Tạp chí KHKT Nông nghiệp, Tập 2,số 3/2004

SO SÁNH CÁC TRUNG BÌNH SAU PHÂN TÍCH PHƯƠNG SAI

Comparison of treatment means after analysis of variance

Nguyễn Đình Hiền[1]

SUMMARY

The paper introduces methods for comparison of treatment means following analysis of variance (ANOVA). If the null hypothesis in the analysis of variance is rejected, the next step is to compare the treatment means (post hoc comparisons). Different tests such as LSD, Scheffé, Tukey, Bonferroni, S-N-K and Duncan are presented with recommendations, especially concerning the cases of equal numbers and non-equal numbers of replicates.

Keywords*:* Analysis of variance, means, replicates, null hypothesis, comparison

ĐẶT VẤN ĐỀ

Phân tích phương sai một nhân tố với a mức: A_1, A_2, . . . , A_a được bảng:

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương trung bình	Giá trị F thực hiện	Giá trị tới hạn F
Nhân tố	a-1 dfA	SSA	MsA = SSA/dfA	F_{tn} = msA/msE	F(α,dfA,dfE)
Sai số	n-a dfE	SSE	msE		
Toàn bộ	n-1	SSTO			

Trong phân tích phương sai một nhân tố msE được ký hiệu là se^2, se được gọi là sai số thí nghiệm, dfE là bậc tự do của sai số.

Nếu $F_{tn} \le F$ (α, dfA,dfE) thì chấp nhận giả thiết H_0: "Các trung bình của các mức bằng nhau". Nếu ngược lại thì bác bỏ H_0, tức là chấp nhận H_1: " Các trung bình của các mức không bằng nhau".

Phân tích phương sai hai hay ba nhân tố thì có nhiều giả thiết ứng với các trung bình khác nhau (trung bình của nhân tố 1, trung bình của nhân tố 2, trung bình của tương tác, . . .).

Sau khi phân tích phương sai và kết luận "Các trung bình của các mức khác nhau" thì vấn đề đặt ra là cần so sánh các trung bình để biết cụ thể các trung bình nào bằng nhau, các trung bình nào khác nhau.

Để kiểm định giả thiết về sự bằng nhau của một loại trung bình phải tính tỷ số F_{tn}. Bình phương trung bình dùng làm mẫu số trong F_{tn} chính là bình phương của sai số dùng trong việc ước lượng và so sánh các trung bình tương ứng, còn bậc tự do tương ứng của mẫu số được gọi là bậc tự do của sai số.

Chúng ta sẽ gọi sai số là se còn bậc tự do là dfE.

[1] Khoa Sư phạm kỹ thuật, Trường ĐHNNI

Sai số của trung bình là: $$s_{\bar{y}} = \frac{s_e}{\sqrt{r}}$$

Sai số của hiệu hai trung bình $$s(\bar{y}_1 - \bar{y}_2) = se\sqrt{\frac{2}{r}}$$

Dưới đây là một số cách so sánh các trung bình sau khi phân tích phương sai.

1. So sánh hai trung bình

So giá trị tuyệt đối của hiệu hai trung bình m_i và m_j với ngưỡng LSD (sai khác có ý nghĩa nhỏ nhất - Least significant difference)

$$LSD_{ij} = t(\alpha/2, dfE) \times \sqrt{msE\left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

trong đó α là mức ý nghĩa của kiểm định (tức xác suất đưa ra kết luận sai lầm: " hai trung bình khác nhau" khi hai trung bình thực sự không khác nhau), dfE là bậc tự do của sai số, $msE = s^2_e$, s_e là sai số thí nghiệm, r_i và r_j là các lần lặp của mức A_i và A_j

Nếu $|\overline{x_i} - \overline{x_j}| \leq LSD_{ij}$ kết luận hai trung bình m_i và m_j bằng nhau

Nếu $|\overline{x_i} - \overline{x_j}| > LSD_{ij}$ kết luận hai trung bình m_i và m_j khác nhau.

Khi số lần lặp bằng nhau (đều bằng r) LSD bằng

$$LSD = t(\alpha/2, dfE) \times \sqrt{msE \times \frac{2}{r}}$$

2. So sánh nhiều trung bình (multiple comparaison) khi số lần lặp bằng nhau

Nếu có a trung bình thì tất cả có a(a-1)/ 2 cặp trung bình cần so sánh. Có nhiều phương pháp, thường gọi là kiểm định (test), để so sánh và được chia thành các nhóm sau:

a. Một ngưỡng so sánh cho tất cả các cặp

Tất cả các cặp trung bình đều được so sánh theo cùng một cách: lấy giá trị tuyệt đối của hiệu hai trung bình rồi so với một ngưỡng. Nếu bé hơn ngưỡng thì coi như hai trung bình bằng nhau, ngược lại thì coi như khác nhau.

a1- Kiểm định LSD

Nếu ước lượng giá trị trung bình thì nửa chiều dài khoảng ước lượng bằng:

$$L = t(\alpha/2, dfE) \times \frac{se}{\sqrt{r}} = t \times s\bar{y} \qquad (1)$$

Nếu so sánh hai trung bình thì dùng ngưỡng kiểm định:

$$LSD = t(\alpha/2, dfE) \times \sqrt{msE \times \frac{2}{r}} = t \times s(\bar{y}_i - \bar{y}_j) \qquad (2)$$

Gọi hệ số t(α/2, dfE) ở hai công thức (1) và (2) là hệ số nhân t
Phương pháp LSD là phương pháp kinh điển được dùng từ lâu và quen thuộc với người dùng.

Các nghiên cứu sau này chứng tỏ nếu dùng kiểm định LSD để so sánh tất cả các cặp trung bình thì xác suất có kết luận sai: " hai trung bình khác nhau" khi hai trung bình thực sự không khác nhau sẽ không còn là α mà lớn hơn nhiều lần. Thí dụ khi so hai trung bình với α= 0,05 nếu có 4 mức, tức là 6 cặp trung bình thì xác suất có kết luận sai có thể lớn hơn 0,2.

Như vậy chỉ nên dùng LSD khi so sánh một cặp (hoặc một vài cặp) trung bình nào đó mà chúng ta có ý đồ so sánh khi thiết kế thí nghiệm chứ không nên dùng để so sánh tất cả các cặp trung bình sau khi xử lý dữ liệu.

a2- Kiểm định Scheffé

Kiểm định Scheffé cũng dùng (1) và (2), nhưng thay cho hệ số nhân t là hệ số nhân s

$$s = \sqrt{dfA \times F(\alpha, dfA, dfE)}$$

trong đó dfA là bậc tự do ứng với bình phương trung bình nằm ở tử số của F_{tn}, dfE là bậc tự do của sai số, F(α,dfA,dfE) là giá trị tới hạn trong phân phối Fisher- Snedecor.

Phương pháp Scheffé dùng để so sánh mọi cặp trung bình và còn mở rộng để kiểm định mọi tương phản (contrast).

Kiểm định Scheffé được gọi là bảo thủ vì ngưỡng so sánh quá lớn.

a3- Kiểm định HSD của Tukey (còn gọi là kiểm định Tukey - Cramer)

Kiểm định HSD (Honestly significant difference) dựa trên việc nghiên cứu cách so sánh theo phương pháp LSD giữa "trung bình nhỏ nhất và trung bình lớn nhất" để đưa ra hệ số nhân w thay cho hệ số **t** trong (1) và (2)

$$w = \frac{q(\alpha, a, dfE)}{\sqrt{2}}$$

q(α,p,dfE) được cho trong các bảng số thống kê với tên gọi bảng các phân vị trong phân phối của phạm vi kiểu Student (hay Student hoá).
(Selected percentiles of Studentized range distributions hayUpper percentage points of the Studentized range), trong đó:

α là mức ý nghĩa của kiểm định
p là tham số của phân phối (ở đây lấy p = a)
dfE là bậc tự do của sai số.

Ngưỡng tính theo HSD lớn hơn ngưỡng tính theo LSD và nhỏ hơn ngưỡng tính theo Scheffé.

a4- Kiểm định Bonferroni

Khi có k cặp trung bình cần so sánh thì dựa trên nhận xét về khuyết điểm của kiểm định LSD Bonferroni thay hệ số nhân t trong (1) và (2) bằng hệ số nhân t'

$$t = t(\alpha/2, dfE) \qquad t' = t(\alpha', dfE) \quad \text{với} \quad \alpha' = \alpha/2k$$

Thí dụ có 5 trung bình cần so sánh, tất cả có k = 5 x 4 / 2 = 10 cặp trung bình.

Giả sử $\alpha = 0{,}05$ và bậc tự do dfE = 65

$\alpha = 0{,}05$ $\qquad\qquad$ $\alpha' = 0{,}05/20 = 0{,}0025$ hay 0,25%

$t = t(0{,}025,65) = 1{,}997$ $\qquad\qquad$ $t' = t(0{,}0025,65) = 2{,}906$

Sau đây là thí dụ về so sánh 5 trung bình với số lần lặp r = 14; $\alpha = 0{,}05$; dfE = 65

$m_1 = 4{,}050$; $m_2 = 4{,}429$; m3 = 4,900; $m_4 = 5{,}343$; $m_5 = 5{,}921$; se = 1,633

Ngưỡng:

LSD $\qquad 1{,}997 \times 1{,}633 \times \sqrt{\frac{2}{14}} = 1{,}997 \times 0{,}6172 = 1{,}233$

Tukey Cramer $\qquad 3{,}975 \times \frac{1}{\sqrt{2}} \times 1{,}633 \times \sqrt{\frac{2}{14}} = 1{,}735$

Bonfferroni $\qquad 2{,}906 \times 1{,}633 \times \sqrt{\frac{2}{14}} = 1{,}794$

Scheffé $\qquad \sqrt{4 \times 2{,}513} \times 1{,}633 \times \sqrt{\frac{2}{14}} = 1{,}957$

b. Nhiều ngưỡng so sánh cho các cặp.

Nếu sắp xếp các trung bình từ nhỏ đến to thì có các kiểm định với nhiều ngưỡng so sánh, còn gọi là kiểm định đa phạm vi (multiple range test). Hai trung bình kề nhau dùng ngưỡng so sánh với tham số p = 2, hai trung bình cách nhau một (tức là ở giữa có một trung bình) dùng ngưỡng với p =3, cách nhau hai thì p = 4 . . .

b1- Kiểm định Student - Newman - Keuls (Gọi tắt là kiểm định S-N-K)

Các ngưỡng so sánh

$$W(p) = q(\alpha,p,dfE)\,\overline{sy} = q(\alpha,p,dfE) \times \sqrt{\frac{se^2}{r}}$$

trong đó q (α, p,dfE) là trị lấy trong bảng các phân vị trong phân phối của phạm vi kiểu Student (hay Student hoá) đã giới thiệu ở mục a3 (*Kiểm định Tukey Cramer*)

α là mức ý nghĩa, p là tham số, dfE là bậc tự do của sai số, $\overline{sy}$ là sai số của một trung bình.

Thí dụ với dfE = 24; $se^2 = 11{,}79$; r = 5 ta có $\quad \overline{sy} = \sqrt{\frac{11{,}79}{5}} = 1{,}536$

Các giá trị W(p) để so sánh 6 trung bình

p	2	3	4	5	6
q(0,025,p,24)	2,92	3,53	3,90	4,17	4,37
W(p)	4.5	5,4	6,0	6,4	6,7

Áp dụng vào 6 trung bình xếp từ nhỏ đến to ta có kết quả sau:

13,3 14,6 18,7 19,9 24,0 28,8

b2- Kiểm định Duncan

Kiểm định Duncan dùng công thức tương tự kiểm định S-N-K nhưng chọn mức ý nghĩa α' tuỳ theo a - số công thức cần so sánh

$$\alpha' = 1 - (1 - \alpha)^{a-1}$$

(thí dụ a = 3 $\alpha = 0{,}05$ $\alpha' = 0{,}0975$; a = 4 $\alpha' = 0{,}14$)

Duncan lập ra bảng số tương tự bảng các phân vị trong phân phối của phạm vi kiểu Student, từ đó tìm được ngưỡng so sánh

$$R(p) = q(\alpha',p,dfE)\ s\overline{y} = q(\alpha',p,dfE) \times \sqrt{\frac{se^2}{r}}$$

p	2	3	4	5	6
q(α',p,24)	2,92	3,07	3,15	3,22	3,28
R(p)	4,5	4,7	4,9	5,0	5,1

Áp dụng vào thí dụ trên cho kết quả tương tự kiểm định S-N-K

3. So sánh nhiều trung bình (multiple comparaison) khi số lần lặp không bằng nhau

Khi ước lượng đối với trung bình m_i theo LSD phải thay $s\overline{y}$ trong công thức (1) đối với nửa khoảng tin cậy L bằng $s\overline{y}_i$ để có công thức (1')

$$L = t(\alpha / 2, dfE) \times \frac{se}{\sqrt{r_i}} = t \times s\overline{y}_i \qquad (1')$$

Đối với các ước lượng theo Scheffé, Tukey Cramer, Bonferroni cũng làm tương tự.

Khi so sánh hai công thức có trung bình m_i và m_j với số lần lặp khác nhau thì phải thay 2/ r trong công thức tính $s(\overline{y}_i - \overline{y}_j)$ bằng $(1/ r_i + 1/ r_j)$ để có (2')

$$LSD = t(\alpha/2, dfE) \times \sqrt{msE \times (\frac{1}{r_i} + \frac{1}{r_j})} = t \times s(\overline{y}_i - \overline{y}_j) \qquad (2')$$

Đối với các kiểm định của Scheffé, Tukey Cramer, Bonferroni làm tương tự.

Đối với các kiểm định S-N-K và Duncan thì thay $\overline{sy} = \sqrt{\frac{se^2}{r}}$ bằng $\sqrt{se^2(\frac{1}{r_i} + \frac{1}{r_j})}$

Một số kiểm định khác

Kiểm định Dunnet

Khi so sánh các trung bình nếu có đối chứng thì nên dùng kiểm định Dunnett (có thể kiểm định một phía và hai phía).

Đối với kiểm định Dunnett thì dùng bảng số riêng để tính hệ số nhân t_D thay cho hệ số t trong công thức (2) của LSD. Nếu số lần lặp không bằng nhau thì thay 2/ r bằng (1/ r_i + 1/ r_j).

Ngoài các kiểm định một ngưỡng chung hoặc nhiều ngưỡng đã nêu trong các chương trình máy tính còn một số kiểm định khác như kiểm định Tukey b, Waller-Duncan, Hochberg GT2, Gabriel, Sidak, kiểm định F của REGW (Ryan-Einot-Gabriel-Welsch), kiểm đinh phạm vi của REGW.

Một số kiểm định có cách tiếp cận rất mới và không theo các lập luận quen thuộc trong thống kê kinh điển với các tính toán khá phức tạp.

Kiểm định khi các phương sai của các mức không bằng nhau

Một trong những giả thiết cơ bản của phân tích phương sai là các sai số phân phối chuẩn với phương sai bằng nhau σ^2. Khi phân tích phương sai phải kiểm định giả thiết các phương sai của các mức của nhân tố bằng nhau (dùng kiểm định Bartlett hay Levene).

Các kiểm định nêu ở các phần trên đều đòi hỏi các phương sai của các mức bằng nhau.

Trong các chương trình máy tính có các kiểm định không đòi hỏi các phương sai của các mức bằng nhau như kiểm định T2 của Tamhane, Dunnett T3 và Dunnett C, Games Howel.

Trong lúc chờ đợi các nghiên cứu mới để đánh giá kỹ hơn các kiểm định thì theo lời khuyên trong một số tài liệu nếu so sánh một số ít cặp trung bình đã có ý đồ so sánh từ trước có thể dùng kiểm định LSD, nếu so sánh với đối chứng thì dùng Dunnett, nếu so sánh tất cả các cặp trung bình thì dùng Tukey HSD. Nếu sắp xếp các trung bình theo thứ tự từ nhỏ đến lớn sau đó phân chia thành một số nhóm đồng nhất thì dùng kiểm định S-N-K hoặc Duncan, tuy nhiên cần chú ý là tuy hai kiểm định này được người sử dụng hoan nghênh vì đơn giản và sát thực tế nhưng lại không được các nhà nghiên cứu lý thuyết thống kê công nhận vì không chặt chẽ trong chứng minh.