện tượng thú vị!

Nếu chỉ nhìn qua số trung bình, chúng ta thấy không có khác nhau đáng kể giữa hai lần đo lường. Nhưng đó có thể là một nhận xét sai lầm, bởi vì không thể lấy số trung bình tổng thể để đánh giá độ tin cậy cho mỗi cá nhân! Đây là loại sai lầm mà người ta hay gọi là ecological fallacy (tôi chưa biết dịch thuật ngữ này như thế nào!)

Có lẽ một cách đánh giá khách quan hơn là chúng ta xem xét độ tương quan giữa hai lần đo lường. Để thấy rõ các số liệu này, chúng ta có thể vẽ biểu đồ tán xạ cho đo lường lần 1 và 2 như sau:



Biểu đồ 1. Tương quan giữa hai đo lường IGF-I ở 20 đối tượng. Hệ số tương quan giữa hai đo lường là $r = 0.817$

Một phương pháp thông thường mà rất nhiều nhà nghiên cứu sử dụng để đánh giá độ tin cậy của đo lường là tính hệ số tương quan (coefficient of correlation) giữa hai lần đo lường. Trong ví dụ trên, hệ số tương quan là 0.817. Chúng ta có thể kết luận gì về hệ số này?

Câu trả lời là khó mà kết luận gì về hệ số này. Thật ra, sử dụng hệ số tương quan để đánh giá độ tin cậy giữa hai đo lường là *sai lầm*, bởi vì hệ số này, như tên gọi, phản ảnh *độ tương quan*, chứ không phải *độ tin cậy*. Hai khái niệm này rất khác nhau, nhưng dễ hiểu lầm. Có thể lấy ví dụ sau đây để chứng minh cho phát biểu đó: giả dụ chúng ta có 5 đối tượng và số liệu đo lường 2 lần như sau:

```
Lần 1: 90, 100, 105, 107, 110
Lần 2: 95, 105, 110, 112, 115
```

Hệ số tương quan là 1. Với một kết quả tuyệt đối như thế, chúng ta có thể kết luận phương pháp đo lường này tuyệt vời không? Câu trả lời là không. Thật ra, phương pháp đo lường này rất tồi, bởi vì hai lần đo lường khác nhau đến 5 đơn vị, và khác nhau một cách nhất quán. Do đó, dù hệ số tương quan là tuyệt đối, nhưng độ tin cậy thì lại rất tồi! **Nói tóm lại, không nên sử dụng hệ số tương quan để đánh giá độ tin cậy của một phương pháp đo lường**.

Các phương pháp phân tích thích hợp cho việc đánh giá độ tin cậy là: giới hạn tương đồng (tức phương pháp Bland-Altman), sai số đo lường chuẩn, hệ số biến thiên cá thể, và hệ số tin cậy. Cách tính các chỉ số này có thể thể hiện trong Bảng 2 sau đây:

Bảng 2. Chi tiết tính toán phương sai và độ khác biệt về đo lường IGF-I

ID (i)	IGFI-1 (x_{i1})	IGFI-2 (x_{i2})	Difference ($d_i = x_{i1} - x_{i2}$)	Mean ($\bar{x}_i$)	Variance (s_i^2)
1	179	156	23	167.5	264.5
2	142	143	-1	142.5	0.5
3	200	156	44	178.0	968.0
4	145	146	-1	145.5	0.5
5	162	169	-7	165.5	24.5
6	138	145	-7	141.5	24.5
7	190	179	11	184.5	60.5
8	100	137	-37	118.5	684.5
9	90	111	-21	100.5	220.5
10	102	124	-22	113.0	242.0
11	123	118	5	120.5	12.5
12	113	168	-55	140.5	512.5
13	124	127	-3	125.5	4.5
14	122	122	0	122.0	0.0
15	78	104	-26	91.0	338.0
16	64	80	-16	72.0	128.0
17	116	104	12	110.0	72.0
18	149	150	-1	149.5	0.5
19	133	116	17	124.5	144.5
20	73	89	-16	81.0	128.0
Mean	127.1	132.2	**-5.1**	**129.7**	**242.35**
SD	37.6	27.2	**21.9**	**30.9**	

Quay lại vấn đề đánh giá đo lường IGF-I. Để dễ dàng hiểu và phân tích, tôi sẽ trình bày bảng số liệu trên theo hàng (thay vì theo cột) như dưới đây:

- Cột 1 chỉ là mã số cho từng đối tượng;
- Cột 2 là IGF-I đo lường lần thứ nhất, kí hiệu x_{i1} (trong đó i = 1, 2, 3, …, 20);
- Cột 3 là IGF-I đo lường lần thứ hai, kí hiệu x_{i2};

- Cột 4 là độ khác biệt giữa hai đo lường, kí hiệu d_i. Dòng 21 ("mean") là số trung bình của tất cả d_i, và dòng 22 ("SD") là độ lệch chuẩn của d_i:

- Cột 5 là số trung bình của hai lần đo lường. Chẳng hạn như với đối tượng 1, số trung bình chỉ đơn giản là (179 + 156)/2 = 167.5. Dòng 21 là số trung bình của tất cả trung bình, và dòng 22 là độ lệch chuẩn của tất cả số trung bình;

- Cột 6 là phương sai (variance) của hai lần đo lường. (Nếu bạn đọc chưa hiểu phương sai là gì, có thể tìm đọc bài Lâm sàng thống kê số 1 của tôi). Ở đây, vì có 2 đo lường, cho nên phương sai (kí hiệu là s_i^2) cũng đơn giản bằng bình phương của độ khác biệt giữa 2 đo lường chia cho 2. Chẳng hạn như với đối tượng 1, phương sai là $s_1^2 = \frac{(179-156)^2}{2} = 264.5$. Dòng 21 là số trung bình của tất cả s_i^2. Chú ý, tôi không tính độ lệch chuẩn của s_i^2 vì chỉ số này không có ý nghĩa gì.

Bay giờ, chúng ta có thể sử dụng các tính toán trên để đánh giá độ tin cậy của phương pháp đo lường.

II. Giới hạn tương đồng (Limit of agreement hay LoA)

Một phương pháp đánh giá độ tin cậy đơn giản nhất (và dễ hiểu nhất) có tên là "**Limit of Agreement**" mà tôi tạm dịch là "giới hạn tương đồng" do Martin Bland và Douglas Altman đề suất từ thập niên 1980s. Theo phương pháp này, chúng ta tiến hành 2 bước:

- Bước 1: tính độ khác biệt giữa hai đo lường cho từng đối tượng và gọi là d_i (tức cột 6 trong **Bảng 1** trên);

- Bước 2: tính số trung bình, độ lệch chuẩn, và khoảng tin cậy của d_i. Khoảng tin cậy 95% của d_i chính là LoA.

Trong ví dụ trên, chúng ta thấy (Bảng 1, cột 4, hàng 21 và 22) độ khác biệt trung bình giữa hai lần đo lường là -5.1 ug/L, và độ lệch chuẩn là 21.9 ug/L. Nói cách khác, khoảng tin cậy 95% (hay LoA) là:

LoA = -5.1 ± 1.96 × 21.9 = −48.0 đến +37.8 ug/L

Chỉ số LoA cho chúng ta biết nếu chúng ta đo IGF-I hai lần ở một đối tượng thì đo lường lần thứ nhất có thể thấp hơn lần thứ hai là 48 ug/L, nhưng cũng có thể cao hơn đến 37.8 ug/L.

III. Sai số đo lường chuẩn (standard error of measurement – SEM)

Chúng ta có thể dễ dàng thấy rằng phương sai ở mỗi đối tượng (cột 6) phản ảnh độ dao động, và do đó, độ tin cậy của đo lường. Độ dao động càng cao, độ tin cậy càng thấp. Và ngược lại, độ dao động thấp có nghĩa là độ tin cậy cao. Nhưng chúng ta có 20 đối tượng, nên cách tóm lược độ tin cậy tốt nhất là lấy trung bình của phương sai của 20 đối tượng. Gọi số trung bình phương sai là s^2, chúng ta có:

$$s^2 = \frac{s_1^2 + s_2^2 + s_3^2 + ... + s_{20}^2}{20} = \frac{264.5 + 0.5 + 968 + ... + 128}{20} = 242.35$$

Số này được trình bày ở hàng số 21 ("mean"). Nhưng đơn vị của phương sai là bình phương, cho nên chúng ta cần phải hoán chuyển về đơn vị gốc bằng cách lấy căn bậc hai:

$$s = \sqrt{242.35} = 15.6 \text{ ug/L.}$$

Đây chính là **sai số đo lường chuẩn** (SEM). Ý nghĩa của SEM là: nếu chúng ta đo nồng độ IGF-I ở **một** đối tượng nhiều lần (giả dụ như 100 lần), thì chúng ta kì vọng rằng 68% các đo lường IGF-I của đối tượng đó có thể cao hơn hay thấp độ trung bình của đối tượng khoảng 15.6 ug/L. Cũng có thể nói rằng 95% các đo lường IGF-I của đối tượng đó có thể cao hơn hay thấp độ trung bình của đối tượng khoảng 31 ug/L (tức lấy 15.6 nhân cho 1.96, vì 1.96 là -- nếu các bạn còn nhớ -- số z của phân phối chuẩn).

IV. Hệ số biến thiên cá thể (within-subject coefficient of variation)

Chú ý cột 5 trong **Bảng 2** là số trung bình cho từng đối tượng. Do đó, số trung bình cho toàn bộ 20 đối tượng chỉ đơn giản là trung bình của 20 số trung bình! Gọi số trung bình tổng thể là m, chúng ta có m = 129.7 ug/L (như dòng 21 của Bảng 1).

Hệ số biến thiên cá thể (wCV) được ước tính bằng cách lấy sai số chuẩn đo lường chia cho số trung bình tổng thể. Nói cách khác:

$$wCV = \frac{s}{m} \times 100$$

Trong ví dụ trên, $wCV = 15.5 / 129.7 \times 100 = 11.9$ (hay 12%). Nói cách khác, thay vì mô tả độ tin cậy bằng SEM, chúng ta mô tả bằng phần trăm qua wCV.

Ý nghĩa của wCV cũng tương tự như ý nghĩa của SEM, nhưng tương đối (thay vì tuyệt đối): nếu chúng ta đo nồng độ IGF-I ở **một** đối tượng 100 lần, thì kết quả này cho biết có khoảng 68% các giá trị IGF-I của đối tượng đó dao động trên dưới 12% so với số trung bình. Cũng có thể phát biểu (hay suy luận) rằng xác suất 95% là các giá trị IGF-I của đối tượng đó dao động khoảng 24% trên dưới giá trị trung bình.

V. Hệ số tin cậy

Trong phần đầu, tôi có nhận xét rằng ứng dụng hệ số tương quan để đánh giá độ tin cậy đo lường là không thích hợp, vì hệ số tương quan bỏ qua độ khác biệt giữa hai đo lường. Trong phân tích độ tin cậy, có một phương pháp tính toán khác rất thích hợp cho việc đánh giá độ tin cậy nhưng ít khi nào được giới thiệu trong sách giáo khoa: đó là hệ số tin cậy (coefficient of reliability), sẽ viết tắt là *R* (chú ý khác với *r* thường sử dụng cho hệ số tương quan).

Hệ số tin cậy được phát triển từ lí thuyết đo lường (theory of measurement), mà tôi không thể mô tả trong bài viết vì nằm ngoài phạm vi. Tuy nhiên, hệ số tin cậy R được phát triển từ các cơ sở lí thuyết sau đây:

Gọi *X* là đo lường cho một cá nhân, lí thuyết đo lường phát biểu rằng *X* có hai phần: phần giá trị thật (true value) và phần nhiễu (random error). Chúng ta sẽ kí hiệu hai giá trị này là *T* và *E*. Nói cách khác, mô hình này phát biểu rằng:

$$X = T + E$$

Nên nhớ rằng, chúng ta không biết (hay không quan sát) được giá trị thật *T* của đối tượng bao nhiêu, nhưng chúng ta chỉ biết *X,* tức là giá trị quan sát hay đo lường được, và *E* là sai số mà chúng ta tính được. Nói cụ thể hơn, trong ví dụ trên, chúng ta không biết IGF-I thật sự của từng đối tượng là bao nhiêu, nhưng chúng ta biết được giá trị qua từng lần đo lường, và qua nhiều lần đo lường, chúng ta biết được *E.*

Giả định rằng *T* và *E* hoàn toàn độc lập (tức không có mối tương quan gì với nhau), và tuân theo luật phân phối chuẩn, thì phương sai của *X* là tổng phương sai của *T* và *E.* Nếu kí hiệu phương sai là σ^2, chúng ta có:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \qquad [1]$$

Từ đó, hệ số tin cậy R được định nghĩa như sau:

$$R = \frac{\sigma_T^2}{\sigma_X^2}$$

Hay cũng có thể viết $R = \sigma_T^2 / \left(\sigma_T^2 + \sigma_E^2\right)$, hay $R = 1 - \sigma_E^2 / \left(\sigma_T^2 + \sigma_E^2\right)$. Nhìn qua công thức trên, chúng ta thấy R chính là tỉ số phương sai của giá trị thật và phương sai của giá trị quan sát. Như vậy, nếu *E* rất nhỏ so với *X,* thì *T* rất lớn, và đó là tín hiệu cho thấy phương pháp đo lường có độ tin cậy cao.

Bởi vì phương sai của *T* không thể cao hơn phương sai của *X,* cho nên giá trị của R phải dao động từ 0 đến 1. Nếu R = 0 thì điều này có nghĩa là phương pháp đo lường vô dụng vì hoàn toàn không có thể tin cậy được; nếu R = 1 thì phương pháp đo lường hoàn hảo. Nhưng trong thực tế, không có phương pháp đo lường nào tuyệt vời với R = 1, cho nên -- tùy theo trường hợp -- chúng ta phải chấp nhận một phương pháp đo lường với giá trị R thấp hơn 1. Trong loãng xương, máy DXA thường có giá trị R là 0.98 hay 0.97, nhưng với các chỉ số sinh hóa, R thấp hơn nhiều (có thể 0.8 thậm chí 0.7).

Phương pháp ước tính R tương đối phức tạp (nhất là khi có nhiều đo lường cho một đối tượng), nhưng nguyên tắc vẫn là phân chi tổng số biến thiên (variation) thành 2 nguồn: between-subject variation within-subject variation mà tôi đã đề cập đến trong phần đầu. Trong trường hợp chúng ta có *N* đối tượng, và mỗi đối tượng được đo hai lần (như ví dụ 1), chúng ta có thể tính toán số trung bình và phương sai cho từng đối tượng như **Bảng 2**. Gọi $\bar{x}_i$ là số trung bình và s_i^2 là phương sai cho từng đối tượng, và gọi $\bar{x}$ là số trung bình tổng thể (tức là trung bình của $\bar{x}_i$), chúng ta có thể ước tính vài thông số như sau:

- Tổng bình phương của đo lường *giữa các đối tượng* (kí hiệu *BSS*):

$$BSS = 2\left(\bar{x}_1 - \bar{x}\right)^2 + 2\left(\bar{x}_2 - \bar{x}\right)^2 + 2\left(\bar{x}_3 - \bar{x}\right)^2 + \ldots + 2\left(\bar{x}_N - \bar{x}\right)^2$$

Chú ý số 2 ở đây là hai lần đo lường. Thật ra, công thức trên có thể viết ngắn gọn (nếu bạn nào thích toán và nhớ toán!): $BSS = \sum_{i=1}^{N} 2\left(\bar{x}_i - \bar{x}\right)^2$

- Bởi vì *BSS* tính từ *N* đối tượng và cần 1 thông số để tính, cho nên chúng ta phải chia *BSS* cho N-1 để có "mean square" (trung bình bình phương) *BMS*.

$$BMS = \frac{BSS}{N-1}$$

- Tổng bình phương của đo lường *ở mỗi đối tượng* (kí hiệu *WSS*):

$$WSS = s_1^2 + s_2^2 + s_3^2 + ... + s_N^2$$

Hay đơn giản hơn: $WSS = \sum_{i=1}^{N}(k_i - 1)s_i^2$, trong đó k_i là số lần đo lường cho đối tượng *i* (trong ví dụ IGF-I, k_i là một hằng số bằng 2 – 2 lần đo lường cho mỗi đối tượng).

- Chú ý rằng mỗi s_i^2 trong *WSS* chúng ta phải "tiêu" một thông số (trung bình cho từng đối tượng). Trong ví dụ trên, chúng ta có *2N* đo lường (mỗi đối tượng đo 2 lần), nhưng chúng ta phải chi tiêu mất *N* thông số, nên bậc tự do (degree of freedom) là *2N – N = N*. Do đó, chúng ta phải điều chỉnh *WSS* bằng cách chia *WSS* cho bậc tự do và kết quả là *WMS*:

$$WMS = \frac{WSS}{N}$$

Từ các tính toán trên, bây giờ chúng ta có thể ước tính σ_T^2 và σ_E^2 trong mô hình [1] trên đây. Nên nhớ rằng chúng ta không biết σ_T^2 và σ_E^2, mà chỉ ước tính thôi. Vì thế, thay vì viết σ_T^2 và σ_E^2, chúng ta dùng kí hiệu s_T^2 và s_E^2 để chỉ ước số của hai thông số đó. Theo lí thuyết thống kê, chúng ta có:

$$s_T^2 = \frac{BMS - WMS}{k}$$

Và:

$$s_E^2 = WMS$$

Ở đây, *k* là số lần đo cho từng đối tượng (tức là 2 trong ví dụ trên). Do đó, hệ số tin cậy là:

$$R = \frac{s_T^2}{s_T^2 + s_E^2}$$

Bạn đọc có thể tự mình tính toán R qua các công thứ trên. Ở đây, tôi chỉ cung cấp các ước số (hay đáp số!) như sau: s_T^2 = 788.3, s_E^2 = 242.3, và R = 0.76.

R = 0.76 có ý nghĩa gì ? Tôi sẽ quay lại câu hỏi này trong một bài sau, vì đây là vấn đề rất quan trọng. Nhưng ở đây, có thể nói ngay rằng vì hệ số tin cậy dưới 0.8, cho nên có thể nói rằng độ tin cậy của phương pháp đo lường IGF-I chỉ trung bình. Sự thật này có ảnh hưởng lớn đến việc sử dụng IGF-I cho việc truy tìm những vận động viên thiếu thành thật (hay nói trắng ra là "ăn gian").

Nói tóm lại, bài này đã trình bày một số phương pháp phân tích độ tin cậy đo lường, một khía cạnh cực kì quan trọng trong nghiên cứu và thực hành lâm sàng. Đo lường không đáng tin cậy hay có độ tin cậy thấp có thể dẫn đến nhiều hệ quả nghiêm trọng như chẩn đoán sai và điều trị không cần thiết. Nhưng rất tiếc, vấn đề này chưa được quan tâm đúng mức (có thể bác sĩ chưa hiểu hay quá tin vào máy móc) cho nên trong thực tế đã xảy ra nhiều trường hợp cười ra nước mắt và tranh cãi không cần thiết.

Vấn đề quan trọng là sau khi biết được độ tin cậy của phương pháp đo lường, bước kế tiếp là sử dụng các thông tin này cho việc thực hành lâm sàng và nghiên cứu. Nhưng tôi sẽ quay lại đề tài này trong một bài tiếp. Còn bây giờ, bạn đọc có thể tự mình làm vài tính toán cho quen dần với các khái niệm vừa giới thiệu trên.

Chú thích kĩ thuật:

Mắc dù các phương pháp phân tích độ tin cậy đo lường có thể tính toán thủ công bằng Excel, nhưng với một chương trình thống kê thì hữu hiệu hơn. Các các mã R sau đây được sử dụng trong bài này:

```
# nhập số liệu cho hai lần đo lường, gọi hai biến là igfi1 và igfi2

igfi1 <- c(179, 142, 200, 145, 162, 138, 190, 100, 90, 102, 123, 113,
           124, 122, 78, 64, 116, 149, 133, 72)

igfi2 <- c(156, 143, 156, 146, 169, 145, 179, 137, 111, 124, 118, 168,
           127, 122, 104, 80, 104, 150, 116, 89)

# vẽ biểu đồ 1, chú ý pch=16 là kí hiệu cho các dấu tròn đen

plot(igfi1, igfi2, pch=16)

# Ước tính độ khác biệt giữa hai đo lường cho từng đối tượng

d <- igfi1 - igfi2

# tính trung bình và độ lệch chuẩn của khác biệt, gọi là md và sd

md <- mean(d)
sd <- sd(d)

# tính khoảng tin cậy 95% của d, tức là LoA

loa.025 <- md-1.96*sd
loa.975 <- md+1.96*sd

# báo cáo kết quả LoA

c(md, sd, loa.025, loa.975)

# Ước tính phương sai cho từng đối tượng

v <- (igfi1-igfi2)^2/2

mean.v <- mean(v)

sem <- sqrt(mean.v)

c(mean.v, sem)

# Ước tính trung bình cho từng đối tượng và tổng trung bình
```

```
m <- (igfi1 + igfi2)/2

overall.mean <- mean(m)

# Ước tính wCV

wcv <- sem / overall.mean

# báo cáo kết quả

c(overall.mean, sem, wcv)
```

 # Các mã sau đây sử dụng để tính hệ số tin cậy R

```
igfi1 <- c(179, 142, 200, 145, 162, 138, 190, 100, 90, 102, 123, 113,
           124, 122, 78, 64, 116, 149, 133, 72)

igfi2 <- c(156, 143, 156, 146, 169, 145, 179, 137, 111, 124, 118, 168,
           127, 122, 104, 80, 104, 150, 116, 89)

k <- 2
n <- length(igfi1)
meani <- (igfi1 + igfi2)/k
overall.mean <- mean(meani)

bssi <- k*(meani - overall.mean)^2
bss <- sum(bssi)
bms <- bss/n

wssi <- (igfi1-igfi2)^2/2
wss <- sum(wssi)
wms <- wss/(k*n - n)

var.t <- (bms-wms)/k
var.e <- wms
r <- var.t / (var.t + var.e)
```

 # báo cáo kết quả

```
c(var.t, var.e, r)
```