

BỔ SUNG VỀ PHÂN TÍCH PHƯƠNG SAI

1- MÔ HÌNH TẮT ĐỊNH, NGẪU NHIÊN VÀ HỖN HỢP

Trong các giáo trình phương pháp thí nghiệm và toán sinh học khi phân tích phương sai các nhân tố thường được coi là **tắt định (cố định)**. Việc phân tích và kết luận được trình bày theo các mẫu định sẵn đã quen thuộc với cán bộ giảng dạy và sinh viên trong các trường.

Những năm gần đây, do đòi hỏi của thực tế và cũng do những tiến bộ của việc nghiên cứu và phân tích các thiết kế thí nghiệm, các nhân tố có thể tắt định hay ngẫu nhiên và mô hình phân tích phương sai được sắp thành 3 loại: **tắt định** (Fixed - nếu tất cả các nhân tố đều tắt định), **ngẫu nhiên** (Random - nếu tất cả các nhân tố đều ngẫu nhiên) và **hỗn hợp** (Mixed - nếu một số nhân tố tắt định, một số ngẫu nhiên).

Việc quyết định xem nhân tố tắt định hay ngẫu nhiên phải làm trước khi bố trí thí nghiệm và căn cứ vào bản chất của nhân tố cũng như ảnh hưởng của kết luận khi ứng dụng trong thực tế.

Sau đây là một số mô hình thường dùng:

1.1 Phân tích phương sai một nhân tố

Mô hình một nhân tố có a mức, mỗi mức lặp lại r_i lần

$$x_{ij} = \mu + a_i + e_{ij} \quad (i = 1, a; j = 1, r_i)$$

μ là trung bình chung

a_i là tác động của mức A_i

e_{ij} là sai số ngẫu nhiên giả thiết độc lập, phân phối chuẩn $N(0, \sigma_e^2)$

Nếu nhân tố A cố định thì mô hình gọi là mô hình tắt định, các a_i là các hằng số thoả mãn điều kiện

$$\sum_{i=1}^a a_i = 0$$

Nếu nhân tố A ngẫu nhiên thì mô hình gọi là mô hình ngẫu nhiên, các a_i là các giá trị của biến ngẫu nhiên phân phối chuẩn $N(0, \sigma_a^2)$

Phương pháp phân tích

Cả hai mô hình đều chung cách phân tích mà nội dung gồm:



a) Tách tổng bình phương toàn bộ SSTO thành hai phần: tổng bình phương do nhân tố SSA và tổng bình phương do sai số SSE

b) Tách bậc tự do của tổng bình phương toàn bộ dfTO thành hai phần: bậc tự do dfA của tổng bình phương SSA và bậc tự do dfE của tổng bình phương SSE

c) Chia tổng bình phương cho bậc tự do được bình phương trung bình msA, msE

d) Tóm tắt toàn bộ cách phân tích vào trong bảng : $n = \sum r_i$

Bảng phân tích phương sai (ANOVA)

Nguồn biến động	Bậc tự do	Tổng BP	BP trung bình	F_{tn}	Kỳ vọng
Nhân tố	$a - 1$ dfA	SSA	$msA = SSA/dfA$	msA/msE	
Sai số	$n - a$ dfE	SSE	$msE = SSE/dfE$		σ_e^2
Toàn bộ	$n - 1$	SSTO			

Cách kiểm định giả thiết

Tuỳ theo mô hình có thể tính các kỳ vọng của msA và msE. Từ đó có cách tính tỷ số F_{tn} và cách kiểm định giả thiết đối với nhân tố A.

Mô hình tất định

Giả thiết H_0 : “các a_i bằng không”, đối thiết: “có a_i khác không”.

Kỳ vọng của msE bằng σ_e^2 , còn kỳ vọng của msA bằng $\sigma_e^2 + \Phi_A$

$$\text{trong đó } \Phi_A = \frac{\sum_{i=1}^a r_i a_i^2}{(a-1)} \quad \text{nếu mọi } r_i \text{ đều bằng } r \text{ thì } \Phi_A = \frac{r \sum_{i=1}^a a_i^2}{a-1}$$

Nếu giả thiết H_0 đúng thì tỷ số $F_{tn} = msA/msE$ phân phối Fiszơ Snêđêcô với $a-1$ và $n-a$ bậc tự do và ta có quy tắc kiểm định:

Tìm giá trị tới hạn $F(\alpha, a-1, n-a)$.

Nếu $F_{tn} \leq F(\alpha, a-1, n-a)$ thì chấp nhận H_0 , nếu ngược lại thì chấp nhận H_1 .

Mô hình ngẫu nhiên

Giả thiết H_0 : “ σ_A^2 bằng không”, đối thiết: “ σ_A^2 khác không”.

Kỳ vọng của msE bằng σ_e^2 , còn kỳ vọng của msA bằng $\sigma_e^2 + r'\sigma_A^2$

$$\text{với } r' = \frac{N^2 - \sum_{i=1}^k r_i^2}{(a-1)N} \quad \text{nếu mọi } r_i \text{ đều bằng } r \text{ thì } r' = r$$

Nếu giả thiết H_0 đúng thì tỷ số $F_{tn} = msA / msE$ phân phối Fisor Snêđêco với $a-1$ và $n-a$ bậc tự do và ta có quy tắc kiểm định:

Tìm giá trị tới hạn $F(\alpha, a-1, n-a)$, nếu $F_{tn} \leq F(\alpha, a-1, n-a)$ thì chấp nhận H_0 , nếu ngược lại thì chấp nhận H_1

σ^2_e được ước lượng bằng msE

Nếu $msA > msE$ thì σ^2_A được ước lượng bằng $(msA - msE) / r'$

1.2 Phân tích phương sai hai nhân tố chéo nhau (crossed)

Mô hình hai nhân tố chéo nhau (hay trực giao)

Nhân tố A có a mức, nhân tố B có b mức, mỗi công thức (a_i, b_j) lặp lại r lần

$$x_{ijk} = \mu + a_i + b_j + ab_{ij} + e_{ijk} \quad (i = 1, a; j = 1, b; k = 1, r)$$

μ là trung bình chung, a_i là tác động của mức A_i của nhân tố A

b_j là tác động của mức B_j của nhân tố B

$(ab)_{ij}$ là tương tác giữa 2 mức A_i và B_j của hai nhân tố A, B

e_{ijk} là sai số ngẫu nhiên giả thiết độc lập, phân phối chuẩn $N(0, \sigma^2_e)$.

Nếu nhân tố A và B tất định thì mô hình gọi là mô hình tất định, các a_i và b_j thỏa mãn các hệ thức:

$$\sum_{i=1}^a a_i = 0 \quad \sum_{j=1}^b b_j = 0 \quad \sum_{i=1}^a ab_{ij} = 0 \quad \sum_{j=1}^b ab_{ij} = 0$$

Nếu A và B ngẫu nhiên thì mô hình gọi là mô hình ngẫu nhiên, các a_i là các giá trị của biến ngẫu nhiên phân phối chuẩn $N(0, \sigma^2_a)$, các b_j là các giá trị của biến ngẫu nhiên phân phối chuẩn $N(0, \sigma^2_b)$ còn $(ab)_{ij}$ là các giá trị của biến ngẫu nhiên phân phối chuẩn $N(0, \sigma^2_{ab})$

Nếu một trong 2 nhân tố tất định, nhân tố kia ngẫu nhiên thì có mô hình hỗn hợp.

Phương pháp phân tích

Cả ba mô hình đều chung cách phân tích mà nội dung gồm:

a) Tách tổng bình phương toàn bộ SSTO thành bốn phần: tổng bình phương do nhân tố A (SSA), tổng bình phương do nhân tố B (SSB), tổng bình phương do tương tác AB (SSAB) và tổng bình phương do sai số (SSE).

b) Tách bậc tự do của tổng bình phương toàn bộ dfTO thành bốn phần: bậc tự do dfA của tổng bình phương SSA, bậc tự do dfB của tổng bình phương SSB, bậc tự do dfAB của tổng bình phương SSAB và bậc tự do dfE của tổng bình phương SSE

c) Chia tổng bình phương cho bậc tự do được bình phương trung bình ms_A , ms_B , ms_{AB} và ms_E

e) Tóm tắt toàn bộ cách phân tích vào trong bảng :

Bảng phân tích phương sai

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương trung bình	F_{tn}	Kỳ vọng
Nhân tố A	$df_A = a-1$	SSA	ms_A	F_{tnA}	
Nhân tố B	$df_B = b-1$	SSB	ms_B	F_{tnB}	
Tương tác	$df_{AB} = (a-1)(b-1)$	SSAB	ms_{AB}	F_{tnAB}	
Sai số	$df_E = ab(r-1)$	SSE	ms_E		σ^2_e
Toàn bộ	$abr - 1$	SSTO			

Cách kiểm định giả thiết

Tuỳ theo mô hình có thể tính các kỳ vọng của ms_A , ms_B , ms_{AB} và ms_E . Từ đó có cách tính tỷ số F_{tn} và cách kiểm định các giả thiết đối với nhân tố A, nhân tố B và tương tác AB.

Mô hình tất định

Giả thiết H_{0A} : Các a_i bằng không, đối thiết H_{1A} : có a_i khác không

Giả thiết H_{0B} : Các b_j bằng không, đối thiết H_{1B} : có b_j khác không

Giả thiết H_{0AB} : Các ab_{ij} bằng không, đối thiết H_{1AB} : có ab_{ij} khác không

Kỳ vọng	Kiểm định giả thiết
$E(ms_A) = \sigma^2_e + (br \sum a_i^2) / (a-1)$	$F_{tnA} = ms_A / ms_E$ so với $F(\alpha, df_A, df_E)$
$E(ms_B) = \sigma^2_e + (ar \sum b_j^2) / (b-1)$	$F_{tnB} = ms_B / ms_E$ so với $F(\alpha, df_B, df_E)$
$E(ms_{AB}) = \sigma^2_e + (r \sum \sum ab_{ij}^2) / ((a-1)(b-1))$	$F_{tnAB} = ms_{AB} / ms_E$ so với $F(\alpha, df_{AB}, df_E)$
$E(ms_E) = \sigma^2_e$	

Mô hình ngẫu nhiên

Giả thiết H_{0A} : σ^2_A bằng không, đối thiết H_{1A} : σ^2_A khác không

Giả thiết H_{0B} : σ^2_B bằng không, đối thiết H_{1B} : σ^2_B khác không

Giả thiết H_{0AB} : σ^2_{AB} bằng không, đối thiết H_{1AB} : σ^2_{AB} khác không

Kỳ vọng	Kiểm định giả thiết
$E(msA) = \sigma_e^2 + r\sigma_{AB}^2 + br\sigma_A^2$	$F_{tnA} = msA/msAB$ so với $F(\alpha, dfA, dfAB)$
$E(msB) = \sigma_e^2 + r\sigma_{AB}^2 + ar\sigma_B^2$	$F_{tnB} = msB/msAB$ so với $F(\alpha, dfB, dfAB)$
$E(msAB) = \sigma_e^2 + r\sigma_{AB}^2$	$F_{tnAB} = msAB/msE$ so với $F(\alpha, dfAB, dfE)$
$E(msE) = \sigma_e^2$	

Có thể ước lượng các phương sai như sau:

σ_e^2 ước lượng bằng msE

σ_{AB}^2 ước lượng bằng $(msAB - msE)/r$

σ_B^2 ước lượng bằng $(msB - msAB)/ar$

σ_A^2 ước lượng bằng $(msA - msAB)/br$

Mô hình hỗn hợp

Giả sử nhân tố A tất định, nhân tố B ngẫu nhiên (kéo theo AB ngẫu nhiên)

Giả thiết H_{0A} : Các a_i bằng không, đối thiết H_{1A} : có a_i khác không

Giả thiết H_{0B} : σ_B^2 bằng không, đối thiết H_{1B} : σ_B^2 khác không

Giả thiết H_{0AB} : σ_{AB}^2 bằng không, đối thiết H_{1AB} : σ_{AB}^2 khác không

Kỳ vọng	Kiểm định giả thiết
$E(msA) = \sigma_e^2 + r\sigma_{AB}^2 + (br\sum a_i^2)/(a-1)$	$F_{tnA} = msA/ msAB$ so với $F(\alpha, dfA, dfAB)$
$E(msB) = \sigma_e^2 + ar\sigma_B^2$	$F_{tnB} = msB/ msE$ so với $F(\alpha, dfB, dfE)$
$E(msAB) = \sigma_e^2 + r\sigma_{AB}^2$	$F_{tnAB} = msAB/ msE$ so với $F(\alpha, dfAB, dfE)$
$E(msE) = \sigma_e^2$	

1.3 Phân tích phương sai hai nhân tố phân cấp

Mô hình hai nhân tố phân cấp (chia thứ bậc(Hierachical) còn gọi là chia ổ hay lồng vào nhau(Nested))

Nhân tố A là cấp trên có a mức, nhân tố B là cấp dưới có b mức, mỗi công thức(a_i, b_j) lặp lại r lần

$$x_{ijk} = \mu + a_i + b_{j(i)} + e_{ijk} \quad (i = 1, a; j = 1, b; k = 1, r)$$

μ là trung bình chung

a_i là tác động của mức A_i của nhân tố A

$b_{j(i)}$ là tác động của mức B_j (dưới mức i của nhân tố A) của nhân tố B

e_{ijk} là sai số ngẫu nhiên giả thiết độc lập, phân phối chuẩn $N(0, \sigma_e^2)$.

Nếu các mức A_i và B_j ngẫu nhiên thì mô hình gọi là mô hình ngẫu nhiên, nếu A tất định B ngẫu nhiên thì có mô hình hỗn hợp.

Trong mô hình ngẫu nhiên nhân tố A ngẫu nhiên, các a_i là các giá trị của biến chuẩn $N(0, \sigma_A^2)$. Nhân tố B ngẫu nhiên, các b_j trong cùng một mức i của nhân tố A là các giá trị của biến chuẩn $N(0, \sigma_B^2)$.

Trong mô hình hỗn hợp nhân tố A tất định, các a_i thoả mãn điều kiện

$$\sum_{i=1}^a a_i = 0$$

Nhân tố B ngẫu nhiên, các b_j trong cùng một mức i của nhân tố A là các giá trị của biến chuẩn $N(0, \sigma_B^2)$.

Phương pháp phân tích

Cả hai mô hình đều chung cách phân tích mà nội dung gồm:

a) Tách tổng bình phương toàn bộ SSTO thành ba phần: tổng bình phương do nhân tố A (SSA), tổng bình phương do nhân tố B trong A (SSB(A)) và tổng bình phương do sai số (SSE).

b) Tách bậc tự do của tổng bình phương toàn bộ dfTO thành ba phần: bậc tự do dfA của tổng bình phương SSA, bậc tự do dfB(A) của tổng bình phương SSB(A) và bậc tự do dfE của tổng bình phương SSE

c) Chia tổng bình phương cho bậc tự do được bình phương trung bình msA, msB và msE

f) Tóm tắt toàn bộ cách phân tích vào trong bảng

Nguồn	Bậc tự do	Tổng bình phương	Bình phương trung bình	F_{tn}	Kỳ vọng
Nhân tố A	dfA = a-1	SSA	msA	F_{tnA}	
Nhân tố B	dfB(A) = a(b-1)	SSB(A)	msB(A)	F_{tnB}	
Sai số	dfE = ab(r-1)	SSE	msE		
Toàn bộ	dfTO = abr-1	SSTO			σ_e^2

Cách kiểm định giả thiết

Mô hình ngẫu nhiên

Giả thiết H_{0A} : σ_A^2 bằng không, đối thiết H_{1A} : σ_A^2 khác không

Giả thiết H_{0B} : σ_B^2 bằng không, đối thiết H_{1B} : σ_B^2 khác không

Kỳ vọng	Kiểm định giả thiết
$E(msA) = \sigma_e^2 + r\sigma_B^2 + br\sigma_A^2$	$F_{tnA} = msA / msB(A)$ so với $F(\alpha, dfA, dfB(A))$
$E(msB(A)) = \sigma_e^2 + r\sigma_B^2$	$F_{tnB} = msB(A) / msE$ so với $F(\alpha, dfB(A), dfE)$
$E(msE) = \sigma_e^2$	

Ước lượng các phương sai:

σ_e^2 được ước lượng bằng msE

σ_B^2 được ước lượng bằng $(msB(A) - msE) / r$

σ_A^2 được ước lượng bằng $(msA - msB(A)) / br$

Mô hình hỗn hợp

Giả thiết H_{0A} : các a_i bằng không, đối thiết H_{1A} : có a_i khác không

Giả thiết H_{0B} : σ_B^2 bằng không, đối thiết H_{1B} : σ_B^2 khác không

Kỳ vọng	Kiểm định giả thiết
$E(msA) = \sigma_e^2 + r\sigma_B^2 + \Phi_A$	$F_{tnA} = msA / msB(A)$ so với $F(\alpha, dfA, dfB(A))$
$E(msB) = \sigma_e^2 + r\sigma_B^2$	$F_{tnB} = msB(A) / msE$ so với $F(\alpha, dfB(A), dfE)$
$E(msE) = \sigma_e^2$	

$$\text{với } \Phi_A = \frac{br \sum_{i=1}^a a_i^2}{a-1}$$

Có thể ước lượng các phương sai:

σ_e^2 được ước lượng bằng msE

σ_B^2 được ước lượng bằng $(msB - msE) / r$

Mô hình tất định

Khi cả A và B đều tất định thì có mô hình tất định (rất ít gặp), cách tính như hai mô hình ngẫu nhiên và hỗn hợp,.

Giả thiết H_{0A} : các a_i bằng không, đối thiết H_{1A} : có a_i khác không

Giả thiết H_{0B} : các $b_{j(i)}$ bằng không, đối thiết H_{1A} : có $b_{j(i)}$ khác không với mọi i

Hai giá trị F thực nghiệm đều có mẫu số là sai số msE (se^2)

$F_{tnA} = msA / msE$

$F_{tnB(A)} = msB(A) / msE$

1.4 Phân tích phương sai hai nhân tố chia ô (Split plot)

Mô hình hai nhân tố chia ô

Trong bố trí thí nghiệm chia ô lớn, ô nhỏ (Split plot) nhân tố A bố trí trên ô lớn, nhân tố B bố trí trên ô nhỏ, mỗi lần lặp là một khối (Block)

$$x_{ijk} = \mu + c_k + a_i + cd_{ik} + b_j + ab_{ij} + e_{ijk} \quad (i = 1, a; j = 1, b; k = 1, r)$$

μ là trung bình chung, c_k là tác động của khối k

a_i là tác động của mức A_i của nhân tố A

cd_{ik} là tương tác giữa khối và nhân tố A

b_j là tác động của mức B_j của nhân tố B

ab_{ij} là tác động của tương tác AB

e_{ijk} là sai số ngẫu nhiên giả thiết độc lập, phân phối chuẩn $N(0, \sigma^2_e)$.

Khối thường được coi là nhân tố ngẫu nhiên

Các c_k phân phối chuẩn $N(0, \sigma^2_K)$ còn A và B thì có thể tất định hoặc ngẫu nhiên, nếu A tất định thì thỏa mãn điều kiện $\sum a_i = 0$ (B cố định thì $\sum b_j = 0$) còn nếu A ngẫu nhiên thì các giá trị a_i phân phối chuẩn $N(0, \sigma^2_A)$ (B ngẫu nhiên thì b_j phân phối chuẩn $N(0, \sigma^2_B)$).

Thường lấy tương tác A x K làm sai số ô lớn và bỏ qua tương tác B x K.

Tuỳ giả thiết hai nhân tố tất định hay ngẫu nhiên hay hỗn hợp mà có cách kiểm định khác nhau.

Phương pháp phân tích

Các mô hình đều chung cách phân tích mà nội dung gồm:

a) Tách tổng bình phương toàn bộ SSTO thành sáu phần: tổng bình phương do khối SSK, tổng bình phương do nhân tố A (SSA), tổng bình phương do tương tác AK (SSAK), tổng bình phương do nhân tố B (SSB), tổng bình phương do tương tác AB (SSAB), và tổng bình phương do sai số SSE.

b) Tách bậc tự do của tổng bình phương toàn bộ dfTO thành sáu phần: bậc tự do dfK của tổng bình phương SSK, bậc tự do dfA của tổng bình phương SSA, bậc tự do dfAK của tổng bình phương SSAK, bậc tự do dfB của tổng bình phương SSB, bậc tự do dfAB của tổng bình phương SSAB và bậc tự do dfE của tổng bình phương SSE

c) Chia tổng bình phương cho bậc tự do được bình phương trung bình msK, msA, msAK, msB, msAB và msE

d) Tóm tắt kết quả vào bảng phân tích phương sai

Bảng phân tích phương sai

Nguồn biến động	Bậc tự do	Tổng bình phương	Bình phương trung bình	F_{tn}	Kỳ vọng
Khối	dfK = r-1	SSK	msK		
Nhân tố A	dfA = a-1	SSA	msA	F _{tnA}	
Tương tác AK Sai số ô lớn	dfAK = (a-1)(r-1)	SSAK	msAK		
Nhân tố B	dfB = b-1	SSB	msB	F _{tnB}	
Tương tác AB	dfAB = (a-1)(b-1)	SSAB	msAB	F _{tnAB}	
Sai số ô nhỏ	dfE = a(b-1)(r-1)	SSE	msE		σ^2_e
Toàn bộ	abr -1	SSTO			

Cách kiểm định giả thiết

Mô hình tất định

Giả thiết H_{0A}: Các a_i bằng không, đối thiết H_{1A}: có a_i khác không

Giả thiết H_{0B}: Các b_j bằng không, đối thiết H_{1B}: có b_j khác không

Giả thiết H_{0AB}: Các ab_{ij} bằng không, đối thiết H_{1AB}: có ab_{ij} khác không

Kỳ vọng	Kiểm định giả thiết
Ô lớn	
$E(msK) = \sigma^2_e + ab\sigma^2_K$ $E(msA) = \sigma^2_e + b \sigma^2_{AK} + (br\sum a_i^2)/(a-1)$ $E(msAK) = \sigma^2_e + b \sigma^2_{AK}$	F _{tnA} =msA/ msAK so với F(α,dfA,dfAK)
Ô nhỏ	
$E(msB) = \sigma^2_e + (ar\sum b_j^2)/(b-1)$ $E(msAB) = \sigma^2_e + (r\sum \sum ab_{ij}^2)/(a-1)((b-1)$ $E(msE) = \sigma^2_e$	F _{tnB} =msB/ msE so với F(α,dfB,dfE) F _{tnAB} =msAB/ msE so với F(α,dfAB,dfE)

Mô hình ngẫu nhiên: nhân tố A và nhân tố B ngẫu nhiên

Giả thiết H_{0A}: σ^2_A bằng không, đối thiết H_{1A}: σ^2_A khác không

Giả thiết H_{0B}: σ^2_B bằng không, đối thiết H_{1B}: σ^2_B khác không

Giả thiết H_{0AB} : σ^2_{AB} bằng không, đối thiết H_{1AB} : σ^2_{AB} khác không

Kỳ vọng	Kiểm định giả thiết
---------	---------------------

Ô lớn

$E(msK) = \sigma_e^2 + b\sigma_{AK}^2 + ab\sigma_K^2$ $E(msA) = \sigma_e^2 + b\sigma_{AK}^2 + r\sigma_{AB}^2 + br\sigma_A^2$ $E(msAK) = \sigma_e^2 + b\sigma_{AK}^2$	Không có kiểm định chính xác (Có thể dùng kiểm định gần đúng)
--	--

Ô nhỏ

$E(msB) = \sigma_e^2 + r\sigma_{AB}^2 + a\sigma_B^2$ $E(msAB) = \sigma_e^2 + r\sigma_{AB}^2$ $E(msE) = \sigma_e^2$	$F_{tnB} = msB / msAB$ so với $F(\alpha, dfB, dfAB)$ $F_{tnAB} = msAB / msE$ so với $F(\alpha, dfAB, dfE)$
--	---

Mô hình hỗn hợp: nhân tố A tất định, nhân tố B ngẫu nhiên

Giả thiết H_{0A} : mọi a_i bằng không, đối thiết H_{1A} : có a_i khác không

Giả thiết H_{0B} : σ^2_B bằng không, đối thiết H_{1B} : σ^2_B khác không

Giả thiết H_{0AB} : σ^2_{AB} bằng không, đối thiết H_{1AB} : σ^2_{AB} khác không

Kỳ vọng	Kiểm định giả thiết
---------	---------------------

Ô lớn

$E(msK) = \sigma_e^2 + ab\sigma_K^2$ $E(msA) = \sigma_e^2 + b\sigma_A^2 + r\sigma_{AB}^2 + (br\Sigma a_i^2)/(a-1)$ $E(msAK) = \sigma_e^2 + b\sigma_{AK}^2$	Không có kiểm định chính xác (Có thể dùng kiểm định gần đúng)
--	--

Ô nhỏ

$E(msB) = \sigma_e^2 + a\sigma_B^2$ $E(msAB) = \sigma_e^2 + r\sigma_{AB}^2$ $E(msE) = \sigma_e^2$	$F_{tnB} = msB / msE$ so với $F(\alpha, dfB, dfE)$ $F_{tnAB} = msAB / msE$ so với $F(\alpha, dfAB, dfE)$
---	---

Mô hình hỗn hợp: nhân tố A ngẫu nhiên, nhân tố B tất định

Giả thiết H_{0A} : σ^2_A bằng không, đối thiết H_{1A} : σ^2_A khác không

Giả thiết H_{0B} : các b_j bằng không, đối thiết H_{1B} : có b_j khác không

Giả thiết H_{0AB} : σ^2_{AB} bằng không, đối thiết H_{1AB} : σ^2_{AB} khác không

Kỳ vọng	Kiểm định giả thiết
Ô lớn	
$E(msK) = \sigma^2_e + b\sigma^2_{AK} + ab\sigma^2_K$ $E(msA) = \sigma^2_e + b\sigma^2_{AK} + br\sigma^2_A$ $E(msAK) = \sigma^2_e + b\sigma^2_{AK}$	$F_{tnA} = msA / msAK$ so với $F(\alpha, dfA, dfAK)$
Ô nhỏ	
$E(msB) = \sigma^2_e + r\sigma^2_{AB} + (ar\Sigma b_j^2)/(b-1)$ $E(msAB) = \sigma^2_e + r\sigma^2_{AB}$ $E(msE) = \sigma^2_e$	$F_{tnB} = msB/msAB$ so với $F(\alpha, dfB, dfAB)$ $F_{tnAB} = msAB/msE$ so với $F(\alpha, dfAB, dfE)$

Các mô hình một nhân tố khối ngẫu nhiên đầy đủ (RCBD) hay ô vuông La tinh chủ yếu dùng với nhân tố tất định và cách phân tích không khác gì cách trình bày trong các giáo trình về phương pháp thí nghiệm và toán sinh học hiện đang dùng.

Trên đây là một số mô hình hay dùng trong phân tích phương sai.

Có một điểm quan trọng cần chú ý là trong mô hình phân tích phương sai một nhân tố việc tách các tổng bình phương SSTO thành các bình phương thành phần SSA và SSE được thực hiện theo như cách đã trình bày ở chương 2 **dù số lần lặp bằng nhau hay không**.

Đối với trường hợp hai nhân tố thì cách phân tích SSTO thành các phương sai thành phần (SSR, SSA, SSB, SSAB, SSE trong mô hình thiết kế chéo nhau Crossed hay mô hình chia băng Strip plot, hay SSA, SSB, SSAB, SSE ô ngang, SSE ô dọc trong mô hình chia ô Strip plot, hay SSA, SSB, SSE trong mô hình phân cấp Hierarchical) như đã trình bày trong chương 3 và chương 4 chỉ đúng **khi số lần lặp bằng nhau**, hay có thể thay đổi một ít để tính cho trường hợp số lần lặp tỷ lệ với nhau, thường gọi là **trường hợp cân đối** (Balanced). Nếu không **cân đối** thì việc tách các tổng bình phương phức tạp hơn nhiều và phải sử dụng đến các thuật toán dùng trong **phân tích hồi quy bội tuyến tính**.

Đối với 3 nhân tố thì có nhiều mô hình và việc phân tích dài hơn. Giáo trình này chỉ tập trung vào mô hình 3 nhân tố chéo nhau (Crossed), phân cấp (Hierarchical), chia ô lớn, ô vừa, ô nhỏ (Split split plot), chia băng rồi chia ô (Strip split plot). Việc phân tích mô hình tất định hay ngẫu nhiên không phức tạp lắm nhưng mô hình hỗn hợp thì phức tạp hơn nhiều.

Trong các bộ chương trình chuyên về thống kê như Minitab, SPSS, Statistica, Irristat Ver 4.0 đối với mỗi nhân tố phải khai báo rõ tất định hay ngẫu nhiên, chéo nhau hay phân cấp để chương trình xử lý và đưa ra kết luận. Một số trường hợp không có kiểm định F chính xác phải dùng các kiểm định F gần đúng trong đó mẫu số là tổng hợp của một số các bình phương trung bình.

2- KIỂM ĐỊNH CÁC GIẢ THIẾT CƠ BẢN TRONG PHÂN TÍCH PHƯƠNG SAI

Trong mô hình phân tích phương sai một nhân tố có 3 giả thiết đối với sai số ngẫu nhiên e_{ij} . Giả thiết các e_{ij} độc lập và phân phối chuẩn với kỳ vọng bằng không thường được chấp nhận trong thực tế, riêng giả thiết các phương sai của e_{ij} bằng nhau (từ đó suy ra các phương sai của các mức bằng nhau) thì hay bị vi phạm. Để kiểm định giả thiết này có nhiều phương pháp nhưng hay dùng nhất là kiểm định Bartlett và kiểm định Levene.

2.1 Kiểm định Bartlett

Kiểm định Bartlett được thực hiện theo bảng sau:

Mức	Số lần lặp r_i	Phương sai s_i^2	$(r_i - 1)s_i^2$	$(r_i - 1)\ln(s_i^2)$	$1/(r_i - 1)$
A_1	r_1	s_1^2	$(r_1 - 1)s_1^2$	$(r_1 - 1)\ln(s_1^2)$	$1/(r_1 - 1)$
A_2	r_2	s_2^2	$(r_2 - 1)s_2^2$	$(r_2 - 1)\ln(s_2^2)$	$1/(r_2 - 1)$
...	...	...	...	...	...
Tổng	n		$\sum_{i=1}^a (r_i - 1)s_i^2$	$\sum_{i=1}^a (r_i - 1)\ln(s_i^2)$	$\sum_{i=1}^a 1/(r_i - 1)$

Tính phương sai chung s_C^2

$$s_C^2 = \frac{\sum_{i=1}^a (r_i - 1)s_i^2}{(n - a)} \quad M = (n - a)\ln(s_C^2) - \sum_{i=1}^a (r_i - 1)\ln(s_i^2)$$

$$C = 1 + \frac{1}{3(a - 1)} \left(\sum_{i=1}^a \frac{1}{(r_i - 1)} - \frac{1}{n - a} \right) \quad \chi_{TN}^2 = \frac{M}{C}$$

Ngưỡng so sánh $\chi_{LT}^2 = \chi^2(\alpha, a - 1)$

Nếu $\chi_{TN}^2 \leq \chi_{LT}^2$ thì chấp nhận giả thiết H_0 : "Các phương sai bằng nhau", nếu ngược lại thì bác bỏ H_0

Mức	r_i	s_i^2	$(r_i - 1)s_i^2$	$(r_i - 1)\ln(s_i^2)$	$1/(r_i - 1)$
A	7	0.0631	0.3785	-16.5801	0.1667
B	8	0.0183	0.1282	-27.9983	0.1429
C	6	0.0281	0.1403	-17.8671	0.2000
D	8	0.0363	0.2542	-23.2077	0.1429
Tổng	$n=29$		0.9013	-85.6532	0.6524
Phương sai chung		$s_c^2 =$	0.03605		
		$M =$	2.58227	$C =$	1.0680

$$\chi^2_{tn} \quad 2.41776 \quad \chi^2_{lt} \quad 7.815$$

Kết luận: Chấp nhận H_0 : Các phương sai bằng nhau

2.2 Kiểm định Levene

Kiểm định Levene được coi là ít phụ thuộc giả thiết e_{ij} phân phối chuẩn hơn là kiểm định Bartlett. Kiểm định này thực hiện như sau:

Nhân tố A có a mức, các số liệu x_{ij} của mức i có trung bình $\bar{x}_i$.

Lấy các độ lệch tuyệt đối $|x_{ij} - \bar{x}_i|$ (cũng có tài liệu lấy độ lệch bình phương)

Phân tích phương sai bằng các độ lệch tuyệt đối trên. Nếu $F_{tn} \leq F_{lt}$ thì chấp nhận giả thiết H_0 : “ Các phương sai bằng nhau”, nếu ngược lại thì bác bỏ H_0

Thí dụ	Số liệu xij								T bình
A	7.4	6.18	6.86	7.76	6.39	5.95	7.48	6.86	
B	8.84	6.69	7.12	7.42	6.83	5.06	5.35	6.76	
C	8.09	7.96	5.31	7.39	0.51	7.84	6.28	6.20	
D	7.55	5.65	6.92	6.5	5.46	7.4	8.37	6.84	
Bảng các độ lệch tuyệt đối									
A	0.54	0.68	0.00	0.90	0.47	0.91	0.62		
B	2.08	0.07	0.36	0.66	0.07	1.70	1.41		
C	1.89	1.76	0.89	1.19	5.69	1.64	0.08		
D	0.71	1.19	0.08	0.34	1.38	0.56	1.53		

Bảng phân tích phương sai						
Nguồn	SS	df	MS	F _{tn}	F _{lt}	
Nhân tố	6.782	3	2.261	2.11	3.01	
Sai số	25.646	24	1.0686			
Toàn bộ	32.428	27				

Kết luận: Chấp nhận H_0 : Các phương sai bằng nhau

2.3 Các biến đổi thường dùng trong phân tích phương sai

Khi các biến không phân phối chuẩn người ta thường dùng một số biến đổi như sau:

Nếu biến p là **tỷ lệ phần trăm%** thì dùng **biến đổi** $\arcsin \sqrt{p/100}$

Nếu biến X có độ nghiêng (Skewness) âm (nghiêng trái tức là sườn trái thoải, sườn phải dốc đứng) thì dùng các **biến đổi lũy thừa** $X^2, X^3, \dots$. Nếu độ nghiêng dương (nghiêng phải) thì **biến đổi căn bậc hai** $\sqrt{X}$ (hoặc $\sqrt{X+a}$ nếu có trị X âm), $\frac{1}{\sqrt{X}}$ (hay $\frac{1}{\sqrt{X+a}}$), **biến đổi lôgarit** $\ln(X)$, $\log(X)$ (hay $\ln(X+a)$, $\log(X+a)$ nếu có trị X âm), biến đổi nghịch đảo $1/X$ (hay $1/(X+a)$)

Các phần mềm thống kê thường dùng **phép biến đổi tổng quát Box-Cox**

$$Y = (X^\lambda - 1) / \lambda \quad (\lambda \neq 1)$$

$$Y = \ln(X) \quad (\lambda = 0)$$

Tham số λ được chọn sao cho biểu thức

$$L = -\frac{v}{2} \ln s_T^2 + (\lambda - 1) \frac{v}{n} \sum \ln X$$

đạt giá trị cực đại.

(s_T^2 là phương sai của biến Y với v bậc tự do)

Biến đổi Box-Cox bao gồm các biến đổi kể trên như biến đổi căn bậc hai ($\lambda = 1/2$), biến đổi nghịch đảo ($\lambda = -1$), biến đổi lôgarit ($\lambda = 0$)

Trong phân tích phương sai khi giả thiết các phương sai của các mức s_i^2 bằng nhau bị bác bỏ (qua kiểm định Bartlett hay Levene) thì có thể thực hiện một số phép biến đổi sau:

Nếu phương sai s_i^2 của các mức tỷ lệ với trung bình cộng $\bar{x}_i$ thì biến đổi căn bậc hai ($\sqrt{X}$ hay $\sqrt{X+a}$)

Nếu độ lệch chuẩn s_i của các mức tỷ lệ với trung bình cộng $\bar{x}_i$ thì biến đổi lôgarit ($\ln(X)$ hay $\ln(X+a)$)

Thí dụ về biến đổi $\arcsin \sqrt{p}$

Công thức	Khối						thành	Psai
	K1	K2	K3	K4	K5	K6	Tbình	Psai
A	42.4	34.3	24.1	39.5	55.5	49.1	40.82	122.07
B	33.3	33.3	5.0	26.3	30.2	28.6	26.12	114.40
C	8.5	21.9	6.2	16.0	13.5	15.4	13.58	31.73
D	16.6	19.3	16.6	2.1	11.1	11.1	12.80	38.28

	Biến đổi $\arcsin \sqrt{p / 100}$						Tbình	Psai
A	40.6285	35.85	29.4	38.94	48.16	44.48	39.58	43.33
B	35.244	35.24	12.92	30.85	33.34	32.33	29.99	72.79
C	16.9506	27.9	14.42	23.58	21.56	23.11	21.25	23.68
D	24.0435	26.06	24.04	8.332	19.46	19.46	20.23	41.16

Nguồn	SS	df	MS	F	xác suất
nhân tố	1462	3	487.2	13.38	0.00016
khối	358.6	5	71.71	1.969	0.142
sai số	546.3	15	36.42		
Toàn bộ	2366	23			

Thí dụ về biến đổi căn bậc hai (khi phương sai tỷ lệ với trung bình cộng)

Công thức	Khối				Tbình	Psai	Tỷ lệ
	K1	K2	K3	K4	xtb	s ²	s ² / xtb
A	438	442	319	380	394.8	3352.9	8.5
B	538	422	377	315	413.0	8868.7	21.5
C	77	61	157	52	86.8	2300.3	26.5
D	17	31	87	16	37.8	1124.9	29.8
E	18	26	77	20	35.3	786.25	22.3

Có thể coi phương sai (s²) tỷ lệ với trung bình cộng (xtb)

Biến đổi căn bậc hai

	K1	K2	K3	K4	xtb	s ²
A	20.93	21.02	17.86	19.49	19.8	2.2

B	23.19	20.54	19.42	17.75	20.2	5.2
C	8.77	7.81	12.53	7.21	9.1	5.7
D	4.12	5.57	9.33	4.00	5.8	6.2
E	4.24	5.10	8.77	4.47	5.6	4.5

Bảng phân tích phương sai

Nguồn	SS	df	MS	F_{tn}	F_{lt}
<i>Nhân tố</i>	866.9556	4	216.739	53.300	3.259
<i>Khối</i>	22.61308	3	7.538	1.854	3.490
<i>Sai số</i>	48.7966	12	4.066		
<i>Toàn bộ</i>	938.3653	19			

Thí dụ về biến đổi lôgarit (khi độ lệch chuẩn tỷ lệ với trung bình cộng)

			Đôi	chứng		Trung bình	Phương sai
X	20	40	50	60	80	50	500
Ln(X)	2.9957	3.6689	3.9120	4.0943	4.3820	3.8105	0.27535
			Thí nghiệm		=1,5x Đôi Chứng		
X	30	60	75	90	120	75	1125
Ln(X)	3.4012	4.0943	4.3175	4.4998	4.7875	4.2201	0.2740

Vì thí nghiệm bằng 1,5 x đôi chứng nên rõ ràng trung bình cộng của thí nghiệm gấp 1,5 lần đôi chứng và độ lệch chuẩn của thí nghiệm cũng gấp 1,5 lần đôi chứng, nói cách khác độ lệch chuẩn tỷ lệ với trung bình cộng.

Sau biến đổi độ lệch chuẩn của đôi chứng và thí nghiệm xấp xỉ bằng nhau

3- SO SÁNH CÁC TRUNG BÌNH SAU PHÂN TÍCH PHƯƠNG SAI

Khi phân tích phương sai một nhân tố với a mức: $A_1, A_2, \dots, A_a$ được bảng phân tích phương sai. So F_{tn} với $F(\alpha, df_A, df_E)$ trong bảng. Nếu $F_{tn} \leq F(\alpha, df_A, df_E)$

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 không bằng nhau**”.

Sau khi phân tích phương sai và kết luận “**Các trung bình không bằng 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.

Trong phân tích phương sai một nhân tố **msE** được ký hiệu là **se²**, se gọi là sai số thí nghiệm, **dfE** là bậc tự do của sai số.

Khi 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, ...)**.

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 chung là: **So sánh các trung bình của nhân tố A với sai số se và bậc tự do dfE**.

3.1 So sánh tất cả các trung bình (multiple comparison) theo một ngưỡng.

Nếu nhân tố có a mức thì có a trung bình và như vậy tất cả có $a(a-1)/2$ cặp trung bình cần so sánh. Sẽ chọn một ngưỡng chung cho mọi cặp so sánh(trước hết xét trường hợp số lần lặp bằng nhau).

Sai số của trung bình của một mức:

$$s\bar{x} = \frac{se}{\sqrt{r}}$$

Có nhiều phương pháp, thường gọi là kiểm định (test), để so sánh các trung bình

Kiểm định LSD

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

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

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

$$LSD = t(\alpha/2, dfE) \times \sqrt{msE \times \frac{2}{r}} = t \times s(\bar{x}_2 - \bar{x}_1) \quad (2)$$

Nếu $|\bar{x}_2 - \bar{x}_1| \leq \text{LSD}$ kết luận hai trung bình m_i và m_j bằng nhau

Nếu $|\bar{x}_2 - \bar{x}_1| > \text{LSD}$ kết luận hai trung bình m_i và m_j khác nhau.

Gọi hệ số $t(\alpha/2, dfE)$ 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 :

Nếu thí nghiệm chỉ có 2 mức thì xác suất có kết luận sai: “hai trung bình khác nhau” khi giả thiết H_0 đúng tức “là hai trung bình bằng nhau” sẽ là α (mức ý nghĩa).

Nếu thí nghiệm có $a > 2$ mức và so sánh mọi cặp trung bình theo LSD với mức ý nghĩa α thì xác suất để có ít nhất một cặp trung bình bị kết luận sai là “khác nhau” trong khi giả thiết H_0 là “các trung bình bằng nhau” sẽ không còn là α mà lớn hơn nhiều lần. Thí dụ khi so hai trung bình với $\alpha = 0,05$ nếu có 4 mức, tức là 6 cặp trung bình, xác suất có ít nhất một 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 đã có ý đồ so sánh từ khi thiết kế thí nghiệm chứ không nên dùng để so sánh tất cả các cặp trung bình sau khi xử lý dữ liệu.

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_{α} , dfE là bậc tự do của sai số, $F(\alpha, dfA, dfE)$ 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.

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 đối với cặp “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).

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

$$q(\alpha, p, dfE) = \frac{(y_{\max} - y_{\min})}{s\bar{y}}$$

$q(\alpha, p, dfE)$ được cho trong các bảng thống kê với tên gọi bảng một số phân vị trong phân phối biên độ kiểu Student (Selected percentiles of Studentized range distributions) hay các điểm phần trăm trên của biên độ kiểu Student (Upper percentage points of the Studentized range), trong đó:

α là mức ý nghĩa của kiểm định

p là tham số của phân phối bằng số trung bình cần so sánh (ở đây $p=a$)

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.

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'

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

Thí dụ có 5 trung bình cần so sánh, tất cả có $k = 5 \cdot 4/2 = 10$ cặp

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

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

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

Thí dụ 1

5 trung bình với số lần lặp $r = 14$; $\alpha = 0,05$; $dfA = 4$; $dfE = 65$

$$t(0,025, 65) = 1,997 \quad F(0,05, 4, 65) = 2,513 \quad q(0,05, 5, 65) = 3,975$$

$$m_1 = 4,050; m_2 = 4,429; m_3 = 4,900; m_4 = 5,343; m_5 = 5,921; \text{se} = \mathbf{1,633}$$

Ngưỡng:

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

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

$$\text{Bonferroni} \quad 2,906 \times 1,633 \times \sqrt{\frac{2}{14}} = 1,794$$

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

Scheffé

Thí dụ 2 6 công thức có trung bình :

A ₁	A ₂	A ₃	A ₄	A ₅	A ₆
13,3	14,6	18,7	19,9	24,0	28,8

$$a = 6 \quad r = 5 \quad dfA = 5 \quad dfE = 24$$

$$s_e^2 = 11,79 \quad F(0,05,5,24) = 2,62$$

$$t(0,025,24) = 2,064 \quad q(0,05,6,24) = 4,37$$

$$\text{Số cặp phải so sánh } k = 6 \cdot 5 / 2 = 15 \quad \alpha' = \alpha/30 = 0,05/30 = 0,00167$$

$$t(0,00167, 24) = 3,28$$

$$\text{Ngưỡng theo LSD} \quad 2,064 \times \sqrt{\frac{11,79 \times 2}{5}} = 4,482$$

13,3	14,6	18,7	19,9	24,0	28,8
a	a				
	b	b			
		c	c		
			d	d	
					e

$$\text{Ngưỡng theo Tukey} \quad \frac{4,37}{\sqrt{2}} \times \sqrt{\frac{11,79 \times 2}{5}} = 6,71$$

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

$$\text{Ngưỡng theo Bonferroni} \quad 3,28 \times \sqrt{\frac{11,79 \times 2}{5}} = 7,123$$

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

$$\text{Ngưỡng theo Scheffé} \quad \sqrt{\frac{5 \times 2,62 \times 11,79 \times 2}{5}} = 7,86$$

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

Ba cách so sánh theo Tukey, Bonferroni và Scheffé cho kết quả chia nhóm tương tự nhau.

3.2 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 biên độ (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 $p = 4 \dots$

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

Các ngưỡng so sánh

$$W(p) = q(\alpha, p, dfE) \quad s\bar{x} = q(\alpha, p, dfE) \times \sqrt{\frac{se^2}{r}}$$

trong đó $q(\alpha, p, dfE)$ là trị lấy trong bảng các phân vị trong phân phối biên độ kiểu Student đã giới thiệu ở mục Kiểm định Tukey Cramer.

α là mức ý nghĩa, p là tham số, dfE là bậc tự do của sai số, $s\bar{x}$ là sai số của trung bình.

Thí dụ 2 với $dfE = 24$; $se^2 = 11,79$; $r = 5$ ta có

$$s\bar{x} = \sqrt{\frac{11,79}{5}} = 1,536$$

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
$W(p)$	4,485	5,422	5,99	6,405	6,712

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
a	a	a			
	b	b	b		
		c	c	c	
				d	d

Kiểm định Duncan

Kiểm định Duncan dùng công thức tương tự kiểm định S-N-K nhưng chọn mức ý nghĩa α' tùy 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 biên độ kiểu Student, từ đó tìm được ngưỡng so sánh

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

p	2	3	4	5	6
$q(\alpha', p, 24)$	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

3.3 So sánh nhiều trung bình khi số lần lặp không bằng nhau

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

$$L = t(\alpha/2, dfE) \times \frac{se}{\sqrt{r_i}} = t \times s\bar{x}_i \quad (1')$$

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

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

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

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 $\sqrt{\frac{se^2}{r}}$ bằng $\sqrt{se^2(\frac{1}{r_i} + \frac{1}{r_j})}$

3.4 Một số kiểm định khác

Kiểm định Dunnett

Khi so sánh các trung bình nếu có đối chứng thì nên dùng cách 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 định phạm vi của REGW.

Một số kiểm định có cách tiếp cận rất mới và các tính toán khá phức tạp.

5- 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 so sánh các trung bình sau phân tích phương sai 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 có những điểm không chặt chẽ trong chứng minh.

Tập I số 4/2003 tạp chí khoa học kỹ thuật nông nghiệp DHNN1. Nguyen dinh Hien.
Mô hình cố định, ngẫu nhiên và hỗn hợp trong phân tích phương sai.
Use of fixed, random and mixed models in analysis of variance