

TÍNH TOÁN BÁCH PHÂN VỊ VÀ Z-SCORE CHO CHIỀU CAO THEO TUỔI, CÂN NẶNG THEO TUỔI VÀ BMI THEO TUỔI

Phương pháp được sử dụng để xây dựng các **chuẩn tham chiếu WHO năm 2007** dựa trên mô hình **GAMLSS** với phân phối lũy thừa Box–Cox (*Box–Cox power exponential distribution*) (Rigby và Stasinopoulos, 2004). Tuy nhiên, các mô hình được lựa chọn cuối cùng được đơn giản hóa thành **mô hình LMS** (Cole và Green, 1992), vì không có chuẩn tham chiếu nào yêu cầu hiệu chỉnh **độ nhọn của phân phối (kurtosis)**.

Do đó, việc tính toán **bách phân vị (percentiles)** và **z-score** cho cả ba chỉ số đều sử dụng các công thức dựa trên phương pháp LMS.

Tuy nhiên, một giới hạn được áp dụng cho tất cả các chỉ số, theo đó chỉ cho phép xác định bách phân vị trong khoảng tương ứng với **z-score từ -3 đến +3**. Cơ sở của giới hạn này là các bách phân vị nằm ngoài ± 3 SD không thay đổi đáng kể khi thay đổi các z-score tương đương. Tác động mất mát do giới hạn này là nhỏ, vì khoảng được đưa vào phân tích tương ứng với **bách phân vị từ 0,135 đến 99,865**.

Đối với tất cả các chỉ số, các giá trị ước lượng đã được lập bảng của **tham số lũy thừa Box–Cox, trung vị và hệ số biến thiên**, tương ứng với tuổi (hoặc chiều cao) tt, lần lượt được ký hiệu là:

- $L(t)$: tham số Box–Cox;

- $M(t)M(t)$: trung vị;
- $S(t)S(t)$: hệ số biến thiên.

1. BÁCH PHÂN VỊ VÀ Z-SCORE ĐỐI VỚI CHIỀU CAO THEO TUỔI

Đối với chỉ số này, $L(t)L(t)$ bằng 1, do đó phân phối chuẩn Box–Cox được sử dụng trong phương pháp LMS được đơn giản hóa thành **phân phối chuẩn**.

Vì vậy, khoảng cách giữa các độ lệch chuẩn liên kề (ví dụ giữa 2 SD và 3 SD) là **không đổi ở một độ tuổi cụ thể**, nhưng thay đổi giữa các độ tuổi khác nhau.

Trong trường hợp này, bách phân vị tại tuổi t có thể được ước tính theo công thức:

$$C_{100\alpha}(t) = M(t)[1 + L(t)S(t)Z_{\alpha}]^{1/L(t)} C_{\{100\alpha\}}(t) = M(t)\left[1 + L(t)S(t)Z_{\alpha}\right]^{1/L(t)}$$

Vì $L(t)=1$, công thức được đơn giản hóa thành:

$$C_{100\alpha}(t) = M(t)[1 + S(t)Z_{\alpha}] C_{\{100\alpha\}}(t) = M(t)[1 + S(t)Z_{\alpha}]$$

hay:

$$C_{100\alpha}(t) = M(t) + StDev(t)Z_{\alpha}, -3 \leq Z_{\alpha} \leq 3 \\ C_{\{100\alpha\}}(t) = M(t) + StDev(t)Z_{\alpha}, \quad -3 \leq Z_{\alpha} \leq 3$$

Trong đó:

- Z_{α} là **độ lệch chuẩn tương đương của phân phối chuẩn (normal equivalent deviate)** đối với phần đuôi có diện tích α ;
- $C_{100\alpha}(t)$ là **bách phân vị thứ 100α** tại tuổi t ;
- $StDev(t)$ là **độ lệch chuẩn tại tuổi t** , được xác định bằng cách nhân $S(t)$ với $M(t)$.

Tính z-score cá thể

Z-score của một cá thể có giá trị đo y tại tuổi t được tính như sau:

$$z_{ind} = \frac{\left[\frac{y}{M(t)} \right]^{L(t)} - 1}{S(t)L(t)}$$

Đối với chiều cao theo tuổi, do $L(t)=1$, công thức trở thành:

$$z_{ind} = \frac{y - M(t)}{StDev(t)}$$

2. BÁCH PHÂN VỊ VÀ Z-SCORE ĐỐI VỚI CÂN NẶNG THEO TUỔI VÀ BMI THEO TUỔI

Các chỉ số dựa trên cân nặng có **phân phối lệch phải (right-skewed distribution)**.

Khi được mô hình hóa một cách phù hợp, độ lệch phải của dữ liệu có tác động làm cho **khoảng cách giữa các z-score dương tăng dần khi càng xa trung vị**, trong khi **khoảng cách giữa các z-score âm giảm dần**.

Phương pháp LMS mô hình hóa khá phù hợp dữ liệu lệch bằng cách sử dụng **phân phối chuẩn Box–Cox (Box–Cox normal distribution)**, phân phối này bám khá sát dữ liệu thực nghiệm.

Tuy nhiên, hạn chế của phương pháp là **hai phần đuôi ngoài của phân phối chịu ảnh hưởng rất lớn bởi các giá trị cực đoan**, ngay cả khi số lượng các giá trị này rất ít (ví dụ dưới 1%).

Theo cùng phương pháp được áp dụng cho **Bộ chuẩn tăng trưởng trẻ em của WHO**, một phiên bản giới hạn của phương pháp LMS được sử dụng cho các chỉ số dựa trên cân nặng trong **chuẩn WHO năm 2007**.

Theo đó, phân phối chuẩn Box–Cox được giới hạn trong khoảng tương ứng với các z-score mà dữ liệu thực nghiệm có sẵn, tức là **từ -3 SD đến $+3$ SD**.

Ngoài các giới hạn này, độ lệch chuẩn tại mỗi độ tuổi được cố định theo khoảng cách giữa:

- $+2SD+2$ SD và $+3SD+3$ SD;
- $-2SD-2$ SD và $-3SD-3$ SD.

Cách tiếp cận này tránh phải đưa ra các giả định về phân phối của dữ liệu nằm ngoài giới hạn của các giá trị quan sát được.

Do sự điều chỉnh này, phân phối z-score có thể **hơi lệch khỏi phân phối chuẩn ở các phần đuôi cực trị** (ngoài ± 3 SD), mặc dù tác động thực tế dự kiến là **tối thiểu**.

2.1. Tính bách phân vị

Bách phân vị được tính theo:

$$C_{100\alpha}(t) = M(t) [1 + L(t) S(t) Z_{\alpha}]^{1/L(t)}, -3 \leq Z_{\alpha} \leq 3 \\ C_{\{100\alpha\}}(t) = M(t) \left[1 + L(t) S(t) Z_{\alpha} \right]^{1/L(t)}, \quad -3 \leq Z_{\alpha} \leq 3$$

2.2. Quy trình tính z-score cho từng trẻ

Quy trình sau đây được khuyến nghị để tính z-score cho một trẻ có giá trị đo yy tại tuổi tt.

Bước 1. Tính z-score LMS ban đầu

$$z_{ind} = \left[\frac{y - M(t)}{S(t)L(t)} \right]^{1/L(t)} - \frac{1}{S(t)L(t)}$$

Bước 2. Tính z-score cuối cùng

Z-score cuối cùng của trẻ đối với chỉ số đó, ký hiệu z_{ind}^* , được tính như sau:

$$z_{ind}^* = \begin{cases} z_{ind}, & \text{nếu } z_{ind} \leq 3 + \frac{y - SD_{3pos}}{SD_{23pos}} \\ z_{ind} - \frac{3 + \frac{y - SD_{3pos}}{SD_{23pos}} - z_{ind}}{3}, & \text{nếu } z_{ind} > 3 + \frac{y - SD_{3pos}}{SD_{23pos}} \\ z_{ind}, & \text{nếu } z_{ind} \leq -3 - \frac{y - SD_{3neg}}{SD_{23neg}} \\ z_{ind} + \frac{-3 - \frac{y - SD_{3neg}}{SD_{23neg}} - z_{ind}}{3}, & \text{nếu } z_{ind} < -3 - \frac{y - SD_{3neg}}{SD_{23neg}} \end{cases}$$

$$z_{ind} > 3 + \frac{y - SD3_{neg}}{SD23_{neg}}, \text{ \& \textit{nếu}} \\ z_{ind} < -3 \end{cases}$$

Nói cách khác:

- Nếu z-score LMS nằm trong khoảng **−3 đến +3**, sử dụng trực tiếp z-score LMS.
- Nếu z-score $> +3$, z-score cuối cùng được hiệu chỉnh dựa trên khoảng cách của giá trị đo y so với ngưỡng $+3$ SD.
- Nếu z-score < -3 , z-score cuối cùng được hiệu chỉnh dựa trên khoảng cách của giá trị đo y so với ngưỡng -3 SD.

3. GIẢI THÍCH CÁC GIÁ TRỊ SD SỬ DỤNG TRONG HIỆU CHỈNH

Trong đó:

SD3posSD3_{pos}

SD3posSD3_{pos} là **giá trị ngưỡng +3 SD** được tính tại tuổi tt bằng phương pháp LMS:

$$SD3pos = M(t) [1 + L(t) \times S(t) \times 3]^{1/L(t)} SD3_{pos} = M(t) \left[1 + L(t) \times S(t) \times 3 \right]^{1/L(t)}$$

SD3negSD3_{neg}

$SD3_{neg}$ là **giá trị ngưỡng -3 SD** được tính tại tuổi tt bằng phương pháp LMS:

$$SD3_{neg} = M(t) [1 + L(t) \times S(t) \times (-3)]^{1/L(t)} = M(t) \left[1 + L(t) \times S(t) \times (-3) \right]^{1/L(t)}$$

$SD23_{pos}$

$SD23_{pos}$ là **khoảng cách giữa ngưỡng $+3$ SD và $+2$ SD**, được tính tại tuổi tt bằng phương pháp LMS:

$$SD23_{pos} = M(t) [1 + L(t) S(t) (3)]^{1/L(t)} - M(t) [1 + L(t) S(t) (2)]^{1/L(t)} = M(t) \left[1 + L(t) S(t) (3) \right]^{1/L(t)} - M(t) \left[1 + L(t) S(t) (2) \right]^{1/L(t)}$$

$SD23_{neg}$

$SD23_{neg}$ là **khoảng cách giữa ngưỡng -2 SD và -3 SD**, được tính tại tuổi tt bằng phương pháp LMS:

$$SD23_{neg} = M(t) [1 + L(t) S(t) (-2)]^{1/L(t)} - M(t) [1 + L(t) S(t) (-3)]^{1/L(t)} = M(t) \left[1 + L(t) S(t) (-2) \right]^{1/L(t)} - M(t) \left[1 + L(t) S(t) (-3) \right]^{1/L(t)}$$

Các giá trị trên được sử dụng để kéo dài z-score một cách tuyến tính ở ngoài khoảng ± 3 SD, thay vì giả định một dạng phân phối chưa được dữ liệu thực nghiệm hỗ trợ.

4. VÍ DỤ MINH HỌA: BMI THEO TUỔI Ở TRẺ TRAI

Để minh họa quy trình trên, tài liệu đưa ra ví dụ về **BMI theo tuổi ở trẻ trai**, được trình bày trong Hình 1.

Chỉ số khối cơ thể (BMI), kg/m²

Hình minh họa thể hiện các trẻ/vị thành niên được phân loại theo **chuẩn tham chiếu BMI theo tuổi của WHO năm 2007**.

Trẻ 1

Trẻ trai 11 tuổi, BMI = 30 kg/m².

Các tham số LMS:

$L = -1.7862$ $M = 16.9392$ $S = 0.11070$

Bước 1: Tính z-score LMS

$$z_{ind} = \frac{\left[\frac{30.0}{16.9392} \right] - (-1.7862)}{0.11070 \times (-1.7862)}$$

Kết quả:

$$z_{ind} = 3.24$$

Do:

$$3.24 > 33.24 > 3$$

nên cần áp dụng bước hiệu chỉnh z-score cuối cùng.

Bước 2: Tính ngưỡng +3 SD

$$SD3_{pos} = 16.9392 [1 + (-1.7862) \times 0.11070 \times 3]^{-1/1.7862} = 16.9392 \left[1 + (-1.7862) \times 0.11070 \times 3 \right]^{-1/1.7862}$$

Kết quả:

$$SD3_{pos} = 28.03 \quad SD3_{pos} = 28.03$$

Tính ngưỡng +2 SD

$$SD2_{pos} = 16.9392 [1 + (-1.7862) \times 0.11070 \times 2]^{-1/1.7862} = 16.9392 \left[1 + (-1.7862) \times 0.11070 \times 2 \right]^{-1/1.7862}$$

Kết quả:

$$SD2_{pos} = 22.45 \quad SD2_{pos} = 22.45$$

Do đó:

$$SD23_{pos} = 28.03 - 22.45 = 5.58 \quad SD23_{pos} = 28.03 - 22.45 = 5.58$$

Tính z-score cuối cùng

$$z_{ind} = 3 + (30 - 28.03558) z_{ind}^* = 3 + \left(\frac{30 - 28.03}{5.58} \right)$$

$$z_{ind} = 3.35 \boxed{z_{ind}^* = 3.35}$$

Như vậy, z-score LMS ban đầu là **3,24**, nhưng z-score cuối cùng sau khi áp dụng phương pháp hiệu chỉnh ngoài giới hạn ± 3 SD là **3,35**.

5. TRẺ 2

Trẻ trai 16 tuổi, BMI = 14 kg/m².

Các tham số LMS:

$$L = -1.3529 \quad L = -1.3529 \quad M = 20.4951 \quad M = 20.4951 \quad S = 0.12579 \quad S = 0.12579$$

Bước 1: Tính z-score LMS

$$z_{ind} = \frac{[14.0 - 20.4951] - (-1.3529) \times (-1.3529)}{0.12579 \times (-1.3529)}$$

Kết quả:

$$z_{ind} = -3.96 \quad z_{ind} = -3.96$$

Do:

$$-3.96 < -3 \quad -3.96 < -3$$

nên cần áp dụng hiệu chỉnh cho z-score nằm ngoài giới hạn -3 SD.

Tính ngưỡng -2 SD

$$SD2_{neg} = 20.4951 [1 + (-1.3529) \times 0.12579 \times (-2)]^{-1/1.3529} = 20.4951 \left[1 + (-1.3529) \times 0.12579 \times (-2) \right]^{-1/1.3529}$$

Kết quả:

$$SD2_{neg} = 16.50$$

Tính ngưỡng -3 SD

$$SD3_{neg} = 20.4951 [1 + (-1.3529) \times 0.12579 \times (-3)]^{-1/1.3529} = 20.4951 \left[1 + (-1.3529) \times 0.12579 \times (-3) \right]^{-1/1.3529}$$

Kết quả:

$$SD3_{neg} = 15.11$$

Do đó:

$$SD23_{neg} = 16.50 - 15.11 = 1.39$$

Tính z-score cuối cùng

$$z_{ind}^* = -3 + (14.0 - 15.11 / 1.39) = -3 + \left(\frac{14.0 - 15.11}{1.39} \right)$$

Kết quả:

$$z_{ind}^* = -3.80$$

Như vậy, z-score LMS ban đầu là **-3,96**, sau khi hiệu chỉnh theo phương pháp giới hạn ngoài -3 SD, z-score cuối cùng là **-3,80**.

6. TRẺ 3

Trẻ trai 9 tuổi, BMI = 19 kg/m².

Các tham số LMS:

$$L = -1.6318 \quad L = -1.6318 \quad M = 16.0490 \quad M = 16.0490 \quad S = 0.10038 \quad S = 0.10038$$

Tính z-score LMS

$$z_{ind} = \frac{\left[\frac{19}{16.0490} - 1.6318 \right] \times (-1.6318)}{0.10038 \times (-1.6318)}$$

Kết quả:

$$z_{ind} = 1.47 \quad \boxed{z_{ind} = 1.47}$$

Vì:

$$-3 \leq 1.47 \leq 3 \quad -3 \leq 1.47 \leq 3$$

nên z-score này nằm trong khoảng cho phép của phương pháp LMS và **không cần hiệu chỉnh**.

Do đó:

$$z_{ind} = 1.47 \boxed{z^*_{ind} = 1.47}$$

Tài liệu ghi nhận trường hợp này là **z-score LMS nằm trong khoảng từ -3 đến +3**.

7. TÀI LIỆU THAM KHẢO

Cole TJ, Green PJ (1992). *Smoothing reference centile curves: the LMS method and penalized likelihood.* Statistics in Medicine, 11:1305–1319.

Rigby RA, Stasinopoulos DM (2004). *Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution.* Statistics in Medicine, 23:3053–3076.

WHO Multicentre Growth Reference Study Group (2006). *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.* Geneva: World Health Organization; 312 trang.

Một lưu ý quan trọng cho anh

Tài liệu này đặc biệt hữu ích nếu anh đang làm **chuẩn hóa thuật toán tính BMI-for-age/weight-for-age theo WHO 2007**. Điểm mấu chốt không chỉ là công thức LMS ban đầu:

$$z_{\text{ind}} = (y/M)^{L-1} L S z_{\text{ind}} = \frac{(y/M)^{L-1}}{L S}$$

mà còn là **cách WHO xử lý các trường hợp $z < -3$ hoặc $z > +3$** bằng các khoảng $SD_{23\text{neg}}$ và $SD_{23\text{pos}}$. Đây là phần rất dễ bị bỏ qua nếu chỉ dùng công thức LMS cơ bản.