

Ứng dụng phương pháp phân tích phương sai hai yếu tố trong điều tra chăn nuôi

VÕ THỊ LAN

Trong tạp chí *Phát triển kinh tế* số 173 tháng 2005 chúng ta đã được giới thiệu về ứng dụng phương pháp phân tích phương sai nghiên cứu mối liên hệ giữa các tiêu thức trong điều tra chăn nuôi: phải chăng các giống bò sữa khác nhau có ảnh hưởng đến năng suất sữa. Trong thực tế đơn giản là năng suất sữa không chỉ ảnh hưởng do bởi giống bò mà còn do rất nhiều nguyên nhân khác nữa. Loại thức ăn cung cấp cũng không kém phần quan trọng, dưới đây là kết quả nghiên cứu năng suất sữa chịu ảnh hưởng đồng thời cả 2 nhân tố: giống bò và loại thức ăn.

Nghiên cứu được thực hiện trên mẫu gồm 27 con bò đang vắt sữa cùng độ tuổi được chọn ngẫu nhiên từ ba giống bò sữa khác nhau: bò lai Hà Lan, bò lai Sind và bò giống khác (mỗi loại 9 con) và được cho ăn ba loại thức ăn khác nhau: tinh, vật chất khô (cỏ ủ, rơm khô, ...), khác (hèm, khoai mì, ...) (mỗi loại thức ăn 9 con gồm: 3 bò lai Hà Lan, 3 bò lai Sind và 3 giống khác).

Sản lượng sữa trung bình mỗi ngày của một bò cái đang vắt sữa phân nhóm theo loại bò và theo loại thức ăn (lít) như trong bảng dưới đây:

Thức ăn	Giống bò								
	Bò lai Hà Lan			Bò lai Sind			Bò giống khác		
tinh	6,9	7,3	7,5	7,1	7,5	7,2	6,5	6,7	6,6
VC khô	6,4	6,8	7,0	7,1	6,8	6,7	6,7	6,5	6,4
khác	6,5	5,9	6,4	6,3	6,5	6,6	5,4	5,3	5,1

Với bài toán phân tích phương sai hai yếu tố, ta có ba giả thuyết sau:

1. Sản lượng sữa trung bình mỗi ngày của một bò cái thuộc các giống bò khác nhau đều bằng nhau.
2. Sản lượng sữa trung bình mỗi ngày của một bò cái ăn các loại thức ăn khác nhau đều bằng nhau.
3. Không có ảnh hưởng tương tác giữa các giống bò và các loại thức ăn đến sản lượng sữa trung bình mỗi ngày của một bò cái. Nói cách khác: ảnh hưởng của giống bò đến sản lượng sữa trung bình mỗi ngày của một bò cái là như nhau đối với các loại thức ăn và ảnh hưởng của các loại thức ăn là như nhau đối với sản lượng sữa trung bình mỗi ngày của một bò cái thuộc các giống bò khác nhau

Bước 1: Tính các trung bình mẫu của từng nhóm bò và từng nhóm thức ăn

- a. Sản lượng sữa trung bình mỗi ngày của nhóm bò (lít):

$$\begin{aligned} \text{- bò lai Hà Lan : } x_1 &= \frac{6,9 + 7,3 + \dots + 6,4}{9} = 6,74 \\ \text{- bò lai Sind: } x_2 &= \frac{7,1 + 7,5 + \dots + 6,6}{9} = 6,87 \\ \text{- bò khác: } x_3 &= \frac{6,5 + 6,7 + \dots + 5,1}{9} = 6,13 \end{aligned}$$

- b. Sản lượng sữa trung bình mỗi ngày của các nhóm bò theo từng loại thức ăn (lít)

$$\begin{aligned} \text{- thức ăn tinh } S_1 &= \frac{7,1 + 7,5 + \dots + 6,6}{9} = 7,03 \\ \text{- thức ăn VC khô } S_2 &= \frac{7,1 + 6,8 + \dots + 6,4}{9} = 6,71 \\ \text{- thức ăn khác } S_3 &= \frac{6,3 + 6,5 + \dots + 5,1}{9} = 6,0 \end{aligned}$$

- c. Sản lượng sữa trung bình mỗi ngày của nhóm bò lai Hà Lan ăn thức ăn tinh (lít):

$$x_{11} = \frac{7,1 + 7,5 + 7,2}{3} = 7,27$$

- Sản lượng sữa trung bình mỗi ngày của nhóm bò lai Hà Lan ăn thức ăn VC khô (lít):

$$x_{21} = \frac{7,1 + 6,8 + 6,7}{3} = 6,87$$

$$\dots \dots \dots x_{33} = \frac{5,4 + 5,3 + 5,1}{3} = 5,27$$

- d. Sản lượng sữa trung bình mỗi ngày của một con bò tính chung cho cả ba giống bò và cả ba loại thức ăn (lít):

$$\bar{x} = \frac{7,1 + 7,5 + 7,2 + \dots + 5,4 + 5,3 + 5,1}{9} = 6,58$$

Bước 2: Tính các tổng độ lệch bình phương

- a. Tổng các độ lệch bình phương toàn bộ (SST)

$$SST = (7,1 - 6,58)^2 + (7,5 - 6,58)^2 + \dots + (6,5 - 6,58)^2 + (6,7 - 6,58)^2 + (6,6 - 6,58)^2 + \dots + (6,3 - 6,58)^2 + (6,5 - 6,58)^2 + (6,6 - 6,58)^2 + \dots + (5,4 - 6,58)^2 + (5,3 - 6,58)^2 + (5,1 - 6,58)^2 = 9,441$$

- b. Tổng các độ lệch bình phương giữa các nhóm (SSG)

$$SSG = 3 \times 3 \times [(6,87 - 6,58)^2 + (6,74 - 6,58)^2 + (6,13 - 6,58)^2] = 2,779$$

- c. Tổng các độ lệch bình phương giữa các khối (SSB)

$$SSB = 3 \times 3 \times [(7,03 - 6,58)^2 + (6,71 - 6,58)^2 + (6,0 - 6,58)^2] = 5,032$$

- d. Tổng các độ lệch bình phương giữa các ô (SSI)

$$SSI = 3 \times [(7,27 - 6,87 - 5,27 + 6,58)^2 + (6,27 - 6,74 - 6,00 + 6,58)^2 + \dots + (7,23 - 6,74 - 7,03 + 6,58)^2 + (6,60 - 6,13 - 7,03 + 6,58)^2] = 0,717$$

- e. Tổng các độ lệch bình phương phần dư (SSE)

$$SSE = SST - SSG - SSB - SSI = 9,441 - 2,779 - 5,032 - 0,717 = 0,913$$

Bước 3: Tính các phương sai

- a. Phương sai giữa các nhóm (MSG)

$$MSG = \frac{SSG}{K - 1} = \frac{2,779}{3 - 1} = 1,389$$

- b. Phương sai giữa các khối (MSB)

$$MSB = \frac{SSB}{H - 1} = \frac{5,032}{3 - 1} = 2,516$$

c. Phương sai giữa các ô (MSI)

$$MSI = \frac{SSI}{(K-1)(H-1)} = \frac{0,717}{(3-1)(3-1)} = 0,179$$

d. Phương sai dư (MSE)

$$MSE = \frac{SSE}{KxHx(L-1)} = \frac{0,913}{3x3x(3-1)} = 0,0507$$

Bước 4: Tính tỉ số F

$$a. F_1 = \frac{MSG}{MSE} = \frac{1,389}{0,0507} = 27,4$$

$$b. F_2 = \frac{MSB}{MSE} = \frac{2,516}{0,0507} = 49,6$$

$$c. F_3 = \frac{MSI}{MSE} = \frac{0,179}{0,0507} = 3,53$$

Tra bảng phân phối F với mức ý nghĩa $\alpha=0,05$, ta có kết quả các F tra bảng lần lượt là:

$$F_{K-1, KH(L-1), \alpha} = F_{2, 18, 0,05} = 3,55$$

$$F_{H-1, KH(L-1), \alpha} = F_{2, 18, 0,05} = 3,55$$

$$F_{(K-1)(H-1), KH(L-1), \alpha} = F_{4, 18, 0,05} = 2,93$$

So sánh các tỉ số F tính ra với các F tra bảng, ta có các kết quả sau:

- Vì $F_1 > F_{K-1, KH(L-1), \alpha} = 3,55$, nên chúng ta có đủ bằng chứng để bác bỏ giả thuyết H_0 thứ nhất. Do đó, sản lượng sữa trung bình mỗi ngày của một bò cái thuộc các giống bò khác nhau thì không bằng nhau.

- Vì $F_2 > F_{H-1, KH(L-1), \alpha} = 3,55$, nên chúng ta có đủ bằng chứng để bác bỏ giả thuyết H_0 thứ hai. Do đó: Sản lượng sữa trung bình mỗi ngày của một bò cái ăn các loại thức ăn khác nhau thì không bằng nhau.

- Vì $F_3 > F_{(K-1)(H-1), KH(L-1), \alpha} = 2,93$, nên chúng ta có đủ bằng chứng để bác bỏ giả thuyết H_0 thứ ba. Do đó, có ảnh hưởng tương tác giữa các giống bò và các loại thức ăn đến sản lượng sữa trung bình mỗi ngày của một bò cái. Nói cách khác: Ảnh hưởng của giống bò đến sản lượng sữa trung bình mỗi ngày của một bò cái là như nhau đối với các loại thức ăn và ảnh hưởng của các loại thức ăn là như nhau đối với sản lượng sữa trung bình mỗi ngày của một bò cái thuộc các giống bò khác nhau.

Tương tự như phân tích phương sai một yếu tố, ta có thể thực hiện phân tích phương sai hai yếu tố nhanh chóng hơn bằng cách sử dụng các phần mềm như Excel hoặc SPSS. Dưới đây là kết quả phân tích phương sai hai yếu tố được thực hiện trên Excel:

Anova: Two-Factor With Replication

SUMMARY	Bò lai Hà Lan	Bò lai Sind	Bò giống khác	Total
<i>tinh</i>				
Count	3	3	3	9
Sum	21,7	21,8	19,8	63,3
Average	7,23	7,27	6,60	7,03
Variance	0,09	0,04	0,01	0,14
<i>Vật chất khô</i>				
Count	3	3	3	9
Sum	20,2	20,6	19,6	60,4
Average	6,73	6,87	6,53	6,71
Variance	0,09	0,04	0,02	0,06
<i>khác</i>				

Count	3	3	3	9
Sum	18,8	19,4	15,8	54
Average	6,27	6,47	5,27	6,00
Variance	0,10	0,02	0,02	0,35
<i>Total</i>				
Count	3	3	3	
Sum	60,7	61,8	55,2	
Average	6,74	6,87	6,13	
Variance	0,25	0,15	0,44	

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	5,0319	2	2,5159	49,5839	0,0000	3,5546
Columns	2,7785	2	1,3893	27,3796	0,0000	3,5546
Interaction	0,7170	4	0,1793	3,5328	0,0269	2,9277
Within	0,9133	18	0,0507			
Total	9,440740741	26				

Trên bảng kết quả này, cột F chính là kết quả tính toán tỉ số F và cột F crit chính là các giá trị F tra bảng với mức ý nghĩa $\alpha = 0,05$.

Tương tự như phân tích phương sai một yếu tố, trong phân tích phương sai hai yếu tố sau khi xác định có sự khác biệt giữa các nhóm so sánh, ta có thể dùng kiểm định Tukey để xác định các cặp trung bình tổng thể khác nhau xét theo yếu tố thứ nhất (so sánh giữa K nhóm) hay xét theo yếu tố thứ hai (so sánh giữa H khối).

Với $\alpha = 0,05$, $K = 3$, $H = 3$, $L = 3$, $MSE = 0,0507$, tra bảng phân phối Tukey, ta có: $q_{0,05,3,18} = 3,61$

a. So sánh giữa các nhóm theo giống bò, giá trị Tukey được tính như sau:

$$T = q_{0,05,3,18} \sqrt{\frac{0,0507}{9}} = 0,27$$

Độ lệch tuyệt đối các cặp trung bình nhóm

$$D_1 = |\bar{x}_1 - \bar{x}_2| = |6,74 - 6,87| = 0,13$$

$$D_2 = |\bar{x}_1 - \bar{x}_3| = |6,74 - 6,13| = 0,61$$

$$D_3 = |\bar{x}_2 - \bar{x}_3| = |6,87 - 6,13| = 0,74$$

Kết quả D_2, D_3 lớn hơn giá trị giới hạn Tukey T, cho nên ta có thể nói rằng bò lai Sind và các giống bò khác có sản lượng sữa trung bình khác với giống bò lai Hà Lan, nhưng chưa có bằng chứng để kết luận sản lượng sữa trung bình là khác nhau giữa giống bò lai Hà Lan và giống bò Sind.

b. So sánh giữa các nhóm theo loại thức ăn, giá trị Tukey được tính như sau:

$$T = q_{0,05,3,18} \sqrt{\frac{0,0507}{9}} = 0,27$$

Độ lệch tuyệt đối các cặp trung bình nhóm

$$D_1 = |\bar{x}_1 - \bar{x}_2| = |7,03 - 6,71| = 0,32$$

$$D_2 = |\bar{x}_1 - \bar{x}_3| = |7,03 - 6,00| = 1,03$$

$$D_3 = |\bar{x}_2 - \bar{x}_3| = |6,71 - 6,00| = 0,71$$

Cả ba kết quả D_1, D_2, D_3 đều lớn hơn giá trị giới hạn Tukey T, vì vậy ta có thể nói rằng các loại thức ăn khác nhau sẽ cho sản lượng sữa trung bình khác nhau.

Một lần nữa chúng ta đã thấy được sự vận dụng của phương pháp phân tích phương sai để nghiên cứu mối liên hệ giữa các tiêu thức – một phương pháp được sử dụng khá phổ biến trong nghiên cứu kinh tế xã hội ■