

# SỰ THAY THẾ GIÁ TRỊ KHUYẾT TRONG HỒI QUY VỮNG: XEM XÉT ẢNH HƯỞNG CỦA CÁC HÀM QUYỀN SỐ VÀ THƯỚC ĐO SỰ HỘI TỤ

*NORO, Tatsuo và WADA, Kazumi*

*Bộ phận nghiên cứu, Phòng Công nghệ thông tin Thống kê,*

*Trung tâm thông tin Thống kê (NSTAC), Nhật Bản*

## *Tóm tắt*

Các giá trị khuyết có thể được thay thế, nếu chúng có nguyên nhân từ sự chệch. Bài báo viết về việc thay thế giá trị khuyết trong quá trình hồi quy sẽ tập trung vào sự tồn tại của các giá trị chệch. Trong quá trình thay thế dữ liệu khuyết, các giá trị chệch không cần thiết sẽ bị coi là giá trị sai, nhưng chúng có thể được loại bỏ khỏi mô hình. Nếu xuất hiện một vài giá trị chệch, thì chúng có thể làm biến dạng sự suy diễn của phương pháp ước lượng bình phương nhỏ nhất. Do đó phương pháp Robust chính là giải pháp hiệu quả với các giá trị chệch này. UNSC và UNECE đã cùng nhau xuất bản hàng loạt các ấn phẩm Biên tập dữ liệu thống kê (Statistical Data Editing-SDE) nhằm chia sẻ những hiểu biết thực tế giữa các cơ quan thống kê quốc gia với nhau. Trong phần hai, kĩ thuật Tukey's EDA (Phân tích các dữ liệu được khám phá) sẽ được đề cập lướt qua gồm có phương pháp hồi quy vững với hàm Tukey's Bisquare giống như một hàm quyền số và độ lệch tuyệt đối trung bình (AAD) với thước đo của điều kiện hội tụ. Phép hồi quy vững là thuật toán bình phương nhỏ nhất lặp lại quyền số (IRLS) với ước lượng cổ điển  $M$ . Sự lựa chọn hàm quyền số ảnh hưởng tới việc suy diễn. Hàm quyền số của Huber được hy vọng là một giải pháp toàn bộ không giống như hàm Tukey và sẽ phù hợp hơn với cả các dữ liệu phức tạp như hồi quy đa mô hình. Bởi thước đo độ lệch tuyệt đối trung vị (MAD) vững hơn so với thước đo AAD và thường được sử dụng hơn. Chúng ta sẽ xem xét ảnh hưởng của cả hai hàm quyền số cũng như các thước đo từ đó quyết định được sự hội tụ.

**Từ khóa:** Giá trị chệch (hay giá trị bất thường), ước lượng  $M$ , bình phương nhỏ nhất lặp lại quyền số, Tukey's bisquare, quyền số Huber

## 1. Giới thiệu

Câu trả lời không có đáp án là vấn đề không thể tránh khỏi trong quá trình điều tra thống kê. Các giá trị khuyết sẽ có thể ước tính được nếu

nguyên nhân của chúng bắt nguồn từ sự chệch ở bảng kết quả thống kê. Có nhiều phương pháp ước tính khác nhau, nhưng trong bài viết này chúng tôi đề cập đến phương pháp ước tính thông qua việc hồi quy.

Thông thường quá trình ước tính được thực hiện sau khi toàn bộ dữ liệu không chính xác đã được loại bỏ hoặc chỉnh sửa lại. Tuy nhiên ảnh hưởng của các giá trị chệch có thể vẫn còn vì trong quá trình thay thế dữ liệu khuyết các giá trị chệch không cần thiết bị coi như là giá trị sai, nhưng chúng có thể được tách ra khỏi mô hình. Đặc biệt nếu có một số điểm ảnh hưởng thuộc dữ liệu bị tách ra việc suy diễn thông kê thông qua phương pháp bình phương nhỏ nhất OLS (bình phương nhỏ nhất) sẽ bị chệch. Một trong những cách giúp giải quyết vấn đề này là bỏ đi toàn bộ các giá trị chệch có ảnh hưởng trước khi thực hiện việc ước tính, và giải pháp thay thế này sẽ giới thiệu một số phương pháp thiết thực hơn thay vì phương pháp OLS.

Ủy ban Kinh tế quốc gia châu Âu (UNCEE) đã tổ chức một cuộc Hội thảo về việc biên tập lại dữ liệu thông kê nhằm tăng cường tính hài hòa giữa các phương pháp và khái niệm, đồng thời để trao đổi kinh nghiệm thực tế về việc ban hành khung quản lý chất lượng dữ liệu trong giai đoạn thu thập. Hội thảo đã công bố một loạt các ấn phẩm có tên gọi SDE nhằm chia sẻ những hiểu biết thực tế giữa các cơ quan thống kê quốc gia với nhau. Trong phần hai, Bienias et al. [2] mô tả cách thức thực hiện kĩ thuật Phân tích các dữ liệu được khám phá (EDA) ở Cục điều tra dân số Hoa Kỳ và giới thiệu phương pháp hồi quy vững.

Phương pháp hồi quy vững trình bày bởi Bienias et al. [2] chính là ước lượng cổ điển M của thuật toán IRLS (bình phương nhỏ nhất lặp lại quyền số). Bởi phương pháp này đặt quyền số cho

mỗi điểm dữ liệu theo sự chệch so với mô hình, quá trình phù hợp cần phải có phương pháp đồ thị và phương pháp này cũng giúp người làm công tác thống kê chú ý tới sự phân bố của dữ liệu, xem liệu mô hình đã phù hợp với dữ liệu chưa, hay có ảnh hưởng tới việc ước tính không từ đó giúp hoàn thiện công việc theo khía cạnh khác.

Bienias et al. [2] giới thiệu phương pháp Tukey's bisquare với hàm quyền số và độ lệch tuyệt đối trung bình (AAD) cho thước đo hệ số mà không cần phải giải thích. Tukey's Bisquare là một trong các phương pháp hay sử dụng hàm quyền số nhất, tuy nhiên, không thể hy vọng phương pháp này có thể đưa ra được giải pháp toàn bộ, khác với quyền số Huber, nhìn chung là một lựa chọn rất phổ biến cho phương pháp hồi quy Robust. Cũng là một thước đo hệ số, độ lệch tuyệt đối trung vị (MAD) thiết thực hơn so với AAD và được sử dụng rộng rãi hơn.

Mục tiêu bài viết nhằm làm rõ tác động của sự điều chỉnh IRLS tới kết quả đầu ra, vì vậy chúng ta có thể đưa ra một sự lựa chọn phù hợp với mục tiêu mà ta ước lượng và/hoặc xử lý bộ dữ liệu.

## 2. Phương pháp luận

### 2.1 Ước lượng M

Chúng ta xem xét mô hình hồi quy tuyến tính chuẩn tắc:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n$$

Trong đó:  $y_i$  là biến trả lời,  $x_i$  là biến giải thích và  $\varepsilon_i$  là tổng giá trị phần dư.  $\mathbf{b}$  là một ước lượng của  $\boldsymbol{\beta}$ , mô hình phù hợp là:

$$\hat{y}_i = a + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} = \mathbf{b}'\mathbf{x}_i$$

Và phần dư được tính như sau:

$$e_i = y_i - \hat{y}_i = y_i - \mathbf{b}'\mathbf{x}_i$$

Thước đo này tương đương với ước lượng M để tối thiểu hóa b

$$\sum_{i=1}^n \rho\left(\frac{y_i - \mathbf{b}'\mathbf{x}_i}{\sigma}\right),$$

$\sigma$  là thước đo hệ số và  $\rho$  là hàm tổn thất. Nếu chúng ta cho  $\psi = \rho'$  thì điều kiện cần thiết để tối thiểu hóa là b phải thỏa mãn công thức sau:

$$\sum_{i=1}^n \psi\left(\frac{y_i - \mathbf{b}'\mathbf{x}_i}{\sigma}\right) \mathbf{x}_i' = \mathbf{0}$$

Giờ thì chúng ta xác định được hàm quyền số  $w(e) = \psi(e)/e$  và đặt  $w_i = w(e_i)$ . Sau đó ước lượng b được chọn chính là nghiệm của:

$$\sum_{i=1}^n w_i \left(\frac{y_i - \mathbf{b}'\mathbf{x}_i}{\sigma}\right) \mathbf{x}_i' = \mathbf{0}$$

### 2.2 Thuật toán IRLS

Tiếp theo chính là quá trình chọn ra ước lượng b được đưa ra bởi Bienias et al. [2]. Họ chọn Tukey's bisquare cho hàm quyền số và AAD cũng giống như thước đo tham số  $\sigma$ .

1) Tính giá trị ước lượng ban đầu  $\mathbf{b}^{(0)}$  thông qua ước lượng OLS như sau với  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$  và  $\mathbf{y} = (y_1, \dots, y_n)'$

$$\mathbf{b}^{(0)} = [\mathbf{X}'\mathbf{X}]^{-1} \mathbf{X}'\mathbf{y}.$$

2) Tại mỗi giá trị nhắc lại j thực hiện tính các phần dư  $e_i^{(j-1)}$ , độ lệch tuyệt đối trung bình của nó (AAD)  $s^{(j-1)}$  và IRLS đặt quyền số  $w_i^{(j-1)}$  theo

hàm quyền số  $w(e_i^{(j-1)})$ .

3) Sau đó thực hiện giải hàm sau cho ước lượng bình phương nhỏ nhất với  $\mathbf{W}^{(j-1)} = \text{diag}\{w_i^{(j-1)}\}$  là một ma trận chéo.

$$\mathbf{b}^{(j)} = [\mathbf{X}'\mathbf{W}^{(j-1)}\mathbf{X}]^{-1} \mathbf{X}'\mathbf{W}^{(j-1)}\mathbf{y}$$

Các bước 2) và 3) được lặp lại cho tới khi  $\left| \frac{s^{(j)} - s^{(j-1)}}{s^{(j-1)}} \right|$  nhỏ hơn 0.01

### 2.3 Các hàm quyền số

Beaton và Tukey [1] đưa ra thuật toán IRLS với hàm Tukey's bisquare. Lý do hàm quyền số này trở lên phổ biến là nhờ đặc tính có thể loại bỏ hoàn toàn ảnh hưởng của các giá trị chệch khỏi việc ước lượng.

$$w_i = \begin{cases} \left(1 - \left(\frac{e_i}{cs}\right)^2\right)^2 & \text{if } |e_i| \leq cs, \\ 0 & \text{if } |e_i| > cs. \end{cases}$$

Hàm quyền số Huber [5] trước hết giúp ta xác định được ước lượng M tiếp theo là mở rộng hàm hồi quy trong Huber [6]. Quyền số Huber cũng được sử dụng rộng rãi luôn luôn cung cấp giải pháp chung không tính đến ước lượng ban đầu.

$$w_i = \begin{cases} 1 & \text{if } |e_i| \leq ks, \\ \frac{ks}{|e_i|} & \text{if } |e_i| > ks. \end{cases}$$

Hai hàm quyền số này đều được so sánh trong bài viết. Cả hai đều có chung một hệ số xác định, là c trong hàm Tukey và k trong hàm Huber để kiểm soát tính thực tế của việc ước lượng, và phụ thuộc vào sự yêu thích của người sử dụng. Các hệ số c và k khác nhau bởi thước đo hệ số.

Chúng được chỉ ra trong bảng 1. Vì Bienias et al. [2] cho  $c$  hàm Tukey chạy từ 4 tới 8 theo thước đo AAD, nên chúng ta cũng tính toán được những thông số tương ứng của  $c$  trong hàm Huber đã được chỉ ra trong Holland và Welsch [4]. Các thông số này là  $c = 4.685$  và  $k = 1.345$  với thước đo (SD).

### 3. Các thí nghiệm của Monte Carlo

#### 3.1 Mô phỏng dữ liệu

Biến giải thích độc lập  $x = (x_1, \dots, x_n)'$  và được phân bố đồng nhất theo số ngẫu nhiên giả trong khoảng  $(0,10)$ . Biến phụ thuộc được tạo ra phù hợp với mô hình hồi quy tuyến tính khoảng cách tổ  $\alpha = 5$ , độ dốc  $\beta = 2$ , và sai số  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$  tuân theo quy luật phân phối  $t$  số bậc tự do  $\nu = (1, 2, 3, 5, 10, \dots)$  độc lập với nhau. Mỗi tập dữ liệu bao gồm 100 điểm dữ liệu và có 100,000 tập dữ liệu được tạo ra bởi số bậc tự do và sai số. Chú ý là phân bố  $t$  với 1 bậc tự do tương đương với phân bố Cauchy, và số bậc tự do ban đầu cũng chính là một phân phối chuẩn.

#### 3.2 Sự so sánh

Những điều kiện điều kiện dưới đây được so sánh với thực nghiệm. Các con số của B-(2) trong thực tế không áp dụng cho thước đo MAD, nhưng lại áp dụng với thước đo SD bởi hàm MAD trong phần mềm R trả về các số liệu điều chỉnh phù hợp với SD. Toàn bộ sự mô phỏng trong bài viết đều được thực hiện trên phần mềm R 2.15.0

### 4. Tóm tắt kết quả

Sự so sánh thực hiện dựa trên hai khía cạnh. Thứ nhất, việc ước tính hiệu quả nhờ thực hiện đếm lặp đi lặp lại. Giá trị trung bình và giá

trị lớn nhất được tính trong mỗi thiết lập thể hiện trong Bảng 3 và Bảng 4. Giá trị ước tính ban đầu của phương pháp OLS và giá trị ước lượng chắc chắn WLS tiếp theo đều được đếm, vì vậy mỗi tập dữ liệu đều được đếm ít nhất hai lần. Giới hạn sự lặp lại là 150, và số lần lặp tối đa khác xa so với mức bình quân trong Bảng 3.

Khía cạnh thứ hai là sự phân tán của việc ước lượng. Sau khi ước lượng các hệ số hồi quy  $a$  và  $b$ ,  $\hat{y}_i$  được tính theo mô hình và các hệ số đã ước lượng sau đó lấy giá trị bình quân của mỗi tập dữ liệu. Độ lệch chuẩn của 100,000 tập dữ liệu với mỗi thiết lập được chỉ ra trong Bảng 5 và Bảng 6. Các giá trị trong bảng càng nhỏ thì việc dự đoán càng tốt.

#### 4.1 Các vấn đề trong quá trình thay thế dữ liệu khuyết

##### (1) Lặp lại vô hạn

Mặc dù quyền số Huber đã đạt tới sự lặp lại tối đa với thước đo MAD và hằng số cộng hưởng như trong *Bảng 4*, chắc chắn nó sẽ hội tụ giống như việc mở rộng giới hạn. Một vài phương pháp ước lượng hồi quy mạnh mẽ hơn phương pháp hồi quy ban đầu OLS sẽ cải thiện tốc độ tính toán, vì nó tạo điều kiện cho sự hội tụ.

Hàm Tukey's bisquare với thước đo MAD đôi khi sẽ không hội tụ khi sai số có chiều dài tương đối. Điều này xảy ra không thường xuyên và có thể dễ dàng giải quyết thông qua việc thay đổi hệ hằng số cộng hưởng. Mặc dù do đặc điểm của hàm quyền số là hàm gán quyền số 0 cho các giá trị ngoại lai, nhưng đáng chú ý là thước đo ít tính thực tế hơn AAD lại không bao giờ lặp lại vô hạn.

## (2) Ước lượng không khả thi

Vấn đề này xảy ra dựa theo hàm Tukey's bisquare với thước đo MAD, tương tự, với các tập dữ liệu phân dư với 1 bậc tự do không kể tới hằng số cộng hưởng. Ước lượng sai trong trường hợp có hai giá trị bất thường trong cùng một đường hồi quy và mức độ ảnh hưởng của cả hai đều thấp. Cả hai giá trị bất thường này cách xa nhau cũng như cách xa các điểm dữ liệu khác. Mức ảnh hưởng thấp của các giá trị bất thường này làm cho đường hồi quy thay đổi mà không cần phải thay đổi hệ số góc quá nhiều. Vì vậy nếu đường hồi quy tiến tới một trong số các giá trị bất thường, mà các giá trị bất thường này đủ cách xa các điểm dữ liệu khác, thì việc tính toán sẽ bị sai với tất cả các điểm dữ liệu ngoại trừ giá trị bất thường có quyền số bằng 0.

### 4.3 Kết luận từ các điều kiện

#### A. Hàm quyền số

Quyền số Huber hội tụ nhanh hơn một chút. Nó cung cấp cho chúng ta giải pháp chung và không bao giờ lặp vô hạn, bởi đặc tính của hàm là không bao gồm quyền số 0 cho nên mức ảnh hưởng của mọi điểm dữ liệu tới việc ước lượng sẽ không bị mất đi. Tuy nhiên, lại một vấn đề khác xuất hiện với các kết quả đầu ra mong đợi, do các giá trị bất thường vẫn giữ nguyên mức ảnh hưởng tới việc ước lượng.

Mặt khác, hàm Tukey's bisquare có thể loại bỏ hoàn toàn được mức ảnh hưởng của các giá trị bất thường, đó là nguyên nhân khiến đây không phải là giải pháp toàn bộ.

Nếu một tập dữ liệu bị hỏng nặng, thì quyền số Huber và thước đo MAD chính là một sự lựa chọn tuyệt vời; tuy nhiên, với việc thay thế dữ liệu điều tra, thì tập dữ liệu đã kiểm tra được kì vọng sẽ tương đối sạch trong giai đoạn thay thế. Bởi vậy thông thường yếu tố quan trọng nhất để chọn ra hàm quyền số sẽ là điều kiện của các giá trị bất thường.

#### B. Thước đo tham số

Việc lựa chọn thước đo tham số ảnh hưởng tới thời gian tính toán. Thước đo AAD giúp cho quá trình hội tụ nhanh hơn so với thước đo MAD với cả hai hàm hàm quyền số. Với hàm Tukey's bisquare, nhìn chung thước đo AAD xem xét tốt các vấn đề tính toán và tính chính xác của hàm số. Đối với quyền số Huber, thước đo MAD tốt hơn thước đo AAD nếu tập dữ liệu bị hỏng nặng.

#### C. Hằng số cộng hưởng

Hằng số cộng hưởng càng nhỏ ước lượng càng trở nên hiệu quả, giúp ta ước lượng chính xác hơn ngay cả khi tập dữ liệu bị hỏng nặng. Mặt khác, nó giúp loại bỏ quyền số dữ liệu khiến ảnh hưởng tương quan tiệm cận (ARE) trở nên trầm trọng hơn khi làm sạch tập dữ liệu.

Nhìn chung trong trường hợp tập dữ liệu bị hỏng vừa phải, chúng tôi đề xuất sử dụng giá trị lớn nhất của hàm Tukey's bisquare với thước đo AAD và giá trị nhỏ nhất của quyền số Huber.

#### D. Tiêu chí hội tụ

Bởi số lần lặp lại tăng ít hơn, nên việc cải thiện tính chính xác có thể không được rõ ràng.

## 5. Kết luận

Dưới đây là kiến nghị của IRLS trong tình huống tổng quát của quá trình thay thế giá trị điều tra khuyết thiếu với tập dữ liệu bị hỏng vừa phải, thay đổi theo phạm vi và thời gian.

Trước tiên, người sử dụng cần đưa ra cách giải quyết các giá trị bất thường. Nếu một người muốn toàn bộ các điểm dữ liệu phù hợp tối đa trong việc suy diễn và tìm ra được giải pháp toàn bộ (ví dụ luôn hy vọng sự hội tụ của số lần lặp), thì hàm quyền số Huber với thước đo MAD sẽ là

một điều hấp dẫn. Ngược lại, nếu một người không thể tránh được ảnh hưởng của các giá trị bất thường và mong muốn hạn chế ảnh hưởng thông qua suy diễn, thì hàm Tukey's Bisquare và thước đo AAD chính là lựa chọn của hầu hết quá trình thay thế.

Các thí nghiệm mô phỏng đã tán thành với việc lựa chọn hàm quyền số và thước đo tham số đưa ra bởi Bienias et al.[2]. Các kết quả đưa ra trong bài viết có thể cũng hữu ích trong việc chọn thiết lập trong tình huống cụ thể khác.

### *Tài liệu tham khảo:*

- [1] Beaton, A. E. and Tukey, J. W. (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, *Technometrics* 16, 147-185
- [2] Bienias, J. L., Lassman, D. M. Scheleur, S. A. & Hogan H. (1997) Improving Outlier Detection in Two Establishment Surveys. *Statistical Data Editing 2 - Methods and Techniques*. (UNSC and UNECE eds.), 76-83.
- [3] Fox, J. & Weisberg S. (2010) Robust Regression, Appendix to An R Companion to Applied Regression. Sage, Thousand Oaks, CA, 2nd ed. 2011
- [4] Holland, P. W. & Welsch, R. E. (1977), Robust Regression Using Iteratively Reweighted Least-Squares, *Communications in Statistics – Theory and Methods* 6(9), 813-827
- [5] Huber, P. J. (1964) Robust estimation of a location parameter, *Annals of Mathematical Statistics* 35, 73-101
- [6] Huber, P. J. (1973) Robust Regression: Asymptotics, Conjectures and Monte Carlo, *Annals of Statistics*.1, 799-821
- [7] Huber, P. J. & Ronchetti, Elvezio M. (2009) *Robust Statistics*, 2nd ed., John Wiley & Sons, Inc., New York
- [8] Rousseeuw, P. J. & Leroy, A. M. (1987) *Robust Regression and Outlier Detection*, John Wiley & Sons, Inc.
- [9] Tukey, J.W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

**Bảng 1.** Các hằng số cộng hưởng

Tukey's c với AAD	4	6	8
Tukey's c với SD	5.01	7.52	10.03
Tukey's c với MAD	7.43	11.15	14.87
Huber's k với AAD	1.15	1.72	2.30
Huber's k với SD	1.44	2.16	2.88
Huber's k với MAD	2.13	3.20	4.27

**Bảng 2.** Các điều kiện để so sánh

A. Hàm quyền số :	(1) Tukey's bisquare	(2) Quyền số Huber		
B. Thước đo hệ số :	(1) Độ lệch tuyệt đối trung bình (AAD)	(2) Độ lệch tuyệt đối trung vị (MAD)		
C. Hằng số cộng hưởng :	Tukey[B-(1)]	(i) TK4: 4	(ii) TK6: 6	(iii) TK8: 8
	Tukey [B-(2)]	(i) TK4: 5.01	(ii) TK6: 7.52	(iii) TK8: 10.03
	Huber[B-(1)]	(i) HB4: 1.15	(ii) HB6: 1.72	(iii) HB8: 2.30
	Huber[B-(2)]	(i) HB4: 1.44	(ii) HB6: 2.16	(iii) HB8: 2.88
D. Tiêu chuẩn hội tụ của sự thay đổi tỷ lệ thuận với quy mô	(a) 0.01	(b) 0.001	(c) 0.0001	

Bảng 3. Số lần lặp trung bình

Thuốc đo	AAD						MAD					
wt & tc	TK4	TK6	TK8	HB4	HB6	HB8	TK4	TK6	TK8	HB4	HB6	HB8
Tỷ lệ cv	0.01						0.01					
df 1	3.72	3.66	3.64	3.75	3.71	3.72	5.90	5.40	5.20	5.55	5.15	5.01
df 2	3.43	3.26	3.15	3.30	3.14	3.04	4.90	4.33	4.04	4.47	4.01	3.78
df 3	3.24	3.03	2.89	3.08	2.91	2.81	4.47	3.86	3.53	4.09	3.61	3.35
df 5	3.07	2.82	2.65	2.90	2.73	2.62	4.10	3.45	3.12	3.80	3.29	2.96
df 10	2.96	2.65	2.48	2.80	2.61	2.47	3.79	3.15	2.86	3.61	3.04	2.60
df Inf	2.86	2.51	2.34	2.72	2.51	2.31	3.50	2.91	2.69	3.42	2.77	2.23
Tỷ lệ cv	0.001						0.001					
df 1	4.73	4.54	4.46	4.61	4.50	4.47	7.61	6.83	6.52	6.98	6.36	6.15
df 2	4.95	4.39	4.10	4.49	4.07	3.85	6.51	5.54	5.09	5.83	5.04	4.70
df 3	4.93	4.22	3.85	4.40	3.90	3.61	6.01	4.97	4.46	5.39	4.56	4.16
df 5	4.86	4.00	3.59	4.31	3.75	3.39	5.55	4.45	3.95	5.06	4.16	3.63
df 10	4.78	3.81	3.37	4.25	3.64	3.22	5.16	4.08	3.63	4.82	3.83	3.09
df Inf	4.69	3.62	3.18	4.23	3.54	3.05	4.78	3.79	3.39	4.58	3.43	2.47
tỷ lệ cv	0.0001						0.0001					
df 1	5.82	5.47	5.31	5.56	5.34	5.26	9.32	8.25	7.84	8.40	7.58	7.30
df 2	6.59	5.59	5.11	5.84	5.10	4.71	8.12	6.77	6.15	7.18	6.09	5.63
df 3	6.80	5.49	4.89	5.93	5.00	4.48	7.56	6.09	5.40	6.71	5.54	4.98
df 5	6.88	5.31	4.61	5.96	4.89	4.26	7.01	5.48	4.78	6.34	5.05	4.30
df 10	6.90	5.12	4.37	5.98	4.81	4.07	6.55	5.04	4.37	6.07	4.64	3.58
df Inf	6.87	4.91	4.16	6.03	4.72	3.88	6.09	4.69	4.09	5.79	4.12	2.71



**Bảng 4. Số lần lặp tối đa**

Thuốc đo	AAD						MAD					
wt & tc	TK4	TK6	TK8	HB4	HB6	HB8	TK4	TK6	TK8	HB4	HB6	HB8
Tỷ lệ cv	0.01						0.01					
df 1	6	6	6	6	6	6	150	150	150	21	53	76
df 2	6	5	5	5	5	5	36	22	150	18	19	13
df 3	6	5	5	6	5	4	23	17	150	11	11	11
df 5	7	5	5	5	5	4	25	16	13	14	12	14
df 10	6	5	4	5	5	4	15	10	8	11	9	8
df Inf	6	5	4	6	5	4	12	9	5	10	8	6
Tỷ lệ cv	0.001						0.001					
df 1	8	8	8	8	7	7	150	150	150	25	63	122
df 2	9	7	7	7	7	6	39	146	150	29	37	19
df 3	10	7	6	8	6	6	37	25	150	17	14	20
df 5	10	7	6	8	6	6	115	27	19	17	19	19
df 10	10	7	6	8	6	5	24	16	11	15	13	12
df Inf	10	7	5	8	6	5	19	14	7	15	11	8
Tỷ lệ cv	0.0001						0.0001					
df 1	11	9	10	9	9	9	150	150	150	30	63	150
df 2	13	10	9	10	8	7	150	150	150	41	54	26
df 3	13	9	8	11	8	7	46	32	150	23	20	30
df 5	13	9	7	11	8	7	150	37	26	22	25	25
df 10	15	9	7	11	8	7	33	21	14	21	17	16
df Inf	14	8	7	11	8	7	33	19	8	20	15	11

**Bảng 5.** Độ lệch tiêu chuẩn của trung bình ước lượng với thước đo AAD

	AAD (tỷ lệ chuyển đổi 0.01)					
	df 1	df 2	df 3	df 5	df 10	df Inf.
OLS	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862
TK4	0.6521	0.5944	0.5914	0.5895	0.5889	0.5879
TK6	0.6803	0.5963	0.5918	0.5893	0.5882	0.5867
TK8	0.7117	0.5986	0.5928	0.5896	0.5882	0.5864
HB4	2.1044	0.5954	0.5914	0.5892	0.5884	0.5872
HB6	3.0941	0.5981	0.5923	0.5893	0.5882	0.5866
HB8	4.1281	0.6010	0.5934	0.5898	0.5882	0.5864

	AAD (tỷ lệ chuyển đổi 0.0001)					
	df 1	df 2	df 3	df 5	df 10	df Inf.
OLS	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862
TK4	0.6522	0.5944	0.5915	0.5898	0.5892	0.5884
TK6	0.6803	0.5963	0.5918	0.5893	0.5882	0.5867
TK8	0.7116	0.5986	0.5927	0.5896	0.5882	0.5864
HB4	2.1038	0.5953	0.5913	0.5891	0.5885	0.5874
HB6	3.0923	0.5981	0.5922	0.5893	0.5882	0.5867
HB8	4.1270	0.6010	0.5934	0.5898	0.5882	0.5864

**Bảng 6.** Độ lệch tiêu chuẩn của trung bình ước lượng với thước đo AD

	MAD ( tỷ lệ chuyển đổi 0.01)					
	df 1	df 2	df 3	df 5	df 10	df Inf.
OLS	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862
TK4	0.7321	0.5946	0.5915	0.5893	0.5882	0.5867
TK6	0.6302	0.5967	0.5925	0.5897	0.5883	0.5863
TK8	0.6273	0.5990	0.5937	0.5902	0.5884	0.5863
HB4	0.6113	0.5955	0.5918	0.5893	0.5882	0.5866
HB6	0.6221	0.5984	0.5932	0.5899	0.5883	0.5863
HB8	0.6334	0.6012	0.5945	0.5905	0.5886	0.5862

	MAD ( tỷ lệ chuyển đổi 0.0001)					
	df 1	df 2	df 3	df 5	df 10	df Inf.
OLS	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862
TK4	0.7292	0.5945	0.5916	0.5893	0.5882	0.5867
TK6	0.6295	0.5967	0.5925	0.5897	0.5883	0.5863
TK8	0.6263	0.5990	0.5937	0.5902	0.5884	0.5863
HB4	0.6105	0.5954	0.5917	0.5892	0.5882	0.5866
HB6	0.6216	0.5984	0.5932	0.5899	0.5883	0.5863
HB8	0.6331	0.6011	0.5945	0.5905	0.5886	0.5863