



CHƯƠNG 4

PHÂN TÍCH PHƯƠNG SAI MỘT YẾU TỐ (ONE WAY ANOVA)

ThS HUỖNH TỔ UYÊN

Phân tích phương sai?

Ví dụ1:

Để xét xem giới tính có ảnh hưởng đến kết quả học tập hay không, người ta so sánh điểm trung bình của nhóm sinh viên nam và sinh viên nữ

Nam	Nữ
6.5	5.7
7.0	6.8
6.7	8.2
8.0	7.7
5.6	6.4
6.5	

Kiểm định xem giới tính có ảnh hưởng đến kết quả học tập?

Kiểm định sự khác nhau giữa điểm trung bình của 2 nhóm sinh viên này

Phân tích phương sai?

Ví dụ2:

Để xét xem thời gian làm thêm có ảnh hưởng đến kết quả học tập hay không, người ta điều tra mẫu sau:

Nhóm 1: làm thêm ít <6 giờ /tuần	Nhóm 2: làm thêm TB 6-12 giờ/tuần	Nhóm 3: làm thêm nhiều >12 giờ/tuần
6.3	7.2	6,3
7.0	6.6	5.8
6.5	6.1	6.0
6.6	5.8	5.5
7.2	6.8	5.2
6.9	7.1	6.5
6.4	5.9	5.3
		6.2

Kiểm định xem thời gian làm thêm có ảnh hưởng đến kết quả học tập không?

Phân tích phương sai?

Phân tích phương sai là phương pháp phân tích sự ảnh hưởng của 1 hay nhiều yếu tố nguyên nhân đến 1 yếu tố kết quả

Phân tích phương sai là phương pháp kiểm định sự bằng nhau của trung bình nhiều tổng thể

1. Phân tích phương sai 1 yếu tố

Phân tích phương sai một yếu tố là phân tích ảnh hưởng của một yếu tố nguyên nhân (định tính) đến một yếu tố kết quả (định lượng)

Ví dụ:

Thời gian làm thêm (yếu tố nguyên nhân – định tính) ảnh hưởng đến kết quả học tập (yếu tố kết quả - định lượng)

1. Phân tích phương sai 1 yếu tố

BÀI TOÁN:

Giả sử ta có k nhóm gồm n_1, n_2, \dots, n_k phần tử được chọn từ k tổng thể.

$\mu_1, \mu_2, \dots, \mu_k$ là các trung bình của k tổng thể đó

x_{ij} là giá trị (quan sát) thứ j của nhóm thứ i.

Bảng giá trị quan sát của k nhóm:

NHÓM			
1	2		k
x_{11}	x_{21}	x_{ij}	x_{k1}
...
x_{1n_1}	x_{2n_2}	...	x_{kn_k}
		...	

VD:
slide 3

Giả sử k tổng thể có phân phối chuẩn, có phương sai bằng nhau, các mẫu là độc lập

1. Phân tích phương sai 1 yếu tố

$$\begin{cases} H_0 : \text{yếu tố kq không bị ảnh hưởng bởi yếu tố đang xét} \\ H_1 : \text{yếu tố kq có bị ảnh hưởng} \end{cases}$$

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_1 : \text{tồn tại ít nhất 1 cặp trung bình khác nhau} \end{cases}$$

Ví dụ 2.1

$$\begin{cases} H_0 : \text{thời gian làm thêm không ảnh hưởng đến kết quả học tập} \\ H_1 : \text{thời gian làm thêm có ảnh hưởng đến kết quả học tập} \end{cases}$$

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 \\ H_1 : \text{tồn tại ít nhất 1 cặp trung bình khác nhau} \end{cases}$$

1. Phân tích phương sai 1 yếu tố

Bước 1: Tính giá trị trung bình \bar{X}_j cho từng nhóm và \bar{X} chung cho tất cả các nhóm

$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k \bar{X}_j n_j$$

$$n = \sum_{i=1}^k n_j$$

Ví dụ 2.2

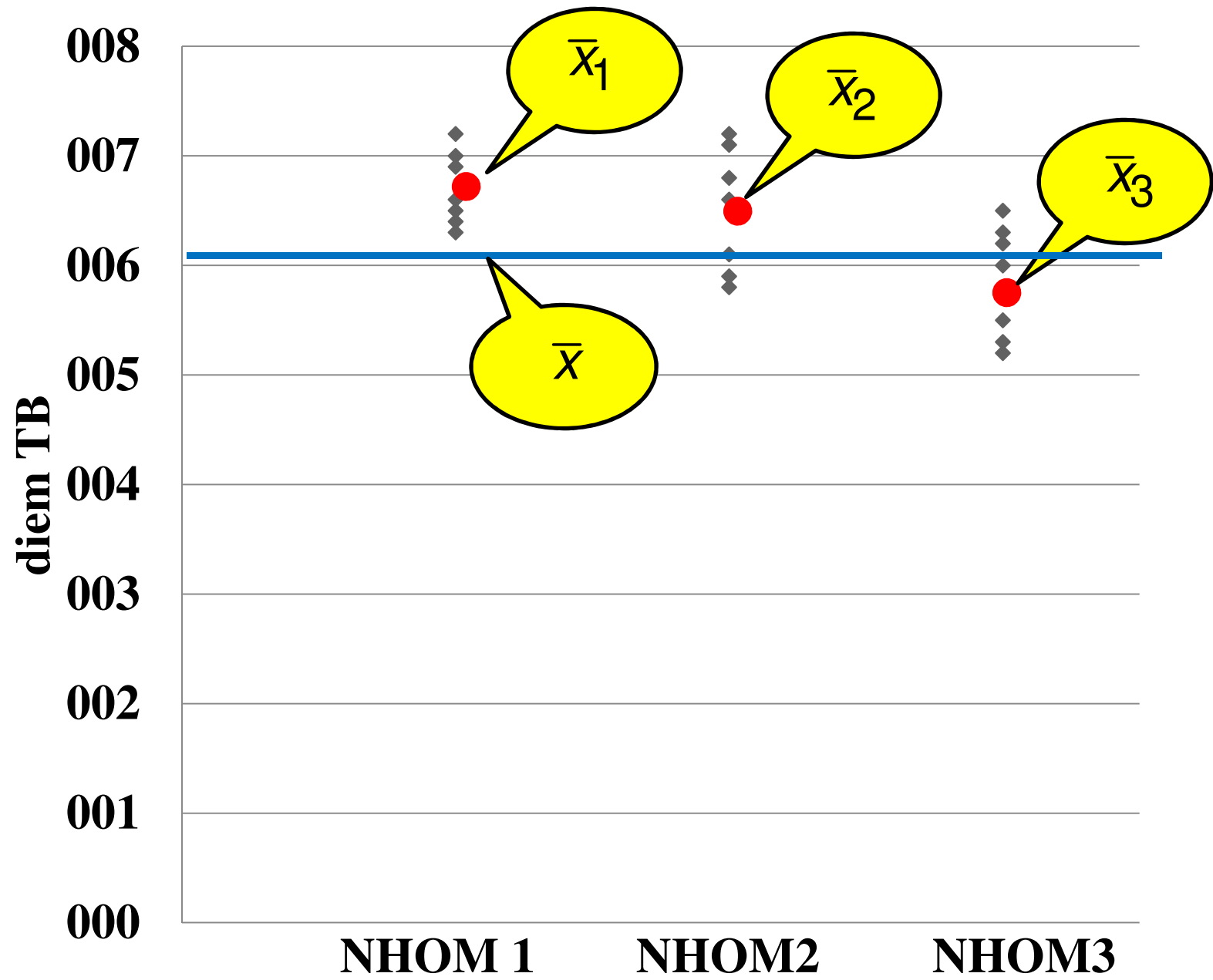
Bước 1. Tính trung bình của từng nhóm và trung bình chung của ba nhóm.

$$\bar{X}_1 = \frac{46,9}{7} = 6,7$$

$$\bar{X}_2 = \frac{45,5}{7} = 6,5$$

$$\bar{X}_3 = \frac{46,8}{8} = 5,85$$

$$\bar{X} = \frac{46,9 + 45,5 + 46,8}{7 + 7 + 8} = 6,3273$$



1. Phân tích phương sai 1 yếu tố

Bước 2:

- Tính sự biến thiên (tổng bình phương độ lệch) giữa nội bộ nhóm

$$SS_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

$$SSW = \sum_{i=1}^k SS_i$$

Tổng biến thiên nội bộ
của tất cả các nhóm

Ví dụ 2.3

Bước 2.

- Tính sự biến thiên giữa nội bộ nhóm:

$$SS_1 = 0,68$$

$$SS_2 = 1,96$$

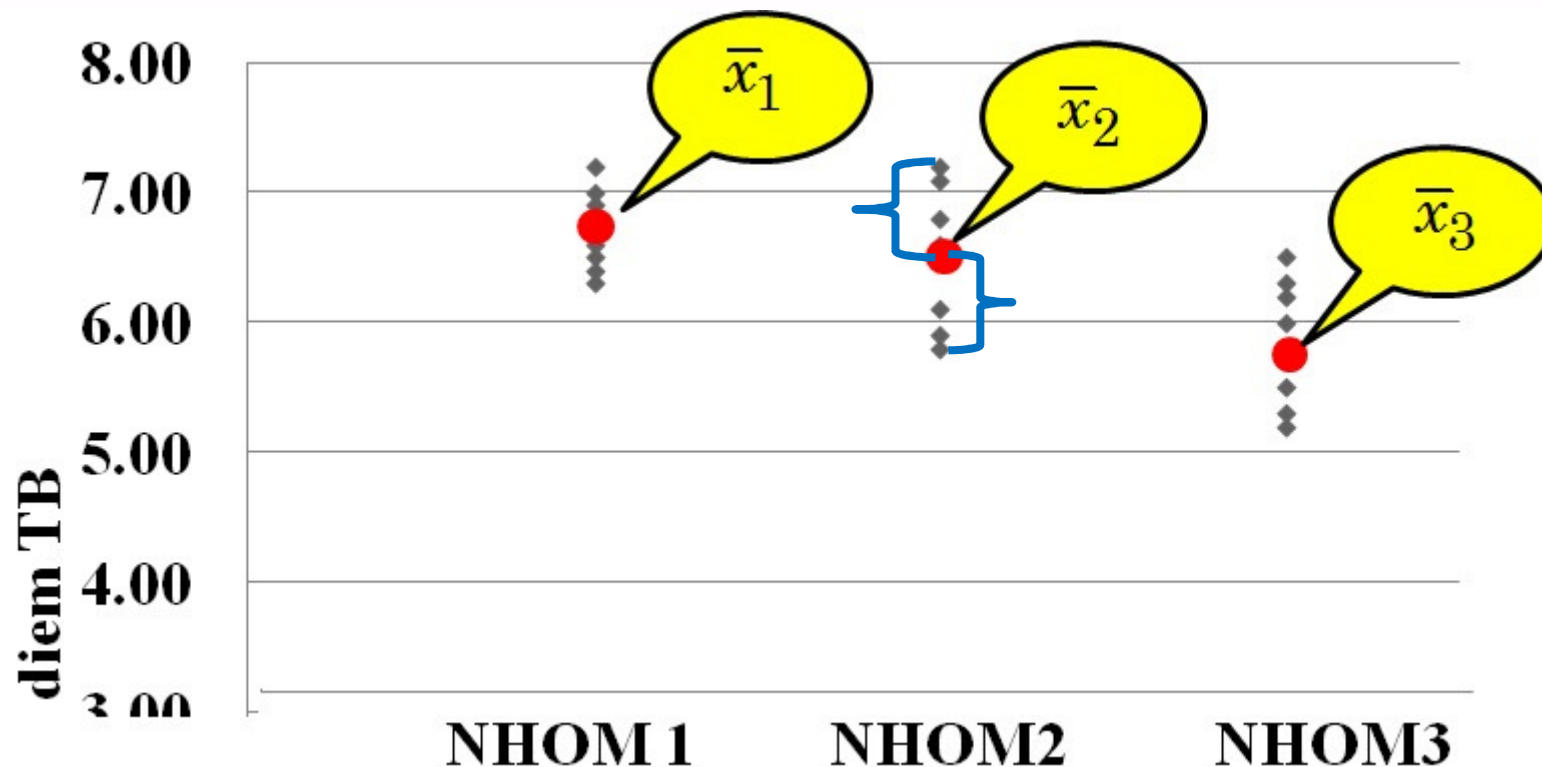
$$SS_3 = 1,62$$

$$\begin{aligned}SSW &= SS_1 + SS_2 + SS_3 \\&= 0,68 + 1,96 + 1,62 \\&= 4,26\end{aligned}$$

1. Phân tích phương sai 1 yếu tố

Nhận xét:

SSW = tổng biến thiên nội bộ của cả 3 nhóm = sự biến thiên gây ra bởi các yếu tố khác yếu tố mà ta đang nghiên cứu



1. Phân tích phương sai 1 yếu tố

Bước 2:(tt)

- Tính sự biến thiên giữa các nhóm

$$SSG = \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 n_i$$

Ví dụ 2.3(tt)

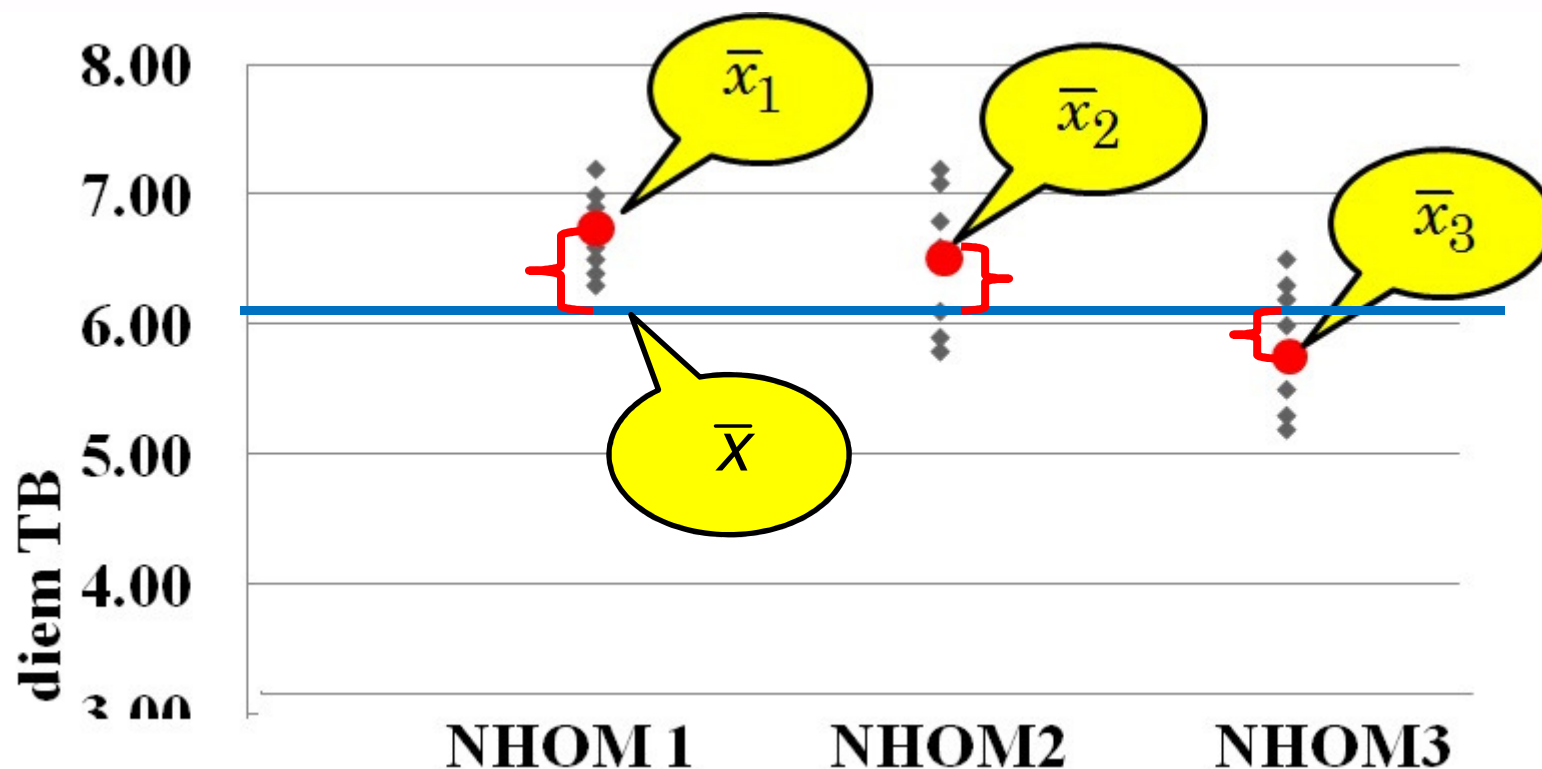
- Tính sự biến thiên giữa các nhóm:

$$\begin{aligned}SSG &= (6,7 - 6,3273)^2 \cdot 7 \\&\quad + (6,5 - 6,3273)^2 \cdot 7 \\&\quad + (5,85 - 6,3273)^2 \cdot 8 \\&= 3,004\end{aligned}$$

1. Phân tích phương sai 1 yếu tố

Nhận xét:

SSG = sự biến thiên gây ra bởi sự khác nhau giữa các nhóm = sự biến thiên gây ra bởi các yếu tố mà ta đang nghiên cứu



1. Phân tích phương sai 1 yếu tố

Gọi SST là tổng biến thiên của 1 quan sát bất kì so với giá trị trung bình

$$SST = SSW + SSG$$

Tổng biến thiên = biến thiên do các yếu tố khác(SSW) + biến thiên do yếu tố đang nghiên cứu(SSG)

Nhận xét:

Nếu phần biến thiên do các yếu tố nghiên cứu tạo ra (SSG) lớn hơn phần biến thiên do các yếu tố khác tạo ra (SSW) thì chứng tỏ yếu tố ta đang nghiên cứu thật sự ảnh hưởng đến yếu tố kết quả
 \Rightarrow tăng khả năng bác bỏ H_0

1. Phân tích phương sai 1 yếu tố

Bước 3:

- Tính các phương sai

$$MSW = \frac{SSW}{n - k}$$

= phương sai
do các yếu tố
khác tạo ra

$$MSG = \frac{SSG}{k - 1}$$

= phương sai do
các yếu tố
nguyên cứu tạo ra

Ví dụ 2.4

- Tính các phương sai

$$MSW = \frac{4,26}{22 - 3} = 0,224$$

$$MSG = \frac{3,004}{3 - 1} = 1,502$$

1. Phân tích phương sai 1 yếu tố

Bước 4:

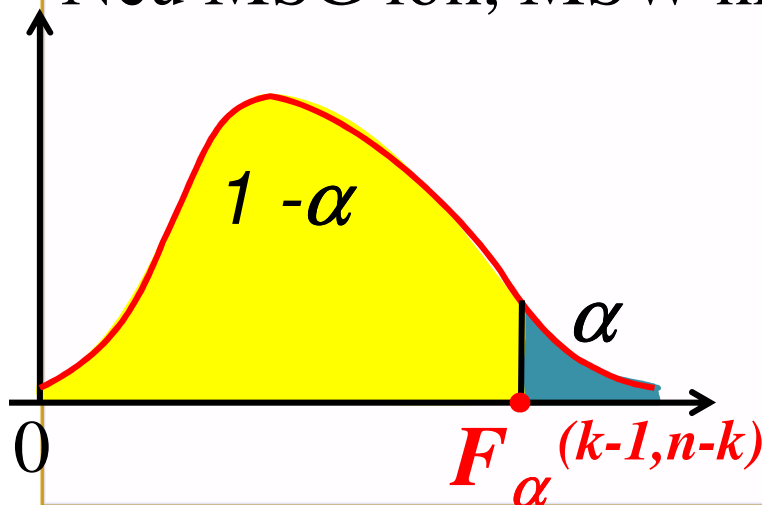
Kiểm định giả thuyết: Xét tỉ số 2 phương sai

$$F = \frac{MSG}{MSW}$$

$$= \frac{\text{phương sai do yếu tố nghiên cứu}}{\text{phương sai do yếu tố khác}}$$

Nhận xét:

Nếu MSG lớn, MSW nhỏ $\Rightarrow F$ lớn \Rightarrow bác bỏ H_0



Nếu $F > F_{\alpha}^{(k-1, n-k)}$
 \Rightarrow bác bỏ H_0

GIẢI VD 2

Ví dụ 2.5

- Tính tỉ số F

$$F = \frac{MSG}{MSW} = \frac{1,502}{0,224} = 6,7$$

$$F_{k-1, n-k, \alpha} = F_{2; 19; 0,05} = 3,52$$

Vì $F > F_{k-1, n-k, \alpha}$ nên ta bác bỏ $H_0 \Rightarrow$ Việc đi làm thêm có ảnh hưởng đến kết quả học tập của sinh viên

BẢNG ANOVA

Bảng ANOVA: Kết quả phân tích phương sai thường được trình bày dưới dạng bảng sau đây

Nguồn biến thiên	Tổng các độ lệch bình phương	Bậc tự do	Trung bình của các độ lệch bình phương (phương sai)	Giá trị kiểm định F
Giữa các nhóm	SSG	k-1	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSW}$
Trong nội bộ nhóm	SSW	n-k	$MSW = \frac{SSW}{n-k}$	
Tổng cộng	SST	n-1		

BẢNG ANOVA

SUMMARY

Groups	Count	Sum	Average	Variance		
Nhóm 1	7	46.9	6.7	0.11333		
Nhóm 2	7	45.5	6.5	0.32666		
Nhóm 3	8	46.8	5.85	0.23142		

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3.00364	2	1.50181	6.69825	0.00628	3.5219
Within Groups	4.26	19	0.22421			
Total	7.26364	21				

F tra bảng

Ví dụ áp dụng

Ví dụ 3: Cho 1 phần bảng ANOVA

a) Hãy hoàn tất bảng b) Với mức ý nghĩa 0,05 cho biết có sự khác biệt về trung bình các tổng thể hay không?

Nguồn biến thiên	Tổng các độ lệch bình phương	Bậc tự do	Trung bình của các độ lệch bình phương (phương sai)	Giá trị kiểm định F
Giữa các nhóm	16,9	6	?(1)	?(2)
Trong nội bộ nhóm	?(3)	41	?(4)	
Tổng cộng	45,2	?(5)		

Ví dụ áp dụng

Ví dụ 4: Nghiên cứu về thu nhập của các hộ gia đình ở ngoại thành, người ta chia ngoại thành thành 7 địa bàn dân cư khác nhau. Chọn ngẫu nhiên các hộ trong từng địa bàn và ghi nhận thu nhập. Địa bàn dân cư thứ ba có 13 hộ được chọn, các địa bàn còn lại đều chọn 19 hộ. Kết quả ANOVA như sau:

Source of Variation	SS	df	MS	F
Between Groups	187,2649	?(1)	?(2)	?(3)
Within Groups	?(4)	?(5)	?(6)	
Total	1269,6891			21

Ở mức ý nghĩa 1% có thể kết luận rằng thu nhập trung bình của các hộ gia đình ở các địa bàn dân cư khác nhau là như nhau được không?

ĐẶT VẤN ĐỀ

Trong trường hợp bác bỏ H_0 ta muốn kết luận về sự hơn kém của các trung bình thì ta cần phân tích sâu hơn

⇒ PHÂN TÍCH SÂU ANOVA
(KIỂM ĐỊNH TURKEY)

2. Phân tích sâu ANOVA – kiểm định Turkey

Mục đích phân tích sâu ANOVA

Trong kiểm định:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_k \\ H_1 : \text{tồn tại ít nhất 1 cặp trung bình khác nhau} \end{cases}$$

Nếu chấp nhận $H_0 \rightarrow$ phân tích xong

Bác bỏ $H_0 \rightarrow$ phân tích sâu hơn để xem trung bình các nhóm nào khác nhau, lớn hơn hay nhỏ hơn.

2. Phân tích sâu ANOVA – kiểm định Turkey

Phương pháp phân tích sâu ANOVA: kiểm định Turkey:

Với cùng mức ý nghĩa α , ta so sánh từng cặp trung bình để phát hiện những nhóm khác nhau.

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array} \right\}, \left\{ \begin{array}{l} H_0 : \mu_1 = \mu_3 \\ H_1 : \mu_1 \neq \mu_3 \end{array} \right\}, \dots, \left\{ \begin{array}{l} H_0 : \mu_{k-1} = \mu_k \\ H_1 : \mu_{k-1} \neq \mu_k \end{array} \right\}$$

Khi có k nhóm, số cặp cần phải kiểm định là tổ hợp chập 2 của k nhóm C_k^2

Ví dụ 2.6: Trong ví dụ 2 ta có k=3, vậy ta cần kiểm định $C_3^2 = 3$ cặp

$$\left\{ \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array} \right\}, \left\{ \begin{array}{l} H_0 : \mu_2 = \mu_3 \\ H_1 : \mu_2 \neq \mu_3 \end{array} \right\}, \left\{ \begin{array}{l} H_0 : \mu_1 = \mu_3 \\ H_1 : \mu_1 \neq \mu_3 \end{array} \right\}$$

2. Phân tích sâu ANOVA – kiểm định Turkey

Phương pháp phân tích sâu ANOVA: kiểm định Turkey:

Bước 1: tính $D_{ij} = |\bar{x}_i - \bar{x}_j|$

Bước 2: tính

$$T = q_{(k, n-k), \alpha} \sqrt{\frac{MSW}{n_{\min}}}$$

$$n_{\min} = \min \{n_1, n_2, \dots, n_k\}$$

trong đó q: phân phối Turkey (bảng tra 9)

Bước 3: Bác bỏ H_0 nếu $D_{ij} > T$

Ví dụ 2.7:

$$B1 : D_{12} = |\bar{x}_1 - \bar{x}_2| = 0,2$$

$$D_{23} = 0,65; D_{31} = 0,85$$

$$B2 : T = 3,59 \sqrt{\frac{0,224}{7}} = 0,6422$$

$$B3 : D_{12} < T \Rightarrow \mu_1 = \mu_2$$

$$D_{23} > T \Rightarrow \mu_2 \neq \mu_3$$

$$D_{31} > T \Rightarrow \mu_3 \neq \mu_1$$

$$\text{vì } \bar{x}_2, \bar{x}_1 > \bar{x}_3 \Rightarrow \mu_2 > \mu_3,$$

$$\mu_1 > \mu_3$$

BÀI TẬP TỔNG HỢP

Ban Giám hiệu một trường đại học muốn nghiên cứu ảnh hưởng của việc đi làm thêm đối với kết quả học tập của sinh viên. Một sinh viên đã thu thập thời gian đi làm thêm và kết quả học tập của một số sinh viên trong trường. Sinh viên có đi làm thêm được chia thành ba nhóm. Nhóm thứ nhất gồm 7 sinh viên có thời gian làm thêm ít, dưới 6 giờ / tuần. Nhóm thứ hai gồm 7 sinh viên có thời gian làm thêm vừa phải, từ 6 đến 12 giờ / tuần. Nhóm thứ ba gồm 8 sinh viên có thời gian làm thêm nhiều, trên 12 giờ / tuần. Điểm trung bình học tập của các sinh viên đó như sau.

Nhóm 1	Nhóm 2	Nhóm 3
6,3	7,2	6,3
7,0	6,6	5,8
6,5	6,1	6,0
6,6	5,8	5,5
7,2	6,8	5,2
6,9	7,1	6,5
6,4	5,9	5,3
		6,2
$\Sigma 46,9$	45,5	46,8

Phát biểu giả thuyết:
thời gian đi làm thêm
không ảnh hưởng đến
kết quả học tập của sinh
viên, nghĩa là điểm
trung bình học tập của
ba nhóm trên là như
nhau: .

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Ta kiểm định giả thuyết
trên với mức ý nghĩa
5%.