

TƯƠNG QUAN THỰC NGHIỆM VÀ PHƯƠNG PHÁP BÌNH PHƯƠNG BÉ NHẤT

Phạm Hữu Sy¹

Tóm tắt: Trong các ngành khoa học nhiều trường hợp thường phải xác định tương quan thực nghiệm. Tương quan thực nghiệm có thể được thể hiện bằng nhiều cách khác nhau như bảng tương quan, đồ thị tương quan, toán đồ tương quan, phương trình tương quan. Tác giả bài báo đặc biệt lưu ý đến các sai sót đã xảy ra trong thời gian qua trong các luận văn thạc sỹ, luận án tiến sỹ, tiêu chuẩn xây dựng liên quan đến phương trình tương quan thiết lập theo phương pháp bình phương bé nhất, với mong muốn các lỗi này không lặp lại trong công tác đào tạo.

Từ khóa: Tương quan thực nghiệm, phương pháp bình phương bé nhất

1. Đặt vấn đề

Trong Địa kỹ thuật nói riêng, trong các lĩnh vực khoa học nói chung tồn tại nhiều tương quan giữa các đại lượng mà không thể xác định được trực tiếp bằng lý thuyết tính toán. Trong trường hợp đó cách duy nhất là xác định bằng thực nghiệm, nghĩa là làm nhiều thí nghiệm để tìm ra tương quan đó. Phương pháp xác định tương quan bằng thực nghiệm đã được sử dụng rộng rãi trong thực tế. Việc xác định tương quan thực nghiệm giữa các đại lượng là một vấn đề rất nhỏ trong khoa học, tuy nhiên, thỉnh thoảng vẫn gặp một số lỗi trong ứng dụng, ngay cả trong tiêu chuẩn Việt Nam và cả ở một số người làm công tác nghiên cứu và giảng dạy. Điều này có thể dẫn đến hệ lụy là các lỗi đó có thể truyền từ thế hệ này sang thế hệ khác. Vì vậy, chúng tôi cho rằng cũng cần phải bàn đến vấn đề tương quan thực nghiệm và phương pháp biểu diễn chúng để vấn đề được hiểu rõ thấu đáo, tránh lặp lại các sai lầm.

2. Lịch sử phát triển của vấn đề tương quan thực nghiệm

Như đã nói ở trên, xác định tương quan bằng thực nghiệm tuy là một vấn đề nhỏ trong khoa học nhưng bản thân nó cũng có lịch sử phát triển của mình. Trước đây, để tìm tương quan người ta làm nhiều thí nghiệm theo một kế hoạch nhất định, sau đó kết quả thí nghiệm được thể hiện dưới dạng bảng. Ở cột thứ nhất thể hiện hàm số - một đại lượng của thí nghiệm theo quy luật tăng hoặc giảm nhất định. Cột thứ hai thể hiện biến số, tức là kết quả thí nghiệm. Một số

tiêu chuẩn ngành, tiêu chuẩn quốc gia hiện nay vẫn còn biểu diễn tương quan giữa các đại lượng dưới dạng bảng, ví dụ như bảng tương quan chỉ số búa N_{30} ~ góc ma sát trong ϕ hoặc N_{30} ~ độ bền nén có nở hông q_u trong tiêu chuẩn TCXD 226: 1999 hoặc tương quan giữa góc ma sát trong ϕ , lực dính C , mô đun biến dạng E của đất sét với độ sệt trong tiêu chuẩn xây dựng TCXD 45-78, tương quan giữa sức kháng mũi xuyên tĩnh q_c và góc ma sát trong của đất ϕ trong tiêu chuẩn 20TCN – 174 – 89.

Về mặt toán học, bảng tương quan là một cách thể hiện đơn giản nhất và vì vậy, là cách thể hiện “yếu nhất”. Vì tương quan được thể hiện bằng từng cặp giá trị riêng biệt nên khi người sử dụng cần tìm tương quan của các giá trị trung gian thì buộc phải tính nội suy theo quan hệ tuyến tính trong khi thực chất nó không phải là tuyến tính. Điểm yếu thứ hai là cách thể hiện tương quan này không cho phép đánh giá mức độ chặt chẽ của tương quan. Điểm yếu thứ ba là tương quan chỉ được xác định và thể hiện rất hữu hạn, chỉ trong phạm vi của các giá trị thí nghiệm. Khi cần xác định tương quan vượt ra ngoài phạm vi thí nghiệm thì không phải là nội suy mà phải ngoại suy và vì vậy, không bảo đảm độ tin cậy của tương quan. Do những điểm yếu đó mà hiện nay chỉ còn được sử dụng rất ít mà được thay thế dần bởi tương quan dưới dạng đồ thị.

Tương quan đồ thị có nhiều ưu điểm hơn tương quan bảng. Lúc này quan hệ giữa các đại lượng được lập thành đồ thị biến thiên. Trên 2 trục của đồ thị thể hiện đường tương quan cấp

¹Đại học Thủy lợi

đôi đồng thời trên biểu đồ có thể thể hiện nhiều đường tương quan biến thiên theo thông số thứ ba. Ưu điểm thứ hai của tương quan đồ thị là thể hiện rất trực quan tương quan giữa các đại lượng và có thể đánh giá xu thế của tương quan ngoài phạm vi kết quả thí nghiệm, ngoại suy kết quả với một độ tin cậy nhất định. Ưu điểm thứ ba là có thể thể hiện đồng thời nhiều tương quan trên cùng một biểu đồ, gọi là phép “chồng đồ thị”. Khi cần biểu thị tương quan của 2, 3 đại lượng với cùng một biến số có thể thể hiện trực hoành là biến số chung và trục tung là các đại lượng phụ thuộc. Cách thể hiện này vừa gọn, lại thể hiện trực quan các xu thế biến đổi khác nhau của các đại lượng với cùng một biến số.

Một cách thể hiện khác của biểu đồ là các toán đồ. Toán đồ cũng thể hiện tương quan đồng thời giữa các đại lượng mà thông thường là 3 đại lượng. Toán đồ thể hiện bằng các trục đứng được bố trí theo các trật tự, khoảng cách và độ so le xác định.

Cách thể hiện tương quan thứ ba là tương quan giải tích. Đây là sự phát triển của tương quan đồ thị. Nếu như trong cách biểu diễn tương quan bằng đồ thị, đường biểu đồ nối trực tiếp các điểm với nhau mà có thể chứa các sai số thí nghiệm thì ngày nay người ta ứng dụng phương pháp bình phương bé nhất trong tương quan đồ thị tìm ra đường trung bình hay là đường xu thế (trendline) biểu diễn tương quan đó. Đường trung bình được xác định với tập hợp chuỗi số liệu đủ lớn và đã được loại trừ sai số thô, biểu diễn tương quan chính xác hơn, đặc trưng hơn. Sau khi có đường trung bình sẽ xác định được phương trình giải tích của đường trung bình đó. Từ chỗ xác định được phương trình giải tích, coi đường trung bình là chính xác nhất, có thể đánh giá mức độ chặt chẽ của tương quan thông qua hệ số tương quan. Hệ số tương quan của chuỗi số liệu thí nghiệm là chỉ số thể hiện mức độ chặt chẽ của tương quan. Đối với tương quan tuyến tính một biến và hai biến hệ số tương quan r được xác định thứ tự theo các công thức sau [1], [2], [7]:

$$R = \frac{\sum (x_i - x_{tb})(y_i - y_{tb})}{\sqrt{\sum (x_i - x_{tb})^2 \sum (y_i - y_{tb})^2}} \quad (1)$$

$$R = \sqrt{(r_{xz}^2 - 2r_{xy}r_{xz} + r_{yz}^2)/(1 - r_{xy}^2)} \quad (2)$$

trong đó r_{xz}, r_{xy}, r_{yz} là hệ số tương quan theo từng cặp.

Tất cả các kiến thức nêu trên thiết nghĩ hầu hết những người làm công tác khoa học đều biết. Thế nhưng trên thực tế thỉnh thoảng vẫn xảy ra các lỗi khó chấp nhận mà như đã nói ở trên, xảy ra ngay cả trong tiêu chuẩn ngành, xảy ra với những người làm công tác giảng dạy, nghiên cứu. Vì vậy, mục đích của người viết bài này là chỉ rõ các sai sót để tránh lặp lại và truyền thụ cho các thế hệ sau.

3. Nội dung phương pháp bình phương bé nhất

Để đề cập đến các lỗi đã xảy ra trong các tiêu chuẩn Việt Nam và với một số người đã phạm phải khi áp dụng tương quan như đã nói ở trên, chúng tôi xin nói đến một cách kỹ hơn phương pháp bình phương bé nhất. Phương pháp bình phương bé nhất đã được nhà toán học người Đức Carl Friedrich Gauss đưa ra vào năm 1795, đã được trình bày kỹ trong giáo trình Toán học cao cấp, tập III của Bộ Giáo dục và Đào tạo. Bản chất toán học của phương pháp bình phương bé nhất có thể tóm lược như sau. Từ các kết quả thí nghiệm ta có các cặp tương quan y_i theo x_i tuân theo một hàm giải tích $y = f(x)$ nhất định. Tương quan giữa các đại lượng có thể là tuyến tính hoặc phi tuyến. Vì đây là tương quan thực nghiệm nên các cặp giá trị x_i, y_i tương ứng chắc chắn không nằm trên cùng một đường đồ thị mà có sự sai lệch nhất định. Gọi độ sai lệch đó là d_i , ta có:

$$d_1 = y_1 - f(x_1), d_2 = y_2 - f(x_2), \dots, d_n = y_n - f(x_n).$$

Phương pháp bình phương bé nhất là tìm một đường $y = f(x)$ khả dĩ nhất, tức là đi qua gần nhất tất cả các điểm thí nghiệm sao cho:

$$F = d_1^2 + d_2^2 + \dots + d_n^2 = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n [y_i - f(x_i)]^2 = \min \quad (3)$$

Để dễ thuyết minh ta lấy tương quan tuyến tính làm ví dụ. Đường cần tìm sẽ có dạng:

$$y = ax + b \quad (4)$$

Theo phương trình (1) ta có:

$$d_1 = y_1 - ax_1 - b \quad (5)$$

$$d_2 = y_2 - ax_2 - b \quad (6)$$

$$\dots\dots\dots$$

$$d_n = y_n - ax_n - b \quad (7)$$

Bình phương hai vế và cộng vế theo vế ta có phương trình:

$$F = \sum_{i=1}^n (y_i - ax_i - b)^2 \quad (8)$$

Để tìm được phương trình của y coi a và b là các ẩn số và xác định điểm cực tiểu của hàm F, có nghĩa là tìm giá trị nhỏ nhất của tổng bình phương các giá trị d. Chính vì vậy phương pháp này gọi là phương pháp bình phương bé nhất. Từ phương trình (8) lấy đạo hàm hai vế lần lượt theo a và b và cho bằng 0 sẽ có hệ phương trình:

$$\frac{\partial F}{\partial a} = \sum (y_i - ax_i - b)(-x_i) = 0 \quad (9)$$

$$\frac{\partial F}{\partial b} = \sum (y_i - ax_i - b) = 0 \quad (10)$$

Giải hệ phương trình trên theo a và b ta được:

$$a = \frac{n \sum y_i x_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (11)$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (12)$$

Thay các hệ số a, b vừa tìm được vào phương trình (4) xác định được phương trình hồi quy cần tìm. Hệ số tương quan được xác định theo công thức (1) ở trên.

Hiện nay phần mềm EXCEL đã lập sẵn các phương trình cho tương quan tuyến tính và phi tuyến, trong đó tương quan phi tuyến bao gồm hàm số mũ, hàm logarit, hàm đa thức bậc cao, hàm lũy thừa.

Trong EXCEL, sau khi vẽ và cho hiện phương trình hồi quy, click vào ô “display R-squared value on chart” sẽ xuất hiện giá trị R. Các tài liệu khác nhau chia khoảng của các mức độ chặt chẽ khác nhau thì khác nhau. Nói chung có thể chấp nhận $R > 0,8$ là tương quan chặt, $0,4 < R < 0,8$ là tương quan trung bình, $R < 0,4$ là tương quan yếu.

4. Một số lỗi khi biểu diễn tương quan thực nghiệm

Bây giờ chúng tôi muốn nói đến một số lỗi

khi ứng dụng phương trình hồi quy để xác định tương quan thực nghiệm mà chúng tôi đã gặp. Sẽ là vấn đề nhỏ không phải bận tâm nếu đó chỉ là lỗi của một vài người, trong một vài tài liệu cá nhân. Tuy nhiên, thực tế đó là lỗi trong tiêu chuẩn ngành và trong các luận văn thạc sỹ, tiến sỹ mà có thể lặp lại và lan truyền. Vì vậy thiết nghĩ cần làm rõ để chấn chỉnh lại.

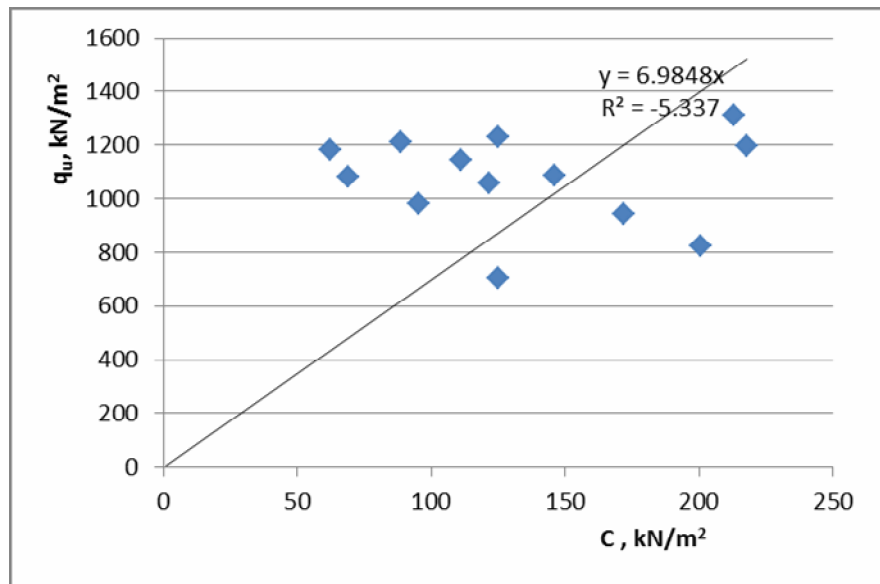
4.1 Lỗi khi chọn phương trình hồi quy

Trước hết nói đến dạng phương trình hồi quy. Như đã nói ở trên, sau khi có tập hợp số liệu thí nghiệm, vẽ biểu đồ phân bố các kết quả đó lên hệ tọa độ Đềcac, quan sát dạng phân bố của các điểm để lựa chọn phương trình giải tích thích hợp. EXCEL đã cho phép lựa chọn các loại phương trình khác nhau. Nếu lựa chọn đúng, khi vẽ đường trendline và click cho xuất hiện phương trình, hệ số tương quan sẽ cao, biểu hiện tương quan chặt. Ngược lại sẽ cho kết quả tương quan yếu. Chúng tôi đã được đọc một luận văn cao học, mặc dầu dạng phân bố của các điểm thể hiện rõ là dạng hàm mũ nhưng học viên vẫn chọn trendline dạng tuyến tính và đương nhiên, cho hệ số tương quan thấp. Lỗi này chứng tỏ rằng người dùng không hiểu các dạng phương trình chính tắc, các quy luật phân phối của chuỗi số liệu và không biết đến khái niệm “hệ số tương quan”.

4.2 Lỗi khi chọn đường trung bình và xác định hệ số tương quan

Lỗi thứ hai mà chúng tôi đã gặp cũng trong một luận văn thạc sỹ liên quan đến đường trendline khi học viên nghiên cứu “Xây dựng mối tương quan thực nghiệm giữa các chỉ tiêu cơ lý ...” nhưng ở một khía cạnh khác. Sau khi có kết quả thí nghiệm (xem bảng), học viên đã vẽ biểu đồ và thêm đường trendline. Ở đây việc lựa chọn tương quan để nghiên cứu về mặt chuyên môn không ổn, tuy nhiên chúng tôi chỉ bàn đến tương quan dưới góc độ toán học. Căn cứ dạng phân bố của các điểm thể hiện quy luật tuyến tính đầy đủ, thay vì thêm đường trendline dạng $y = ax + b$ thì người dùng lại vô tình chọn dạng $y = ax$ bằng cách click vào ô “set intercept = 0,0” và như vậy buộc đồ thị đi qua gốc tọa độ (xem hình đồ thị bên dưới). Vì vậy, R^2 cho giá trị âm.

C, kN/cm ²	q _u , kN/cm ²
62,17	1183,02
125,14	705,55
69	1080,31
95,47	982,34
125,14	1230,77
111,353	1145,74
121,794	1057,39
88,739	1210,8
218,1	1196
213,1	1309
200,8	825
146,3	1089
171,9	943



Hình 1. Lỗi khi chọn đường trung bình bắt đi qua gốc tọa độ

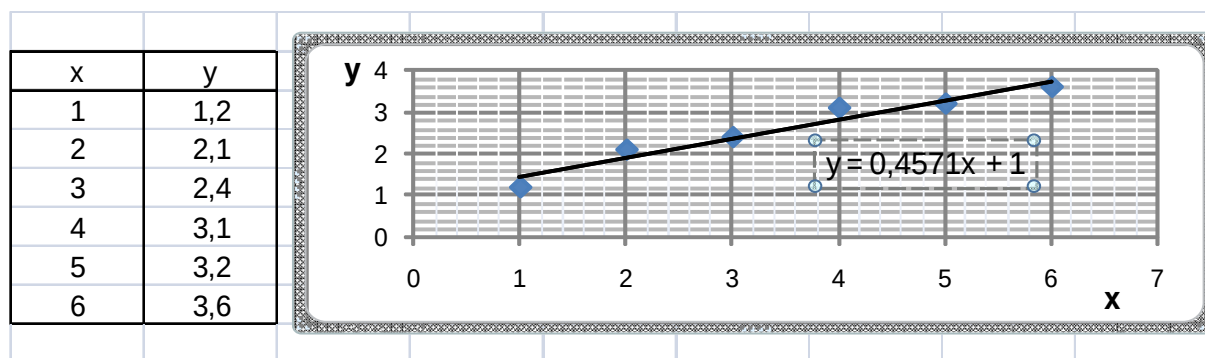
Như chúng tôi đã nói ở trên, sẽ không phải bạn tâm nếu đây là lỗi ở một người, một chỗ nào đó ít được biết đến nhưng đây lại là lỗi trong luận văn tốt nghiệp cao học. Khi lỗi này không được chỉnh sửa kịp thời những người khác tham khảo luận văn này cũng sẽ lặp lại sai lầm này.

4.3 Đổi vai trò hàm và biến trong phương trình hồi quy

Lỗi thứ ba cũng trong luận văn này và đồng thời cũng xảy ra y như vậy trong tiêu chuẩn Việt Nam TCVN 4197 – 1995, liên quan đến vai trò hàm – biến khi thiết lập tương quan.

Như đã nói ở trên, để xác lập tương quan

thực nghiệm bằng giải tích người ta tìm cực tiểu của hàm F, tức là cực tiểu của tổng các Δy^2 để xác định được hàm $y = f(x)$. Vì vậy, không thể tùy tiện đổi vai trò hàm và biến. Nếu trong toán học, đối với một công thức lý thuyết có thể tùy ý đổi vai trò hàm và biến thì đối với công thức thực nghiệm muốn đổi vai trò biến thành hàm phải xác định lại từ đầu, có nghĩa là phải dùng phương pháp bình phương bé nhất giải và tìm cực tiểu của biến Δx . Để làm rõ hơn vấn đề này ta hãy làm một ví dụ cụ thể một tập hợp số liệu trong bảng dưới đây.

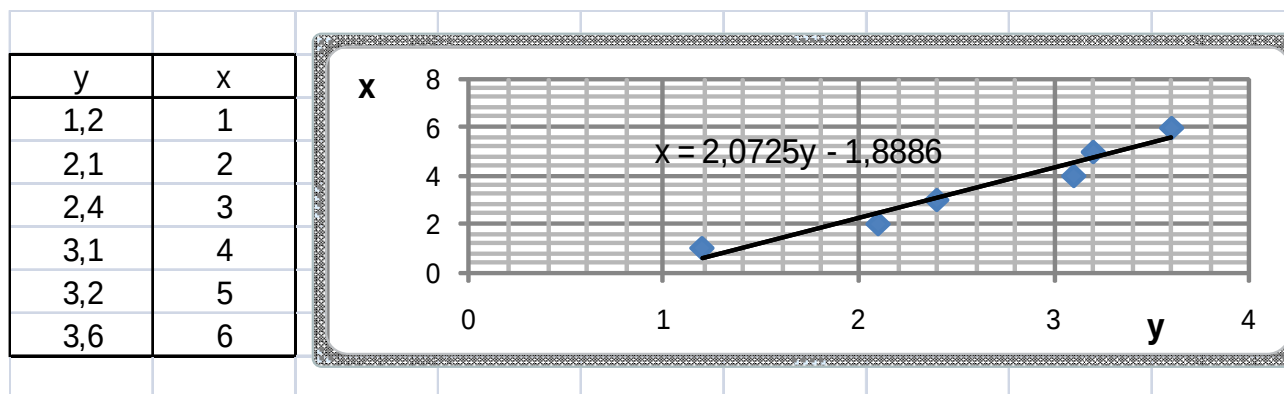


Hình 2. Biểu đồ tương quan hàm $y = f(x)$

Phương trình hồi quy của đường trendline là: $y = 0,4571x + 1$. Nếu thuận tụy tính toán để đổi vai trò biến thành hàm, tức là $x = f(y)$ ta sẽ có:

$$x = \frac{1}{0,4571}y - \frac{1}{0,4571} \rightarrow x = 2,188y - 2,188 \quad (13)$$

Bây giờ ta đổi thứ tự cột trong bảng trên và vẽ lại đồ thị, EXCEL sẽ tìm cực tiểu của Δx :



Hình 3. Biểu đồ biểu diễn tương quan sau khi đã đổi vai trò $x = f(y)$

Như vậy, phương trình của đường trendline bây giờ là:

$$x = 2,0725y - 1,8886 \quad (14)$$

Rõ ràng là phương trình (13) khác phương trình (14). Điều đó là do cùng một tương quan nhưng phương trình biểu diễn cực tiểu của Δy sẽ khác với phương trình biểu diễn cực tiểu của Δx . Đối với công thức thực nghiệm, muốn biến đổi qua lại giữa hàm và biến thoả mái như đối với công thức lý thuyết thì khi xác lập phương trình hồi quy thay vì tìm cực tiểu của hàm $F = \sum (\Delta y)^2$ ta hãy tìm cực tiểu của hàm $F = \sum [(\Delta y)^2 + (\Delta x)^2]$. Trong luận văn trên, học viên đã xác lập tương quan giữa cường độ kháng nén nở hông q_u theo góc ma sát trong của đất: $q_u = f(\varphi)$. Tuy nhiên ở chương cuối, khi ứng dụng cho một công trình cụ thể thì lại là $\varphi = f(q_u)$.

Một dạng sai sót khác là xảy ra với nghiên cứu sinh. Sau khi làm một thí nghiệm và vẽ được biểu đồ quan hệ, nghiên cứu sinh này đã dùng phương pháp thử dần để xác định các hệ số, đưa tương quan của các đại lượng thí nghiệm về dạng phương trình thực nghiệm của một tác giả khác đã công bố. Việc làm này thể hiện nghiên cứu sinh không hiểu ý nghĩa toán học của phương trình thực nghiệm và cách xây dựng nó.

Như trên đã nói, sai lầm về việc đổi vai trò

hàm và biến không chỉ có học viên cao học mắc phải mà xảy ra cả trong tiêu chuẩn Việt Nam. Điều này gây ra sai số hệ thống đối với những ai sử dụng tiêu chuẩn này. Trong tiêu chuẩn TCVN 5747-1993 đã đưa ra công thức thực nghiệm xác định giới hạn chảy theo Casagrande từ kết quả thí nghiệm bằng chày Vaxiliep:

$$w_{cs} = \frac{1}{a}(w_{vl} + b) \quad (15)$$

Tuy nhiên, trong tiêu chuẩn TCVN 4197 – 1995 lại thuận túy giải ngược công thức (15) để đưa ra công thức:

$$w_{vl} = aw_{cs} - b \quad (16)$$

5. Kết luận:

Từ các nội dung trình bày ở trên có thể rút ra một số kết luận sau:

1. Mặc dầu vấn đề biểu diễn tương quan thực nghiệm là một vấn đề nhỏ trong khoa học nhưng lại được sử dụng trong hầu hết các lĩnh vực, kể cả khoa học tự nhiên và xã hội, vì vậy, việc chỉ ra các sai sót khi ứng dụng là cần thiết để không lặp lại về sau.

2. Các lỗi sai sót khi ứng dụng phương trình hồi quy sử dụng phương pháp bình phương bé nhất thường là chọn phương trình hồi quy, chọn đường trung bình không đúng và tự tiện đổi vai trò hàm và biến.

3. Các lỗi này ảnh hưởng đến kết quả nghiên cứu, vì vậy, cần phải được khắc phục, nhất là đối với những người làm công tác giảng dạy và nghiên cứu vì nếu không có thể ảnh hưởng lâu dài.

TÀI LIỆU THAM KHẢO

1. Hệ số tương quan. Wikipedia tiếng Việt.
2. Phạm Hữu Sy “Xác lập tương quan giữa mô đun tổng biến dạng và một số chỉ tiêu vật lý của đất xây dựng bằng phương trình hồi quy”. Tạp chí Nông nghiệp và Phát triển Nông thôn số 3-2002
3. Phạm Hữu Sy “Nghiên cứu đối chứng và xác định quan hệ độ ẩm giới hạn chảy theo phương pháp Casagrande và Vaxiliev” Tạp chí Địa kỹ thuật số 01-1998.
4. Tiêu chuẩn xây dựng TCXD 45-78 “Tiêu chuẩn thiết kế nền nhà và công trình”
5. Tiêu chuẩn ngành 20TCN – 174 – 89 “Đất xây dựng – Phương pháp thí nghiệm xuyên tĩnh”
6. Tiêu chuẩn xây dựng TCXD 226: 1999 “Đất xây dựng – Phương pháp thí nghiệm hiện trường – Thí nghiệm xuyên tiêu chuẩn”
7. Tạ Ngọc Đạt và nnk. “Toán học cao cấp, tập III” NXB Giáo dục

Abstract

EMPIRICAL RELATION AND MINIMUM SQUARE METHOD

In the different science fields there are many cases must be used empirical relation between terms. Empirical relation may be present by different ways as table, graphic, mathematic graphic and analysis formula. The author of the paper specially notices the faults that happen in master and doctorate thesis, government standards which concern relation presented in formula type used minimum square method. These faults were discussed in order will not repeated again.

Keywords: *empirical relation, minimum square method*

Người phản biện: **GS.TS. Nguyễn Chiến**

BBT nhận bài: 12/3/2013

Phản biện xong: 13/6/2013